Properties and processing of natural scenes

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of

Doctor of Natural Sciences

presented by

CHRISTOPH KAYSER

Dipl. Math. ETH

born 18.01.1977

citizen of

Germany

accepted on the recommendation of

PD. Dr. Peter König, examiner
Prof. Dr. Kevan AC. Martin, co-examiner
Prof. Dr. Mathew E. Diamond, co-examiner

2004
Table of contents

Summary .................................................................................................................................................. 5
Zusammenfassung .................................................................................................................................. 7

I. Introduction .......................................................................................................................................... 9
I.1 The choice of the stimulus .................................................................................................................. 10
I.2 What are natural stimuli? .................................................................................................................. 12
I.3 Sensory systems and natural stimuli ............................................................................................... 12
I.4 Experimental aspects of an adaptation to natural stimuli ............................................................... 13
I.5 Theoretical aspects of an adaptation to natural stimuli ................................................................. 14
I.6 The next chapters ............................................................................................................................. 15

II. Properties of natural scenes ............................................................................................................ 17
II.1 Recording movies from a camera mounted to a cats' head ......................................................... 17
II.2 The second order statistics ............................................................................................................. 18
II.3 The motion in natural movies ......................................................................................................... 19
II.4 Statistics of oriented contours ....................................................................................................... 20
II.5 Summary of the results .................................................................................................................. 25
II.6 Directions of future work & conclusion ......................................................................................... 25
II.7 Publications ..................................................................................................................................... 27

III. Neuronal responses to complex stimuli in the primary visual cortex of awake cats ......................... 29
III.1 Recording population activities .................................................................................................... 29
III.2 Recording in the visual system of alert cats ................................................................................. 30
III.3 Set of stimuli .................................................................................................................................. 31
III.4 Basic properties of the local field potential .................................................................................. 33
III.5 Local field potential results for all stimuli .................................................................................... 33
III.6 Response properties of the high frequency activity ...................................................................... 35
III.7 Locking to the temporal structure of the stimulus ........................................................................ 35
III.8 Feature tuning and locking are prominent in complementary frequency ranges ...................... 37
III.9 Eye movements ............................................................................................................................ 38
III.10 Summary of the results ................................................................................................................. 39
III.11 Directions of future work & conclusion ..................................................................................... 40
III.12 Publications ................................................................................................................................... 42

IV. Hemodynamic responses to complex stimuli in the cat visual system ........................................... 43
IV.1 Imaging the cat visual cortex ......................................................................................................... 43
IV.2 BOLD responses to the different stimuli ....................................................................................... 45
IV.3 The origin of the BOLD signal ...................................................................................................... 47
IV.4 Comparing BOLD and electrophysiological measures of neuronal activity ............................. 48
IV.5 Summary of the results ................................................................................................................. 51
IV.6 Directions of future work & conclusion ....................................................................................... 52
IV.7 Publications ................................................................................................................................... 53

V. Optimality criteria linking properties of V1 cells and image statistics 55
V.1 Redundancy reduction and sparse coding ....................................................................................... 55
V.2 Predictive and stable coding .......................................................................................................... 57
V.3 Decorrelation - a form of lateral interaction ................................................................................... 58
V.4 Stable coding and simple cells ....................................................................................................... 58
V.5 Stable coding and complex cells .................................................................................................... 60
V.6 Learning the non-linearity .............................................................................................................. 62
Summary

Our brains continuously process sensory stimuli emerging from everyday interactions with our environment. The investigation of such sensory processing is a central part of today’s neuroscience research and involves many experimental and theoretical approaches. Classically, experiments use stimuli that are well chosen for the system under investigation. However, the typical stimuli used in most experiments are much simpler than the natural stimuli we encounter during everyday life. Motivated by the question as to whether and to what extent results obtained using artificial stimuli generalize to processing under everyday conditions, interest recently shifted towards using more natural stimuli. The present thesis follows this line of research and uses the visual system as a model to study properties of natural stimuli and their processing in sensory cortex.

Using a database of movies recorded from the perspective of a freely moving cat, we study statistical properties of natural scenes. Existing results on the power-law distribution of spatio-temporal frequencies are confirmed and it is shown that the motion in these natural movies is very irregular and has a high degree of temporal variation. Next, it is demonstrated that orientations are correlated across several degrees of visual angle and several hundreds of milliseconds. Collinear contours are more prevalent than parallel contours, and this arrangement is preserved over time. Overall, we demonstrate a good match between these interactions of orientations in natural scenes and known properties of the anatomical connectivity in V1 as well as with physiological and psychophysical effects. Thus, the statistical properties of natural scenes have many counterparts in the biological hardware and knowledge of this structure enhances our understanding of sensory processing. Especially in the temporal domain, natural movies have a complex structure whose impact on visual processing is largely unexplored.

In a series of electrophysiological experiments, we recorded local field potentials (Lfp) and multi-unit spikes in the primary visual cortex of alert cats. The Lfp response to classical drifting gratings differs in two respects from the response to various complex stimuli. First, the frequency range of the Lfp activated is narrow and restricted to the classical gamma band. Complex stimuli, on the other hand, activate a broad range of Lfp frequencies. Second, the temporal structure of the response to gratings is uniform, whereas it shows strong modulations during the presentation of natural movies. Similarly, in the spiking response there is a significant difference between gratings and natural movies. Thus, both average response strength and response timing are different for drifting gratings compared to natural scenes. This difference in basic response properties raises the question of which results actually can be generalized across different types of stimuli.

In the same experiments, natural movies are compared with a stimulus having the same second order, but random higher order statistics. The frequency range of the Lfp activated by these two stimuli is identical, as is the Lfp modulation strength. The spiking response for these two stimuli shows only a small difference. Thus, Lfp and multi-unit spiking
activity are not sensitive to the higher order statistical structure of the stimulus. Either this structure is only extracted on the level of single neurons, but the selectivity to it is lost a soon responses are averaged across neurons, or this structure is only relevant for the processing at higher stages.

Using the same set of stimuli as in the electrophysiology, we performed a series of fMRI scans of the visual cortex of anaesthetized cats. In primary visual cortex, on average, gratings lead to the strongest responses. Similarly to the electrophysiology, no clear difference in the responses to the natural movies and to the stimulus with random higher order structure is found. Thus the fMRI measurements support the electrophysiological findings that large-scale neuronal activity in the primary visual areas is not sensitive to the higher order statistical structure of natural scenes.

Next, a detailed comparison is made between the responses in the blood oxygenation level dependent (BOLD) signal and the different measures of neuronal activity. Such a comparison depends on the detailed definition of the activity measure. For example, the match between BOLD and lfp depends on the lfp frequency band considered. In general, the BOLD signal matches the lfp best within the gamma frequency range. Thus it might well be that the BOLD signal is most sensitive to this frequency range often studied in conjunction with visual processing. The spiking activity, in contrast, shows a worse match to the BOLD signal. Overall the best match between BOLD and lfp is obtained when the comparison is based only on complex stimuli and the dissociation between spiking activity and lfp is strongest upon stimulation with natural stimuli. In conclusion, these results support the notion that the lfp corresponds more to the input to a cortical region than to its output.

Last, artificial neural networks are trained using natural stimuli as input. Using a criterion that requires a neuron’s activity to vary slowly over time, the neurons develop receptive field properties similar to those of complex cells in V1. Furthermore, using combinations of objective functions, the network successfully segregates an initially homogenous population of neurons into two sub-populations with distinct and complementary feature selectivities. We extend the learning scheme so that the network can learn parameters controlling the neurons nonlinear transfer functions simultaneously with the linear receptive fields. This result is important for creating networks than can model less well-understood systems and require only little a priori knowledge. In conclusion, there are several computational principles connecting the receptive fields of cortical neurons to properties of every day stimuli and the optimal receptive fields of the model neurons share many properties of real cortical neurons.
Zusammenfassung


neuronaler Aktivität wirft die Frage auf, welche Resultate sich über verschiedene Typen von Reizmustern verallgemeinern lassen.

In den gleichen Experimenten vergleichen wir die Antworten für natürliche Videos mit denen für ein Reizmuster mit gleicher Statistik zweiter Ordnung, aber zufälliger Statistik höherer Ordnung. Sowohl der Frequenzbereich des LFP, der durch diese Reize aktiviert wird, als auch die mittlere Stärke der Aktivierung sind für beide Reize identisch. Ebenso zeigt sich auch in den Feuerraten nur ein kleiner Unterschied. Beide Signale sind also nicht sensitiv für die Struktur höherer Ordnung. Entweder wird diese Struktur nur auf der Ebene einzelner Zellen extrahiert und diese Selektivität geht beim Mitteln über eine Zell-Population verloren, oder diese Struktur wird nur in höheren kortikalen Arealen analysiert.


I. Introduction

During every day life our brain continuously processes information about sensory stimuli originating from different modalities. Prominent examples are visual, auditory, olfactory, somatosensory and taste stimuli. Usually many of these sensory systems are active at the same time and often activated by a common source. Most of these stimuli arise from ‘natural’ interaction with our environment. For example we touch an object that we are using, we look somewhere to search for food or we hear somebody calling us. The brain continuously processes these impinging stimuli, interprets them and forms a coherent percept of our environment. The question, how sensory stimuli are transformed into electrical impulses in the brain, how these are transmitted between different areas and how this information finally reaches consciousness is a central part of today’s and past neuroscience research.

To investigate the properties of sensory transduction and processing of sensory signals in cortical areas, many experiments have been carried out. In order to perform an experiment, one has to decide on a particular protocol: What type of stimulus should be used, what signals should be recorded and from which preparation.

Sensory processing has been studied using a large variety of stimuli. It quickly became clear that the choice of the stimulus influences the results discovered and for some sensory modalities a particular stimulus advanced to become ‘the’ standard stimulus for this system. As an example, studies in the primary visual cortex were dominated for many years by sine-wave gratings. In general, the choice of the stimulus was usually guided by the effort necessary to create the stimulus and by the properties of the neuronal responses elicited. For example, some stimuli were frequently used because they elicit vigorous responses and thus lower the demand on the signal to noise ratio of the recorded signals. However, for a long time little attention was paid to the relation between the type of stimulus used and the question how the results obtained using a particular stimulus can be generalized. Especially, the question which stimuli should be used to ultimately understand the processing under every day conditions has received little attention.

There exist a large number of signals related to neuronal function that can be measured using different methods. The most prominent example is the spiking activity of single or small populations of neurons. This signal has dominated studies on sensory processing for many years and its relation to neuronal activity is well defined. Another measure that can be recorded using microelectrodes is the local field potential (Lfp). The local field potential is well suited to study temporal properties of the activity and thus is often used in studies investigating synchrony of neuronal firing. In contrast to the spiking activity, the Lfp is not only determined by the local spiking activity, but it also reflects sub-threshold processes. However, the origin of the Lfp is still debated. A third measure of neuronal activity, which became very popular in the past years, is the hemodynamic response measured using
functional magnetic resonance imaging (fMRI). This signal can be measured non-invasively and thus is one of the few methods that can be used with human subjects; hence its popularity. However, in contrast to the other measures of neuronal activity, the response properties of the fMRI signal are less well characterized. Furthermore, are its origin and its relation to the other two measures of neuronal activity unclear. Thus several measures of neuronal activity exist and these capture different aspects of the underlying processes. However, the relation between these different activity measures is still debated and it is unclear which results can be generalized across different methods.

The last point mentioned above is the choice of the experimental preparation: What animal should be used and how the animal is prepared to record the neuronal responses. Signals measuring neuronal activity related to sensory processing can be obtained from a wide variety of species, ranging from drosophila to humans. Usually the species is chosen depending on the sensory system of interest. Furthermore, signals can be obtained from anaesthetized animals or from awake behaving subjects performing complex tasks. Most studies investigating passive sensory processing record form anaesthetized subjects because the anaesthesia eases the technical demands and prolongs the time available for data acquisition. However, the effect of the anaesthesia on the recorded signals is not well understood and it is not clear how the results obtained from anaesthetized preparation generalize to the awake animal.

I.1 The choice of the stimulus

Today’s research uses a set of well-chosen stimuli that are adapted to the details of the question at hand and to the sensory system under investigation. The prominent type of stimulus used, however, changed considerably over the years. For example, as vision research became popular in the fifties, people used small spots of light to stimulate cells in the retina or in the cortex (Barlow 1953; Barlow and Kuffler 1957; Kuffler 1953; Daniel & Whitteridge 1959). Shortly later, Hubel & Wiesel discovered that small bars or edges are much better stimuli for cells in the primary visual cortex (Hubel & Wiesel 1959, 1962). Because of their oriented structure, these drive cortical cells much better than non-oriented spots of light. Using such simple stimuli a large number of results and phenomena was discovered and a rudimentary understanding of visual processing was obtained (Kuffler 1953; Hubel & Wiesel 1959, 1962, 1969; Cleland & Enroth-Cugell 1968). In the 70’s it was observed that cells in the primary visual cortex can be described as linear filters extracting the frequency content of the stimulus (Campbell & Maffei 1974). Because linear filters can be well described in frequency space, stimuli that are very simply in this domain were introduced: drifting sine-wave gratings. Using this stimulus, a great deal of work over the last decades was devoted to characterizing the filter properties of cortical cells in different areas along the visual hierarchy.
(Glezer & Ivanoff 1973; Maffei & Fiorentini 1973; Schiller et al. 1976; Movshon et al. 1978).

All these stimuli have several properties in common. First, they are simple, easy to generate and mathematically well defined. These are critical points, given that the presentation of a stimulus was a technically demanding problem before the advent of modern computer controlled systems. Furthermore, several of these stimuli were adapted to evoke strong responses from the system investigated. However, there is one problem associated with them: they differ strongly from the typical stimuli encountered during everyday life. The typical stimuli used in vision experiments are much simpler than the complex scenes we encounter during everyday vision.

While from today’s perspective, these frequently used stimuli seem very artificial, is the idea old that the stimuli used in experiments should share several properties with those processed on a daily basis. Barlow, for example, associated stimuli with behavioural properties and suggested that the experimenter needs to know what behaviour a chosen stimulus elicits (Barlow 1961a). Around the same time, Lettvin and colleagues (Lettvin et al. 1959) used photographs of the frog’s natural environment to stimulate its retinal cells. Following the same lines, Creutzfeldt and Nothdurft (Creutzfeldt & Nothdurft 1978) used natural photographs to investigate response properties of cortical cells. While these studies paved the way towards the use of natural stimuli, most other researchers relied, and many still do, on simple and artificial stimuli.

However, in recent years the notion that natural stimuli are an important tool for the understanding of visual processing became increasingly popular (Simoncelli & Olshausen 2002; Simoncelli 2003; Touryan & Dan 2001). Based on the observation that naturally occurring sensory stimuli are much more complex than those used in experiments, people begun to roughly classify sensory stimuli into two groups: artificial and natural. It is not so difficult to find a formulation for the intuitive idea behind this classification, e.g. ‘a natural stimulus is a stimulus that could be encountered during every day life a long time ago (lets say 10'000 year)’. However, when we search for a quantitative criterion for such discrimination, we encounter severe problems.

---

1 In a similar way, many studies in the auditory domain use pure tones consisting of a single frequency as stimulus (Evans et al. 1965). Studies on the somatosensory cortex often use point like all or nothing stimulation on the skin (Mountcastle & Powell 1959) or all or nothing movements of sensory organs like whiskers (Simons & Woolsey 1979).
I.2 What are natural stimuli?

The problem of classifying stimuli as natural or not turns out to be quite intriguing. One way to address this issue is to systematically study the properties of those stimuli that we think are natural. However, before doing so we have to create a database of natural stimuli. In the case of the visual system it is generally accepted that if we take a camera and capture (random) pictures in an outdoor environment, we obtain a good approximation to a natural stimulus. The use of computers and modern equipment like digital cameras eases this process and allows us to acquire a larger amount of such material. Motivated by signal processing techniques, researchers begun to investigate the statistical properties of such natural stimuli. Given the huge number of parameters involved, it turned out that a real classification of natural stimuli is difficult, if not impossible at present. However, a larger number of properties of natural stimuli were discovered.

Stored on a computer, images are composed by individual pixels, which thus offer a straightforward starting point for analysis. A quantity that can be computed efficiently and is easy to interpret is the second order correlation of the pixel intensities. Assuming that natural scenes are translation invariant in space and time, the second order statistics is represented by the power spectrum of an image or movie. It turned out that the power spectra of natural visual stimuli have a particular shape (Rudermann & Bialek 1994, Dong & Atick 1995). This particular statistical property distinguishes natural movies already from a very large class of artificial stimuli that have been used in experiments.

While the second order statistics of natural scenes gives us some idea about their structure, most of the information contained in an image is embodied in the higher order structure. In contrast to the second order correlations, the correlations of higher order are difficult to compute given the usually limited database and are also difficult to interpret and visualize. While some progress was made on their characterization (Thomson 2001a, 2001b), we still have a limited knowledge about the higher order structure of natural scenes and thus also about what distinguishes them from randomly created images.

I.3 Sensory systems and natural stimuli

Based on the observation that our sensory systems perform very successful during every day use it was suggested that these should be adapted to the statistical properties of natural stimuli. Studying the detection of motion by the frog’s retinal ganglion cells, Barlow suggested that evolutionary processes had resulted in a match of detector and stimulus (Barlow 1961). Because of this match between neuronal properties and the typical stimulus seen by these, the nervous system processes these stimuli very efficiently (Barlow 1989). This line of thought was later strengthened by the observations that both neurons in the lateral geniculate and in the visual cortex have filter properties that are optimal with respect to the statistical properties of natural scenes (Atick & Redlich 1990; Field 1994; Simoncelli 2003). Considering the huge number of possible stimuli that could reach our sensory systems it is
surprising that these systems operate so fast and efficient. Trying to build such a system our self it seems reasonable to include a priori knowledge about the typical stimuli to be processed. In the same way it seems reasonable that our sensory systems adapted over evolutionary timescales to the typical every day stimulus.

What are the consequences of such an adaptation? The processing of natural stimuli should be faster, more reliable and more efficient than the processing of artificial stimuli (Barlow 1989). Furthermore, the properties of single processing stages should reflect the corresponding properties of the natural stimuli. Thus there are two lines of work relating natural stimuli and properties of the brain (Simoncelli 2003): First, it should be possible to experimentally measure differences in processing of natural and artificial stimuli. Second, theoretical considerations starting from the image statistics should give predictions about properties of the brain. In the next sections these two lines of work are presented in more detail for the visual system.

I.4 Experimental aspects of an adaptation to natural stimuli

Support for an adaptation of the visual system to the properties of natural scenes comes from the following observations. Single neurons in the primary visual system transmit information about large patches of natural scenes more efficiently than about small patches (Vinje & Gallant 2000, 2002). Neurons in the thalamus reduce the redundancy in their output to natural scenes but not to noise movies and thus perform more efficiently under natural conditions (Dan et al. 1996). Responses of neurons in the fly visual system are less ambiguous with respect to behaviourally relevant parameters when probed with natural stimuli (Kern et al. 2001). Psychophysical experiments show that search tasks are performed better in the context of natural images than on images whose statistics was manipulated (Tolhurst & Tadmor 2000). Furthermore, a larger number of analogies were found between interactions of higher order features in natural scenes (Kaschube et al. 2001; Krueger 1998), their interaction on the level of single neurons (Kapadia et al. 1995) and in perceptual studies (Polat & Sagi 1994; Kapadia et al. 1995). Last, the properties of the depth structure in natural images can be used to explain psychophysical anisotropy in the perception of length (Howe & Purves 2002). Concluding, there is strong evidence that the visual system processes natural scenes in a different way than other stimuli.

The results mentioned above strongly indicate that the brain is somehow adapted to the properties of real world stimuli. However, most of these results were obtained in a very special preparation or stimulus paradigm, e.g. by changing the size of a small stimulus patch. Overall, these results do not prove that the visual system is adapted to natural scenes at all, nor do they satisfactorily demonstrate that natural scenes are processed differently than other stimuli. As a consequence we do not know which results obtained using simple stimuli generalize to the processing of natural scenes and many simple but important questions remain open. For example, the question whether there are differences in the processing of
natural and artificial stimuli can be formulated in the following way: “Are the properties of V1 responses to natural stimuli different from those to the classical stimuli, and if so, which properties”. A possible approach to address this is to take a set of natural stimuli and a set of artificial stimuli and to compare the responses in different measures of neuronal activity. As seen in the introduction, there is a large variety of measures that can be used to study neuronal responses. However, it is unclear which of these measures is the most suited to answer this question. As a consequence, several of these should be investigated. Another question that can be formulated is related to the definition of a natural scene: “What is the influence of the higher order structure of the stimulus on the cortical activity”. In contrast to the above question, avoids the latter the differentiation between natural and artificial stimuli, which as we saw above is problematic. Instead, it modifies a property, which is one of the properties that can be used to distinguish natural from other stimuli. Such an influence of the higher order stimulus structure could be tested by manipulating the spatio-temporal properties of natural movies and comparing the responses.

1.5 Theoretical aspects of an adaptation to natural stimuli

Theoretical considerations can be used to link properties of neurons in sensory areas to the properties of natural stimuli. For example, based on an assumption of efficient coding of natural scenes, the shape of an optimal activity distribution was inferred (Barlow 1961; Field 1994; Barlow 2001). From this the concept of sparse coding emerged and it was shown that simple cells in primary visual cortex actually define a sparse code of natural scenes. First, filters with similar properties as simple cells receptive fields do have a sparse activity distribution when applied to natural scenes (Field 1994). Second, starting with a model neuron having an unspecific random receptive field, training this neuron to have an optimally sparse response yields a simple cell like receptive field (Olshausen & Field 1996). Thus the principle of efficient coding or redundancy reduction defines a link between image statistics and properties of neurons in V1. Similar relations were found between the stimulus statistics and properties of neurons in the LGN (Dong & Atick 1995), retina (Atick & Redlich 1990) and also in other sensory systems (Koerding et al. 2002, Lewicki 2002). Thus there exist a number of optimal coding criteria that relate the properties of individual neurons to the stimulus statistics and successfully form a theoretical basis for the understanding of the adaptation of sensory systems to the properties of the real world.

While this approach was very successful at the lower processing stages like the retina, the LGN and simple cells in V1, it still is open how to relate the receptive fields of neurons at higher stages to the properties of natural scenes. For example, already the properties of complex cells in the primary visual cortex cannot be inferred from the properties of natural scenes under the assumption of sparse code. It is likely that other computational principles than efficient coding are necessary in order to explain the response properties of this type of neurons. While such other principles have been proposed in the past, they have not been evaluated in the context of natural scenes. Furthermore, many of the model networks that have been optimised
based on their responses to natural scenes capture the properties of only one type of neuron; in biology, however, at a given processing stage neurons with different properties co-exist. This raises the question how an initially homogeneous population of neurons segregates into groups with distinct and complementary response properties. To achieve such segregation, a mechanism acting like a lateral interaction in the brain is required. Last, most studies address only the properties of the linear receptive field of the model neuron while other parameters are kept fixed. Especially the transfer function of the neuron, describing a possible nonlinear relation between the intrinsic state and the output of the neuron, is kept fixed. In the brain, however, neurons show strongly nonlinear responses. To describe such unknown nonlinear systems we need networks that adaptively change parameters describing their nonlinearity simultaneously with their linear receptive fields.

I.6 The next chapters

There are two lines of research in this thesis, similar to the distinction between experimental and theoretical consequences as above. The first measures neuronal responses in the visual system. Addressing the questions raised above, we use a larger set of stimuli including classical gratings, natural movies and stimuli with altered global structure. This allows quantifying differences between natural and classical stimuli and we can directly address the impact of global stimulus structure on the response properties. Following the discussion of the choices for the experimental protocol above, we record different measures of neuronal activity, local field potentials and spiking activity in the visual cortex of awake animals (Chapter 3). Because these measures of neuronal activity are very localized, we performed high-field fMRI experiments on anaesthetized animals measuring BOLD responses in different visual areas (Chapter 4). There we compare the responses for the different stimuli in this more global measure of neuronal activity. Furthermore, this allows us to compare the neuronal responses measured using microelectrodes to the hemodynamic responses in detail.

The second line of research investigates statistical properties of natural movies and uses these stimuli to train models of visual cortical cells. A set of natural movies was recorded using a camera mounted to a cats’ head and the space-time statistical properties are analysed (Chapter 2). Then networks, either implementing a sparse or stable code, are trained on these natural stimuli (Chapter 5). To address the different points raised above, we investigate how the response properties of complex cells can be learned in such networks using different optimality criteria. We furthermore study the emergence of distinct response properties in an initially homogenous network. Last, we propose a learning scheme in which not only linear receptive fields are optimised but the nonlinearity of the neurons transfer function is adapted as well.
II. Properties of natural scenes

When we move through the world, movements of our body, head and eyes change the direction of gaze and thus impose a temporal structure on the visual input. Different species of animals have a different pattern of body and eye movements. Furthermore, different species live in different environments, which do not necessarily have the same typical spatial structure. As a result, differs the typical every day visual input between species. However, most studies investigating statistical properties or cortical processing of natural scenes do not take this into account. This problem is most prominent for the temporal structure, because the same spatial scene can be projected on the retina with various motion patterns. Actually, many studies work with still images only, or create time varying stimuli by artificially imposing a temporal dimension onto still images. While we have a great deal of knowledge about the spatial properties of natural scenes, their temporal properties are only little explored. In the following we address this point and study the statistical properties of natural time varying images that were recorded from a camera mounted to a cats’ head.

II.1 Recording movies from a camera mounted to a cats’ head

We used the following approach to record natural movies from the perspective of an animal, which in this case is a cat. A small CCD camera was mounted to the head of freely moving animal and short movie sequences were recorded while the animal was exploring different outdoor environments. These environments include forests, meadows and parts of the university campus. Since the camera was mounted to the head of the cat, the recorded movies incorporate all the body movements of the animal.

There is one problem associated with this procedure to record movies from the perspective of a cat. While these include the body and head movements of the animal, the movies do not include the eye movements. This point requires discussion. It is known that human observers looking at natural scenes make roughly 3 saccades per second and in the mean time fixate well on a single point (Land 1992). Cats, in contrast seem to make much fewer saccades. Under laboratory conditions, head-fixed animals watching natural movie clips make a saccade roughly every 3 seconds (Moeller et al. 2003). Similar saccade frequencies were reported in other studies (Evinger & Fuchs, 1978). One might argue that this low number of saccades and thus this difference between humans and cats is related to the fact that the animal is head fixed. However, experiments with human subjects in the same eye-tracking setup where the cat experiments were performed yielded a similar high saccade frequency than measurements of head-free humans (Moeller et al. 2003). Measurements of eye movements of a freely moving animal are technically difficult. A recent experiment made a step into this direction. Using a second CCD camera the animal’s right eye was monitored while animal was freely exploring a local park. From this video of the animal’s eye, eye movements were extracted. The frequency of saccades was similar to a head-fixed animal (G. Moeller, personal
communication). Thus neglecting the eye movements while acquiring the video sequences still gives a good approximation of the input to the eye.

Figure II.1. Example frames taken from different natural movies.

II.2 The second order statistics

The natural movies consist of a series of single images that are discretized into pixels. Studying the statistical properties of the intensity of these pixels is a straightforward analysis and is the regime in which natural movies are best explored. It was shown that the spatial correlations of pixel intensities have a particular pattern. This can be best seen by computing the power of the Fourier spectrum which decays as \(1/f^2\) as a function of spatial frequency (Ruderman & Bialek 1994, Ruderman 1994, Mumford & Huang 1999). This was taken as an indication of the self-similar structure of natural movies (Ruderman 1997). Furthermore, the power is not equally distributed across the different orientations, but horizontal and vertical are more prominent than the oblique orientations (Ruderman 1994).

In a similar way the temporal structure of pixel intensities can be studied. The power of the temporal frequencies decays as a function of the frequency roughly as \(1/w^2\) (Dong & Atick, 1995, Dong 1997). However, this approximation holds well only in the regime of high temporal and low spatial frequencies. In general, spatial and temporal structures are dependent and the power spectrum cannot be factorised into spatial and temporal components. The source for this dependence of spatial and temporal components was found in the motion of objects of different size at different depths in space (Dong & Atick 1995).

Using our natural movie database we confirmed these results. The spatial power in our natural movies decays as \(1/f^2\) and the interaction between spatial and temporal frequencies follow roughly the results described by Dong & Atick (1995). In contrast to their results, however, the temporal correlations in our natural movies seems to be weaker and thus the power spectrum is more ‘white’. This might be attributed to the difference in movies analysed. Dong
& Atick used commercial Hollywood movies that, at least subjectively, have a different temporal structure than the cat-cam movies in our database.

II.3 The motion in natural movies

The temporal structure of our natural movies recorded using a camera mounted to a cats’ head is to a large part determined by the movements of the animal. There are two reasons why it is interesting to know what this typical motion is. First, knowledge about this structure allows constructing stimuli having an artificial spatial structure but a ‘natural’ temporal structure. In this way, the influence of the motion independent of the spatial properties of the stimulus can be investigated. Such stimuli will be used in chapter 3. Second, in the visual hierarchy there are areas selective for motion. In the cat this is area PMLS. We can compare the typical motion pattern to properties of the neurons in this area.

The motion of a natural movie clip can in principle be extracted from the spatio-temporal power spectra. In order to have a direct measure of this motion, we used a different approach and quantified the local optic flow in different movie clips. Following standard methods (Baeuchemin & Barron 1995) we estimated the local optical flow on a grid of points covering the whole image frame. A patch of 30 by 30 pixels centred on the grid point was compared to patches at different positions in a range of 70 pixels in each direction in the next video frame. The comparison was based on the mean square difference after removal of the overall mean luminance of each patch. The best match defined the local optical flow. The global motion vector of the video was computed from this locally defined flow field as the arithmetic average.

![Figure II.2](image)

*Figure II.2.* Properties of the flow fields in a 60 sec movie clip. Left: average amplitude of the flow field in [deg/sec]. Middle: average direction of the local flow. Right: distribution of amplitudes.

Looking at the average flow field for different movie sequences, it turned out that the dominant contribution to the motion comes from the forward walking of the animal: The direction of flow is directed radially away from the centre of the image. While the dominance of the forward walking could have been expected beforehand, the magnitude of the flow field
can only be measured. With an average of 43 deg/sec the amplitude of the motion is pretty large - indicating it’s importance.

The main contribution to the average flow field in our natural movies comes from the forward walking of the animal. This by itself is not very surprising, but can be compared to known receptive field properties of neurons in the motion selective areas PMLS of the cat. Electrophysiological studies revealed that there are two populations of cells in PMLS. One group is strongly activated by radially expanding patterns that have their origin near the area centralis. Another group of cells prefers flow fields that have their origin in the centre of the receptive field of the respective neuron. Thus using the responses from these two groups of neurons the brain can identify the animal’s own body motion and segment relative motion from exterior sources (Merabet et al. 2000; Brosseau-Lachaine et al. 2001).

![Figure II.3. Time course of the flow field. The motion in the natural movies follows a very irregular pattern with many sharp peaks.](image)

Looking at the temporal trace of the flow field reveals a strong variation over time. The flow field is characterized by phases of slow motion, which are separated by short intervals of rapid image translations. Overall the temporal structure of the flow field is highly variable as indicated by an average coefficient of variation around one (1.06+-0.09). This result highlights the importance of studying the temporal structure of natural scenes. Only form static images, it is hard to imagine that the typical temporal structure shows such irregularity. This is an important result with respect to the design of stimuli for the experimental projects. We will see further below that this temporal structure has a strong impact on the activity patterns found in the primary visual cortex.

### II.4 Statistics of oriented contours

The above analysis studies the properties of natural movies based on the luminance of individual pixels. However, these pixel values are not a property of the real word and dependent on the parameters of the discretization. More relevant to the response properties of neurons in the primary visual cortex are higher order features such as edges, contours and
objects. These features are composed of a large number of pixels and thus not dependent on the details of the discretization. Furthermore, because they are directly related to the response properties of visual cortical cells, knowledge about the statistical properties of these features in natural scenes allows a direct comparison of image statistics and properties of the biological hardware.

One prominent feature that is extracted by cells in the primary visual cortex is the orientation of a local image patch (Hubel & Wiesel 1959): The responses of most cells in V1 are modulated by the orientation of a small bar presented within the receptive field. In a similar way, given a digital image, the information about local orientation can be extracted by convolving the image with Gabor filters. From the amplitudes of such filters one can then quantify their statistical properties and correlations.

Studying the properties of pixel luminance in the Fourier domain above, we saw that the cardinal orientations are more prominent than the oblique orientations. This is also visible in the average amplitudes of the different filters.

![Figure II.4. Distribution of orientations in natural scenes](image)

From the amplitudes of different orientations at all positions we computed the pattern of spatial correlations. Several results need to be highlighted here. First, as can be expected, the correlations decay with spatial separation but are still significant after several degrees of visual angle. For the cardinal orientations, they persist over longer distances than for the oblique. Second, on average the correlations of one orientation with itself are stronger than across orientations. Third, the spatial pattern of the correlations is not isotropic. The iso-orientation correlations are extended along the direction determined by the orientation of the respective filter. For example, the correlations of vertical orientations across space are most prominent along the vertical axis and decay faster along the horizontal axis. Thus the pattern of cross-orientation correlations favours collinear arrangements of oriented contours over parallel contours. Overall these results agree well with what other studies investigating the correlations of orientations found (Kaschube et al. 2001; Krueger 1998). However, other studies using randomly collected still images (usually from photo CD’s). These images
contain a large number of man-made objects and thus induce an anthropocentric bias. While the results form our cat-cam database agree qualitatively well with the other studies, we find that the correlations extend over much larger distances than reported previously (several degrees compared to less than one degree).

**Figure II.5.** Correlations of orientations across space. Left panel: On a grid of points the amplitude of different orientations is computed. The prominent orientation of one point is chosen as reference (pink) and the amplitude of this orientation is correlated with the amplitudes of a set of other orientations (e.g. green) at neighbouring points. Right panel: Results for all combinations of orientations.

**Figure II.6.** Left: Parallel and Collinear arranged bars. Right: The iso-orientation correlations are dominated by collinear arrangements. The surfaces of constant correlation are extended into the direction of the orientation (here 45° oblique).

So far, we confirmed results previously reported by other investigators on our large database of natural movies. However, because we can include the temporal domain in our analysis, we can now go one step further and study whether and how the spatial correlation pattern changes over time. We found that a given orientation at a fixed position shows correlations of its amplitude over several hundreds of milliseconds. This holds for all different orientations, but is again most prominent for the cardinal orientations. From this result we can deduce a
time-scale, during which the spatial properties of the stimulus do not change significantly. In our database this is on average 100-300 ms, depending on the orientation. Further, we computed the temporal correlations across orientations and spatial positions. The spatial pattern of the correlations computed at non-zero time lags is very similar to the instantaneous correlations. As a result, the spatial pattern of correlations, especially the prevalence of collinear over parallel contours, is preserved over time with the time scale reported above. Thus, the spatial pattern does not change qualitatively, instead all correlations decay over time with a similar decay constant.

These results allow interesting comparisons of the image statistics with anatomical and electrophysiological properties of the primary visual cortex. First, a prevalence of cardinal orientation is found in the brain as well. More cells in V1 are tuned to the cardinal orientations than to the oblique orientations (Kalia & Whitteridge, 1973; Poggio & Fischer 1977; De Valois et al. 1982; Li et al. 2003) and visually evoked potentials measured in the cat are stronger for the cardinal orientations (Bonds 1982). A similar prevalence of cardinal orientations was also found in optical imaging studies (Sengpiel et al. 1999). Our recordings of local field potentials in the awake cat similarly show that most recording sites prefer cardinal orientations; most of the orientation tuning curves constructed from the local field potential peak at horizontal or vertical orientations. The spatial pattern of correlations can be compared with the pattern of intra-cortical connectivity. The long-range connections between orientation tuned cells in V1 preferentially connect cells with the same orientation preference (Bosking et al. 1997, Gilbert & Wiesel 1989). The figure below shows the distribution of boutons from long-range connections (> 500 um) with respect to the difference in preferred orientation. Overlaid is the cross correlation of the amplitudes of the different orientations at a fixed spatial distance. The two curves match pretty well, indicating a close match between image statistics and intra-cortical connectivity.

The pattern of correlations shows a preference for collinear arrangements than for parallel. A similar preference was also found in the V1 connectivity. Bosking et al. (1997) aligned the pattern of synaptic connections with the spatial arrangement of the cells in the visual field. It turned out that cells not only connect to cells of similar orientation preference, but also that these cells are aligned collinearly in the visual field.
Figure II.7. Comparison of results from image statistics and anatomy. The grey curves indicate the distributions of boutons (at least 500 µm from the soma) relative to the orientation preference of the neuron (defined as zero). The black curve represents the mean across all cells. Most boutons are formed with neurons of similar orientation preference (from Bosking et al., 1997). The red curve displays the strength of cross-orientation correlations (averaged across different spatial distances). The red curve was scaled to match the height of the black curve.

This fact that collinear arrangements are more prominent than parallel was also described in psychophysics. A prominent example is the Gestalt laws, rules that describe how we group objects into a coherent image (Kofka 1935, Wertheimer 1938). One of these laws, the law of good continuation, suggests that collinear arrangements are more likely to be grouped together and thus seen as one object than parallel arrangements. Experiments evidence for such a preference or dominance of collinearity comes from measurements of detection thresholds. These were lower for collinear arrangements of oriented segments than for other configurations (Polat & Sagi 1994, Kapadia et al. 1995).

The firing rate of single neurons in V1 is determined primarily by the local orientation within the receptive field. However, it is known that stimuli outside the classical receptive field can alter the responses to stimuli within the classical receptive field (Maffei & Fiorentini 1976). Kapadia et al. (1995) investigated the effect of an oriented bar outside the classical receptive field on the response to an oriented bar within the receptive field. It turned out that these interactions follow a similar pattern than found in the measurements of detection thresholds. Collinear arrangements of the two bars lead to the strongest facilitation of the response, while two parallel bars did not lead to facilitation of the response. Thus not only the rules for perception, but also the anatomical and physiological substrate in the primary visual cortex shows the same pattern of feature interaction as found in natural scenes.

In the temporal domain we find long lasting correlations of orientations to extend several hundreds of milliseconds preserving their spatial structure, i.e. the collinearity. These persist sufficiently long to allow bottom up and long range lateral input to be coactive and driven by the same orientated structure. This result is of importance assuming that the anatomical connectivity is established using learning schemes that rely on temporal associations, like in a Hebbian learning rule (Hebb 1949). Assume an oriented structure excites a neuron and we wish to strengthen a long-range connection with a neuron preferring a similar orientation. The activity of the first neuron will travel to the second neuron and the speeds of lateral connections in V1 can be quite slow (Bringuier et al. 1999). In order to strengthen the connection, the second neuron must be activated by a similar stimulus as the first neuron. Now, the results from the image statistics tell us that such an orientation is likely to occur, also after several hundreds of milliseconds. Therefore the spatio temporal interactions of orientations seem to fully cover the range of tangential connections and provide a substrate...
that could guide the development of orientation maps and long-range connections in primary visual cortex.

**II.5 Summary of the results**

We analysed the statistical properties of a large database of natural scenes recorded from a camera mounted to a cat’s head. These movies closely resemble the input to the cats’ eye and thus incorporate the rich temporal structure determined by the motion of the animal. The analysis concentrated on two parameters: the pixel-intensity and the amplitudes of oriented filters.

Studying the properties of pixel intensities, we confirmed previous results demonstrating that the Frequency spectra of natural scenes have a particular shape. The power decays with increasing frequency following a $1/f$ power-law, both in the spatial and in the temporal domain. Given that the coupling of spatial and temporal dimensions is determined by the motion in the movie, we computed the flow fields describing this motion on a frame-by-frame basis. Analysis of these revealed that the motion is on average quite strong (43 deg/sec), that the distribution of amplitudes has a long tail extending to high speeds and that this amplitude varies considerably on short time scales (CV>1). These results show that the typical stimuli during every day vision have a complex temporal structure whose impact is largely unexplored and thus should be integrated both in experimental and theoretical work.

The amplitudes of the oriented filters are correlated across space and across different orientations. We confirmed previous results showing that the spatial pattern of the correlations has a particular shape favouring collinear compared to parallel arrangements. Extending this analysis into the temporal domain, we showed that the correlations of an orientation with itself persist for several hundreds of milliseconds. Furthermore, the pattern of cross-orientation correlations is preserved over time. Thus, the dominance of collinear contours is not only an instantaneous phenomenon but prevails over significant time spans.

Overall, we demonstrate a good match between these results and known properties of the anatomical connectivity in V1 and physiological and psychophysical cross-orientation interactions.

**II.6 Directions of future work & conclusion**

Orientation is not the only feature extracted by cortical cells from the visual input. Other features include disparity (Hubel & Wiesel 1962; Ohzawa et al. 1996, 1997; Cumming & DeAngelis 2001), motion (Zeki, 1974; Maunsell & Van Essen, 1983a; Albright, 1993) and colour (Zeki 1980; Lueck et al. 1989). Cells at higher stages are sensitive to complex shapes and objects (Tanaka et al. 1991; Kobatake & Tanaka 1994). Similar to orientation above, the statistical properties of these features in natural scenes should be related to properties of the corresponding processing stages in the brain. For example, the extent of the correlations of these features across space and time should give and indication of the receptive field size of the neurons extracting these features. Other properties of the correlations should have
counterparts in the anatomical wiring or in the response properties of the cells. Furthermore, interactions between different features in the input, like colour and orientation, could give an indication of how different feature maps are related to each other or how different features interaction in psychophysical experiments.

It seems promising to extend such a comparison of stimulus statistics and properties of the sensory systems to other sensory modalities. A large variety of receptive field types are known in different cortical areas and these define a larger number of stimulus features could be extracted from databases of natural stimuli. Compared to the visual system, especially V1, the response properties of neurons in other cortical areas are less well understood. Assuming that other sensory cortices are build in analogy to the primary visual area, predictions made from the analysis of natural stimuli could actually guide experiments and make concrete predictions. These predictions do not need to be restricted to physiology or anatomy but could be used to predict psychophysical effects as well.

As mentioned in the introduction, our different sensory systems are stimulated continuously and often by stimuli arising from the same source. Thus, some events cause multi-modal stimulation. Conversely, at higher stages of the cortex cells are not only stimulated by the input from one sensory modality but integrate across modalities. For example cells in the posterior parietal cortex, the superior temporal sulcus, the in the parahippocampal region and the superior colliculus receive input from different sensory cortices (Welch & Warren 1986; Kandel et al. 1996; Calvert et al 2001). Such cross-modal interaction can reduce reaction times (Miller et al. 1982; Frens et al. 1995) or our ability to locate and detect a stimulus (Welch & Warren 1986; Stein & Meredith 1993). Analysis of multi-modal natural stimuli should given an insight into the common features that often give rise to such multi-modal stimulation. Investigating the statistical properties of these features then should allow making predictions about the properties of neurons that are sensitive to multi-modal stimuli. After all we have very limited knowledge about these effects and we simply now that multi-modal interaction takes place. The details however are unclear. To investigate these phenomena further, theoretical predictions would ease the design of experiments.

The models and optimality criteria for receptive fields of visual cortical neurons discussed in chapter 5 relate the statistical properties of the visual input to the properties of the neurons. Here, a close link between receptive field development and feature statistics is given. Our simulation results for example show that non-oriented colour cells tend to be more ‘stable’ than oriented cells. This suggests that colour is correlated across longer intervals than orientation. Thus studying the statistical properties of multiple features also gives an insight into how neurons selective to these features emerge in a model of V1 learning from the statistical properties of natural scenes.
II.7 Publications

The results about the second order statistics of the pixel intensities and the flow fields were published only in abstract form:

- **The natural visual environment of cats.** Kayser C, Betsch B, Einhäuser W, Körding KP, König P. *Annual Meeting of the German Physiological Society, 2002*


The results about the statistics of orientations were published in the following paper:

III. Neuronal responses to complex stimuli in the primary visual cortex of awake cats

In the introduction we saw that the visual system is adapted to the statistical structure of natural stimuli and that this adaptation affects the response properties of neurons in V1. However, the details of this adaptation and many questions relating to it are unresolved. For example, we do not have the answers for the following two simple questions: “Is the average cortical response to a classical grating and a natural movie different in any respect” and “how does the global structure of a stimulus influence the response properties”. To address these questions we recorded a number of different signals related to neuronal activity in the primary visual cortex of the cat. We used several manipulations of natural movies with altered spatio-temporal properties in order to study the effect of the higher order statistics. Before going into the details of the results, the reason for the choice of the used preparation and stimuli will be given.

III.1 Recording population activities

A difference in the activity pattern for natural and artificial stimuli could reside on various levels of neuronal activity. For example on the average number of spikes per neuron, the timing between individual spikes, the number of neurons activated or the relative spike timing between different neurons. In order not rely on a special assumption we decided to record several measures of neuronal activity that can be extracted from microelectrode recordings. Usually these signals are based on the activity in different frequency ranges. Spikes, for example, are events found in the regime of high frequencies, while slow somatic potentials are found at lower frequencies. From our recordings we extracted three different types of signals: Local field potentials, analog-Multi-unit activity (analog-Mua) and multi-unit spikes. Each of these measures characterizes a different aspect of the neuronal responses.

The local field potential is an aggregate measure of local electrical activity around the electrode tip (Freeman 1975). This signal is influenced by spikes and also by dendritic and somatic potentials (Mitzdorf 1985, 1987). Thus, it represents not only the output of brain region, as the spiking activity, but also the input to a local region (Logothetis, 2003). Because the Lfp is an aggregate measure of activity, it is influences by a large number of signals, and does can be used to detect overall differences in the mean activity level. Furthermore, the Lfp is a continuous signal and thus well suited to study temporal properties of the responses (Abeles 1982).

The analog-Mua is an aggregate measure of high frequency activity. Similar to the local field potential, it is influenced by all nearby neurons and by signals from different processes of a neuron. However, in contrast to the Lfp it is mostly dominated by local spiking activity. Although this measure of neuronal activity is less well explored than the Lfp and the spiking
activity, it is often used when a measure of spiking activity is desired but individual spikes cannot be recorded (e.g. Gail et al. 2000; Super et al. 2003).

Multi-unit spiking activity comprises the spikes of a small population of neurons around the electrode tip. In contrast to the analog-Mua, which is an unspecific measure of high frequency activity, the multi-unit spiking activity is influenced only by spikes.

![Signal extraction from microelectrode recordings](image)

**Figure III.1.** Signals extracted from the microelectrode recordings. Top: Spectrogram quantifying the power at different frequencies of the LFP. The power at each frequency was normalized to show the change compared to the blank before the stimulus. Middle: Analog-Mua. Bottom: PSTH constructed from multi-unit spikes. Both were normalized with respect to the blank.

### III.2 Recording in the visual system of alert cats

Classically, experiments on the visual cortex are performed in anaesthetized animals. This has the advantage that one can record for a long time from the same site with good stability. The anaesthesia, furthermore, eases the problem to control for eye movements. However, anaesthetics can influence the activity levels and properties of spontaneous activity in general (Erchova et al 2002) and also the properties of receptive fields (Roberston 1965; Lee 1970; Friedberg et al. 1999). Especially higher order effects like figure-ground segmentation seem to be influenced by the anaesthesia (Lamme et al. 1998). Furthermore, to ultimately understand the processing under natural conditions, i.e. how every day stimuli are processed
by the visual system of an awake behaving animal, one has to understand the activity patterns in the awake animal. Based on this, we decided to record from alert cats using chronically implanted electrodes. However, in order to ensure well-controlled stimulation, some restrictions were necessary. In order to limit the visual range, avoid recording artefacts from muscle activity and ease the measurement of eye movements, the animals were head fixed.

There are two types of electrodes and drives that are commonly used for recordings in awake animals. Either a recording chamber is implanted which allows inserting a new electrode at the beginning of each session. This allows making many different penetrations and allows good electrode quality for each recording. However, working with cats, the time available for recording in one session is rather short (below 40 minutes). Thus a design is needed in which the electrodes stay in the brain all the time. We developed a small microdrive that can be implanted on the skull and allows manually adjusting the position of four electrode bundles. Using this drive, we could sample a large number of recording sites at different depths along four penetrations in each animal.

**III.3 Set of stimuli**

In the introduction we formulated several general questions concerning the difference of activity patterns for classical lab stimuli and natural movies. To answer these questions we need to compare activity patterns for gratings to those for natural movies. Further, we would like to study the effect of global stimulus structure, e.g. higher order statistics or motion, on the cortical activity. To meet these requirements we used different groups of stimuli: classical gratings, natural movies, movies with altered global statistics and stimuli with altered motion.

The gratings were drifting sine- and square-wave gratings. These were also used to characterize classical tuning properties of the recorded signals, especially those of the local field potential.

The natural movies used were a subset of those used for the analysis if the image statistics described in chapter 2. In total we used three short movie clips, each recorded in a different environment.

We constructed a set of modified natural movies with altered global statistics based on the following reasoning. Natural scenes differ from classical lab stimuli in both their spatial as well as temporal properties. A uniformly drifting sine wave grating for example is characterized by a single spatial as well as a single temporal frequency. Time varying natural scenes on the other hand contain a wide range of spatial and temporal frequencies and the contrast of these can be computed from the amplitudes of the stimulus’ Fourier spectrum. As explained in chapter 2 natural scenes have a characteristic amplitude distribution of both, spatial and temporal frequencies. While the amplitudes of the Fourier spectrum characterize the second order statistics, the phases determine the alignment of different frequencies and
thus the higher order structure. Artificial images constructed from the amplitude spectrum of a natural image but with random phases of the different frequencies have a quite different appearance compared to the original image from which the amplitude spectrum was taken. Such images, known as pink noise, have a foggy like appearance lacking any visible object due to their random higher order statistics. We applied phase randomisations to natural movies in both the spatial and temporal domain simultaneously. This was done by computing the space-time Fourier transform over all movie frames and replacing the phase at each frequency by a random value between 0 and $2\pi$. The inverse Fourier transform was applied to obtain the new stimulus. In total, three pixel noise stimuli were used, each constructed from of the three natural movies. The obtained stimuli have the same spatio-temporal power spectrum and thus the same spatial and temporal frequency distribution but lack the higher order structure of the original natural movies.

The same principle, using an original stimulus and one with altered higher order structure, was also applied to natural movies filtered with Gabor wavelets. These stimuli, which are based on a reconstruction of a natural movie from a wavelet representation, have a reduced content of spatial frequencies. The corresponding manipulated movies have the property, that the alignment of different wavelets defining the stimuli is changed. In this way, the local contrast edges defined by the Gabor wavelets are left unchanged but their global alignment is randomised. The reason to use these stimuli was that individual cells in V1 should be activated strongly by these local contrast edges. Manipulating the alignment of these edges should reveal effects of the global correlations on the population response.

Figure III.2. Example frames from the different types of stimuli.

Simple stimulus properties like mean luminance and contrast can strongly influence the activity elicited. Thus, the different stimuli needed to be normalized with respect to these parameters. However, normalizing all stimuli to have the same mean luminance and r.m.s. contrast introduces one problem. While the overall contrast is the same, the contrast at
individual frequencies can be different. This is most prominent for the gratings that contain only one frequency. After normalization the contrast at this frequency will be much higher than the contrast of the same frequency in the other stimuli. Similarly, the wavelet-filtered stimuli do not contain very high and very low frequencies. Thus the power at intermediate frequency ranges after normalization is higher than for the natural movies. While this causes a problem for the comparison of the activity patterns for natural movies and wavelet noise, each movie can be directly compared to the respective noise.

### III.4 Basic properties of the local field potential

Although the LFP is an aggregate measure of neuronal activity, it shares many properties with the responses of single cells. In the primary visual cortex, it is well known that neurons are tuned to basic image features like orientation, spatial frequency and temporal frequency (Hubel & Wiesel 1962). Some studies investigated similar tuning properties of the LFP, most often orientation tuning, in anaesthetized (Gray & Singer 1989) and awake animals (Frien et al. 2000; Siegel & König 2003). However no coherent picture exists so far (see also chapter III.8). In our experiments we found that the LFP is tuned to all three features and shows similar specificity as the responses of individual neurons.

**Figure III.3.** Tuning curves constructed form the average power in the LFP between 30 and 100 Hz. Each tuning curve was normalized such that the response is expressed relative to the feature value with the minimal response (this was set to zero).

### III.5 Local field potential results for all stimuli

The above figure summarizes the LFP activity for the different stimuli. Clearly, the activity pattern for gratings differs from those for the other stimuli. The range of frequencies activated is restricted to a narrow band and the activity is uniform over time. In contrast to this are the activity patterns for the other stimuli very irregular and include a much larger range of frequencies. We analysed these two aspects, the frequency range activated and the temporal structure in more details.
Figure III.4. Spectrograms for the different types of stimuli averaged across all recording from one animal. Spectrograms were normalized with respect to the blank period by computing the standard deviation at each frequency during the blank and dividing the time-course of the respective frequency by this standard deviation. To better quantify the differences in frequency bands activated by the different stimuli, so called modulation curves were computed: The spectrogram was averaged over time in a window starting 200ms after stimulus onset lasting till stimulus offset. These modulation curves are shown on the right of each spectrogram.

Figure III.5. Modulation curves from one subject. Each modulation curve was divided by its mean. This corrects for overall differences in the activity between stimuli. As a result all modulation curves from the complex stimuli (grey curves) overlap and the difference to gratings (black) becomes obvious.

To analyse the ranges of frequencies activated, we computed modulation curves describing the relative activation of different frequencies for a given stimulus and subject. These reveal that, except for difference in mean activity, the natural movies and their manipulations activate different LFP frequencies with the same relative amplitude. Thus, except for an overall scaling of the mean activity, the relative amplitude of different LFP frequencies is the same for natural movies and all their modifications. For these complex stimuli, the stimulus causes increase in the power over a wide range of frequencies, extending well beyond 100 Hz. This
is in contrast to gratings, for which the activated frequency band is restricted to a more narrow range between 30 and 80 Hz. This corresponds to the classical gamma frequency range that often has been reported to be activated by visual processing.

The two main results following from these results are: First, the activity pattern elicited by gratings is markedly different from that for natural and complex noise stimuli. Second, the lfp power is only weakly sensitive to the higher order structure of the stimuli. The last conclusion is based on the finding that neither the activity pattern for natural movies differs from that for the pixel noise, nor do the two wavelet-manipulated stimuli lead to different lfp activations.

Looking at the temporal activity profile for the different stimuli, there is clear influence of the temporal structure of the visual stimuli onto the temporal structure of the lfp power. The grating is a uniformly moving stimulus and similar uniform is the lfp activity elicited by this stimulus. Natural movies contain a large variety of movements with different speeds and directions. This irregularity seems to transfer to the lfp power, which shows a very patchy pattern. This locking to the temporal structure of the stimulus is analysed in more detail below (see chapter III.8).

III.6 Response properties of the high frequency activity

We recorded two signals characterizing the high frequency activity: the analog-Mua and the multi-unit spiking activity. Our results show that both of these share many response properties with the lfp.

First, the average responses for the different stimuli described above for the lfp were very similar also for the analog-Mua and the spiking activity. Concerning the latter, natural movies elicited slightly higher firing rates than the pixel noise (average of 6.6%, n=33 sites). The firing rates for gratings in contrast were much lower. Overall it seemed that the multi-unit activity is slightly stronger for the natural compared to the noise stimuli, however, this effect is weak compared to the overall difference between stimuli.

Second, the temporal activity profile for natural movies and their manipulations show many irregularities that are related to the temporal structure of the stimulus. Similar to the analysis for the lfp below (III.7) it can be shown that these activity fluctuations are locked to changes in the velocity profile of the stimuli.

III.7 Locking to the temporal structure of the stimulus

As noted above, the power of the lfp shows an irregular temporal activity pattern upon stimulation with natural movies. Our natural movies recorded by a camera mounted to a cat’s head have a rich temporal structure. This structure is characterized by steady passages interrupted by fast and short movements, e.g. head saccades, and passages of more uniform
continuous motion in one direction, e.g. forward walking (c.f. Chapter II). It seems reasonable that the fluctuations of the lfp power are actually locked to the velocity profile of these movies. Using correlation analysis we confirmed this.

![Cross correlation between the local field potential power and the velocity curve of a natural movie. The cross correlation was computed for each frequency of the local field potential independently (values on the x-axis). The lag of the cross-correlation is shown on the y-axis (in ms) and the correlation strength is colour-coded.](image)

**Figure III.6.** Cross correlation between the local field potential power and the velocity curve of a natural movie. The cross correlation was computed for each frequency of the local field potential independently (values on the x-axis). The lag of the cross-correlation is shown on the y-axis (in ms) and the correlation strength is colour-coded.

The above figure shows the cross correlation between the lfp power at different frequencies and the velocity of the stimulus. At many frequencies is the power of the local field potential correlated with the velocity curve of the stimulus. Across animals we found prominent locking in two frequency ranges, between 23 and 39 Hz and above 109 Hz.

Given that natural movies have a very rich spatial structure, we asked whether this locking phenomenon is related to this spatial structure or whether it occurs also for much simpler stimuli. We constructed two types of stimuli moving according to the motion pattern extracted from a natural movie: A sine-wave grating and a still image taken from a natural movie. Both lead to the locking of the lfp power to the stimulus motion. Thus, this locking is not restricted to a particular spatial structure of the stimulus.

Locking of neuronal activity has been described in previous experiments. Rager & Singer (1998) reported locking of the phase of the local field potential to an oscillating full field stimulus. The same type of locking was observed upon stimulation with irregularly moving gratings (Kruse & Eckhorn 1996). In contrast, here we investigate modulation of the power in different frequency bands of the lfp. The modulation of the power of a given lfp frequency can take place on much slower time scales than the oscillation of this lfp frequency. For example, the power of the 100 Hz frequency component of the lfp can be modulated by a 4 Hz signal in the visual stimulus. As a consequence, the details of visual stimulation, such as the monitor refresh rate and the flicker fusion frequency limit only the signal in the stimulus that could cause the locking, i.e. the 4 Hz signal. The lfp frequency at which locking is observed, i.e. 100 Hz in this example is only limited by the Nyquist frequency of the sampling of the electrode signals.
This locking and the related irregular temporal activity profile is thus an inevitable property of the processing of natural stimuli. Either, the visual system directly extracts information form these activity fluctuations or it must be insensitive to these and extracts other information from the activity. Many studies report modulations of the mean activity in the tonic part of neural responses. Examples of such modulations in primary visual cortex are for example contextual effects (Nothdurft et al. 2000) and figure ground modulations (Lamme 1995; Roelfsema et al. 1998), which both are especially relevant for the processing of natural scenes. These studies report an average modulation of single unit firing rates of the order of 25%. The larger the activity fluctuations are the longer one needs to observe a given signal in order to reliably estimate the mean. Thus, the activity fluctuations set a certain limit with which one could extract such differences in the mean activity. To quantify the impact of the fluctuations in our population responses we measured the length of the interval necessary to reliably detect a 25% modulation of the recorded mean activity. We found that on average it takes roughly 400 ms in order to reliably detect at a 25% modulation. The intervals necessary to detect changes in the average activity reported above are long compared to visual latencies of cells in higher visual areas (Perrett et al. 1982; Oram and Perrett 1992) and reaction times of experimental subjects (Thorpe et al. 1996, van Rullen & Thorpe 2001). Thus, the transient fluctuations of the activity evoked by the motion the natural stimuli impose constraints on theories of cortical representations relying on differences in mean activations.

III.8 Feature tuning and locking are prominent in complementary frequency ranges

As described in the last section, the power of the LFP can be locked to the temporal structure of the stimulus. This locking occurs in two frequency ranges and in these frequency ranges the LFP activity is driven directly from external sources. In the complementary frequency ranges, the LFP power is dominated by internal processes. These are not locked to the stimulus and also not phase locked across repeats of the same stimulus. What influences the LFP power in these frequency bands?

In section III.4 we showed that the LFP power is selective to the classical stimulus features like orientation and spatial and temporal frequency. A previous study in our lab determined the optimally orientation tuned frequency band (Siegel & König, 2003). Having these and the present results in mind, we realized that the frequency regimes that do not show locking, coincide with those that do show orientation tuning.

Following this observation, we systematically studied tuning to orientation, spatial and temporal frequency of the LFP power. For each of these features, we measured the strength of tuning in different frequency bands of the local field potential. It turned out that the LFP is tuned to these features with similar selectivity and that the tuning occurs in overlapping frequency bands.
Comparing the frequency bands that show frequency locking to those that show tuning to the stimulus features it turns out that they are complementary (Fig). Prominent tuning occurs between 8 and 23 Hz and between 39 and 109 Hz. Locking in contrast is occurs between 23 and 39 Hz and above 109 Hz. Furthermore, these frequency ranges together cover the entire frequency axis. Thus, only by considering phenomena directly related to stimulus processing one can attach a ‘function’ to nearly all frequencies investigated. This is in contrast to previous literature where often single frequency bands have been reported to be activated by a particular phenomenon and many frequency bands were left as ‘silent’ aside. This might be explained because many studies use stimuli without an intrinsic temporal structure, as for example static textures or uniformly moving bars. These stimuli do not cause stimulus locking, and thus the corresponding frequency bands are silent.

Figure III.7. Locking and feature tuning are prominent in complementary frequency ranges that together cover the entire frequency axis. The solid curve describes the average tuning indices for orientation, spatial and temporal frequency. The dashed curve describes the correlation of lfp power and stimulus motion.

III.9 Eye movements

A major drawback of the experiments presented above is that we did not take the eye movements of the animals into account. Instead we averaged across all positions and movements of the eye. Of course, the visual input for the cats’ eye is altered by eye movements and these impose a new temporal structure on the retinal stimulus that originally was not intended. However, there are several reasons why we think that eye movements did not strongly confound our results.

First, cats do not make many eye movements. In our recording setup, a study comparing eye movements of cats and humans found that head-fixed cats make a saccade roughly every 3.6 seconds, while head-fixed humans make a saccade every 300 ms (Moeller et al. 2003). We confirmed this number in some recording sessions using the identical stimuli as in the physiology experiments. The stimuli used in the physiology experiments lasted 2 seconds.
Assuming that eye movements are independent of the stimulus presentation one would expect that only a smaller fraction of all eye movements occur during a stimulus.

The assumption that eye movements are independent of the stimulus presentation and especially the motion in the stimulus might seem unrealistic. However, a systematic study using different stimuli revealed that stereotyped eye movement patterns like the optokinetic nystagmus are difficult to elicit in cats. While slowly drifting gratings with a low spatial frequency do cause an OKN like pattern of eye movements, the gain of this OKN decreases strongly with increasing frequency. Furthermore, as soon as an irregular motion is superimposed on the uniform drift, the stereotyped OKN breaks down (Moeller et al. 2003). Thus, all our natural and complex stimuli, that have a very irregular motion, do not elicit stabilizing eye movements. Furthermore, most grating stimuli used have temporal frequencies above that a strong OKN can be elicited. We confirmed this in a number of recording sessions measuring tuning to temporal frequency and eye movements at the same time (see Kayser et al. EJN, 2003). The eye movements did not show any significant coupling to the stimulus presentation. Because eye movements occur rarely and are not systematically locked to the stimulus presentation, averaging the responses to many repeats of the stimulus should eliminate the influence of eye movements on our results.

### III.10 Summary of the results

We used microelectrode recordings in areas 17/18 of awake cats to compare the neuronal responses to classical drifting gratings, natural and other complex stimuli. From these recordings we extracted several measures characterizing different aspects of the underlying neuronal activity: Most prominently local field potentials (lfp) and multi-unit spiking activity. The lfp response to gratings differs in two respects form the response to various complex stimuli. First, the frequency range activated by gratings is narrow and restricted to the gamma band (roughly 30-80 Hz). In contrast, all complex stimuli activated a broad range of lfp frequencies. Second, the temporal structure of the lfp response to gratings is uniform over time whereas it shows strong modulations during the presentation of natural movies. Considering the spiking activity, there was a significant difference in the firing rates for optimal gratings compared to the natural movies. Thus, both the averaged response strength and the response timing are different for classical drifting gratings compared to natural movies.

We used natural movies and a stimulus with the same second order, but random higher order structure (pink pixel noise) to probe the effect of the higher order structure on the responses. The frequency range of the lfp activated by these two stimuli is identical as is the average response. Considering the firing rates, there was only a small difference between the responses for these two stimuli. Thus, the lfp and also the multi-unit spiking activity are not sensitive to the higher order statistical structure of the stimulus.

Examining the temporal profile of the lfp response to natural movies, we showed that fluctuations of the lfp power are locked to the velocity profile of the natural stimulus. Further
experiments showed that such a locking is not restricted to natural stimuli but occurs for many different spatial stimuli moving according to an irregular velocity profile. This locking of modulations of the LFP power is most prominent in two frequency bands of the LFP. Several proposed schemes for the coding of visual information in V1 rely in differences in the mean responses. Our analysis shows that the induced modulations of the LFP power severely limit the read-out of possible differences in mean response. As a result, under natural conditions there exists no steady state response, but the activity pattern consists of a series of transient evoked responses.

The firing rates of single units show tuning to several features of the stimulus: spatial and temporal frequency and orientation. We find that the LFP power as well is tuned to these classical features with similar specificity and that the shape of the tuning curves resembles those known from the firing rates of single units. The tuning occurs only in two frequency bands of the LFP. These are the same for all three features and are complementary to those in which the LFP power is locked to the temporal structure of the stimulus. Together these frequency bands cover the entire frequency axis. Thus, only by considering phenomena directly related to stimulus processing one can attach a ‘function’ to all frequencies investigated.

III.11 Directions of future work & conclusion

The difference in the modulation of the local field potential between natural and the noise stimuli was insignificant. Given that natural movies and the noise stimuli have a very different higher order statistics and a different appearance, there must be a difference at some point in the visual system. It could be that either the activity measures used here, especially the LFP, are not sensitive enough or the difference resides at higher processing stages. An indication for the first point could be seen in the fact that the firing rates for natural movies were slightly, but significant higher than for the noise stimuli. So it might be that individual neurons show a significant difference, however due to the pooling over several neurons this difference is lost. To elucidate this further, new experiments are needed, that record the spiking activity of well-isolated cells and in different areas of the visual system.

Furthermore, it would be interesting to compare the response properties of the local field potential to those of other population measures of neuronal activity. The EEG has very similar properties as the local field potential and thus would allow a direct comparison. The EEG can be measured from human subjects and would allow making a comparison across species. Another measure of neuronal activity, which is influence by a large number of neurons, is the BOLD signal measured with fMRI. Measuring the responses to these different stimuli in the fMRI would not only allow an investigation of the responses in many different visual areas simultaneously, but would also allow a direct comparison of results from the different methods.
To better understand the difference between the activity for gratings and natural movies, new stimuli are necessary. While natural stimuli are broadband, gratings cover only a very narrow range of spatio-temporal frequencies and the different frequencies follow a global phase alignment. For natural stimuli, the different spatio-temporal frequencies are aligned such that the resulting motion has a very variable structure as discussed in Chapter 2. Periods with no motion are interspersed with rapid translations. This irregular motion structure leads to the locking phenomenon described above which is most prominent at frequencies above 100 Hz. However this is exactly the frequency range in which the different in lfp modulation between natural stimuli and gratings is strongest. Thus, the type of motion is the most promising candidate feature of the stimulus that needs to be varied in order to see a smooth transition from the activity patterns for gratings to that for natural movies.

The present results show that the activity pattern for gratings differs from that to natural stimuli. This answers one of the questions posed in the introduction and directly leads to the next question: Which results that were established using gratings do generalize to the processing under every day conditions? We have no answer for this question. However, these results highlight the importance of using stimuli with different global structure for the analysis of sensory systems. Otherwise it will not be clear which mechanisms or response properties are important under every day conditions and which are just an artefact from the use of artificial stimuli.

Analysing the general response properties of the local field potential, we found that the power is tuned to the classical stimulus features orientation, spatial and temporal frequency. Tuning to these features is prominent in two frequency bands and these frequency bands are complementary to those in which the power is tightly locked to the stimulus. Together these frequency ranges are complementary and together cover the whole frequency band investigated. Thus, the whole frequency band of the local field potential can be assigned a correlated of visual processing and no frequency range remains ‘silent’. However we do not want to create the impression that the stated frequency bands serve a single purpose. From numerous studies, either recording lfp or EEG, it is well known that state changes of the subject modulate low frequency activity in very narrow bands (McCormick and Bal, 1997; Destexhe et al., 1999). The most prominent modulation of the EEG occurs in the wake sleep cycle in the range of frequencies below 10 Hz (Acherman & Borbely 1995). Similar, higher order effects like attention can have profound effects on the activity recorded in EEG studies (Herrmann & Knight, 2001; Paller et al. 2003; Fell et al. 2003). Further, task related memory was shown to modulate the EEG (Tallon-Baudry et al., 1998; Jensen et al., 2002). Thus, there are a variety of higher order phenomena that can modulate different frequency ranges of the EEG and the local field potential besides the modulation by either feature tuning or stimulus locking described above. However, our results show that passive visual processing is enough to activate all frequency bands of the local field potential.
The present results show that the local field potential is an interesting measure of neuronal activity that can be used to study both, temporal properties of activity patterns but also classical selectivity such as tuning to different features. Many studies that measure local field potentials in the visual system study only the gamma frequency range. Our results indicate why this is so: Gratings activate mostly frequencies in this band but not outside. However, the results also indicate that upon stimulation with broadband stimuli the entire frequency axis of the lfp becomes interesting.

III.12 Publications

The results presented in this chapter are published in the following papers:


IV. Hemodynamic responses to complex stimuli in the cat visual system

Microelectrode recordings allow measurements of localized neuronal activity such as the spiking activity of individual neurons. While this method of measuring neuronal responses has yielded a large amount of knowledge about brain function, it has some disadvantages. First, when inserting the electrode into the brain, some parts of the tissue will be destroyed. This effect might be negligible due to the miniature size of today’s electrodes; its impact, however, is unknown. Second, while good localization is desirable for some questions, for other questions a method quantifying the response strength in many different brain areas is more suitable.

The most prominent technique in neuroscience allowing non-invasive measurements of brain function is functional magnetic resonance imaging (fMRI). This method became very popular in the last decade and is ‘the’ method used with human subjects. It measures hemodynamic responses related to neuronal activity. Its most popular variant, measurements of the blood oxygenation level dependent signal (BOLD), relies on a close link of neuronal activity and blood oxygenation. The BOLD signal measures the local concentration of deoxyhemoglobin, which is influenced by the rate of deoxygenation of haemoglobin and changes in blood flow and blood volume (Buxton 2001). However the question how these three factors in turn are related to neuronal activity remains an open question (c.f. reviews in J. Neurosci. 23(10), 2003, pp. 3959-4011).

The fact that using fMRI one can measure a signal related to neuronal activity in different brain areas simultaneously, makes fMRI an interesting technique for the questions posed in chapter 1. The comparisons of the neuronal responses to natural and other stimuli presented in chapter 3 are based on microelectrode recordings and thus localized to very small regions of the brain. Such recordings notoriously suffer from a sampling problem. Using fMRI, however, we could extend the comparison to a larger area of the visual system. Furthermore, since we already have measurements of neuronal activity from microelectrode recordings, these data can also be used to conduct a more detailed comparison of electrophysiological measures of neuronal activity with BOLD responses.

IV.1 Imaging the cat visual cortex

We measured BOLD responses in the cat visual cortex to the same set of stimuli that was used in the physiology experiments described in chapter 3. We made use of high-field fMRI that allows measurements of BOLD signals with high spatial resolution. Compared to the microelectrode recordings, however, there is one important difference. While the
electrophysiology was obtained from awake animals, this is not possible with the fMRI. The subjects need to rest within the scanner for a long time and any movement can lead to severe artefacts. Thus these experiments had to be conducted on anaesthetized and paralysed animals.

In contrast to the physiology experiments, where responses were sampled in small regions of the brain along different electrode tracks, the BOLD responses were measured in different areas of the visual system. Significant responses were found in the lgn, areas 17,18,19,21 and in PMLS.

![Example slices from one subject. Shown is the correlation of the BOLD signal with the stimulus paradigm colour coded on top of the anatomy map.](image)

**Figure IV.1.** Example slices from one subject. Shown is the correlation of the BOLD signal with the stimulus paradigm colour coded on top of the anatomy map.

The response in each area was calculated as the average change of the BOLD signal from blank to stimulus. This percent change was then averaged across all voxels that showed significant correlations of the BOLD signal with the stimulus paradigm.

![Average BOLD response in areas 17 and 18 for the three types of stimuli and one subject.](image)

**Figure IV.2.** Average BOLD response in areas 17 and 18 for the three types of stimuli and one subject (Black: gratings; dark Gray: natural movies; light gray: pixel noise).

The response pattern was very similar in the two primary visual areas 17 and 18 (Figure 2). Given that cells in these two areas have only slightly different preferences for spatial and temporal frequencies (Movshon et al., 1978) and that most of our stimuli contain a large range
of frequencies, this is not unreasonable. A similar relation was found between the responses in area 19 and area 21a. We therefore averaged the activities across areas 17 and 18 and across areas 19 and 21a. Besides making the results more transparent, this has two further advantages. First, defining the different visual areas from the anatomy scans is subjective and somewhat arbitrary. Combining two areas reduces the number of borders that are critical for the results and thus reduce the number of voxels that could possibly be assigned to the wrong areas. Second, in contrast to the border of areas 17 and 18, the border of 18 and 19 contains the representation of the perimeter of the visual field. Given that the stimulation was restricted to the central 60 degrees, small errors in the definition of this border will not strongly affect active voxels.

**IV.2 BOLD responses to the different stimuli**

![Graphs showing BOLD responses in different areas](image)

**Figure IV.3.** Average BOLD responses in the different areas for the three types of stimuli and all subjects (Black: gratings; dark gray: natural movies; light gray: pixel noise).
The above figure displays the BOLD activations for the different types of stimuli and five animals. Clearly, the overall response strength depends on the animal and area. The strongest responses occurred in the thalamus (1.69% averaged across animals and stimuli) and the weakest responses in the primary visual cortex (0.49% on average).

Comparing the responses to different stimuli, at first glance it is difficult to find any consistent pattern. To get a rough overview over the response properties of the BOLD signal, we used the following method to extract ‘consistent’ results. We compared the activity for two different stimuli in the same animal and counted in how many animals a given stimulus was stronger than another stimulus. With this method one finds that:

- In the LGN: no relations between stimuli holds in at least four of five subjects
- In areas 17/18:
  - gratings lead to the strongest responses
  - the pixel noise leads to stronger responses than the natural movies.
  - The wavelet filtered stimulus leads to stronger responses than the wavelet noise
- In areas 19/21a: no relations between stimuli holds in at least four subjects
- In PMLS:
  - Gratings leads to stronger responses than the pixel noise
  - The wavelet filtered stimulus leads to stronger responses than the wavelet noise

As a general result we see that gratings (with roughly optimal frequency parameters for the areas studied) are at least as efficient in eliciting hemodynamic responses than complex stimuli. Furthermore, there is no clear difference between the BOLD responses to the natural and the respective noise stimuli. While in the primary visual cortex the response to the pixel noise is stronger than to the natural movie, the wavelet noise elicits a weaker response than the wavelet filtered stimulus. In the higher visual areas no difference is consistent across animals. Thus, there is no clear effect of the higher order structure on the average BOLD responses.

The results obtained from the BOLD responses in the primary visual cortex can be compared to the responses measured in the local field potential and the multi-unit firing. The responses caused by gratings in the local field potential were restricted to a narrow frequency band. Within this band, however, the responses to gratings were stronger than the responses to any other stimulus. From this one could suggest that the BOLD signal is most sensitive to this frequency range of the local field potential or stated differently that the responses properties of local field potentials and the BOLD signals match best in the gamma range.

Comparing the BOLD activations in the primary visual cortex to natural movies and the pixel noise, one finds that in most subjects the noise causes stronger responses. This is in contrast to the local field potential were there was only a very small and insignificant difference, at least
when using the entire frequency band of the local field potential for the comparison. However, the multi-unit firing rates in these areas showed a significant difference between natural movies and pixel noise. The firing rates for the natural movies were stronger. Thus the BOLD responses yield the opposite result than the spiking activity.

In the following we will compare the responses properties of the different measures of neuronal activity in more detail. First, however, we will give a short review about the known relations between the neuronal activities measured by the different techniques.

IV.3 The origin of the BOLD signal

Neuronal activity is accompanied by a large consumption of energy (for a recent estimate see Lennie, 2003). First, neuronal firing leads to an increased metabolic demand for oxygen. Further, astrocytes recycling the neurotransmitter glutamate from synaptic clefts require large amounts of glucose. This increased need for oxygen and glucose upon synaptic transmission results in an increased blood flow (Heeger & Ress, 2002 for a review). However, this is only a vague picture of the complicated and interrelated processes caused by neuronal activity. In the last years, increased effort was made to study both, the molecular processes that finally cause an increase of oxygen consumption and blood flow, and the relationships between hemodynamic and neuronal responses to different stimuli.

For its application in neuroscience, the relation of the fMRI-BOLD signal to other measures of neuronal activity is of great importance. Recently, a number of studies made a step toward a detailed comparison of the BOLD signal with other popular measures of neuronal activity. For example, combinations of fMRI with both simultaneous and separately recorded electroencephalographic signals in humans and animals (Menon et al., 1997; Bonmassar et al., 1999; Brinker et al., 1999; Ogawa et al., 2000) showed qualitative and quantitative agreement between these two methods. Recording a direct measure of neuronal activity, the firing rate of individual cells, another group of studies compared spiking activity and BOLD signals in the visual system. Heeger and co-workers (Heeger et al., 1999) recorded firing rates of neurons in the motion selective area MT of monkeys and BOLD responses from the MT homologue in humans. They found similar levels of motion opponency in both experiments, indicating a qualitative agreement. Using a similar paradigm, Rees et al (2000) found that BOLD and firing rates were directly proportional by changing the coherence of stimulus motion in a parameterised way. A similar result was obtained in a different experiment comparing contrast response curves of BOLD and spikes in the primary visual cortex (Heeger et al., 2000). Summarizing, these studies suggest that the BOLD signal is a good representative of spiking activity and that fMRI can be used to infer properties of spiking activity in a non-invasive manner.

However, there is increasing evidence that the match of fMRI and spiking activity might not hold in general and that other measures of neuronal activity than firing rates might better.
correspond to the BOLD signal. Evidence for this comes from simultaneous recordings of local field potentials, multi-unit activity and BOLD signals in the primary visual cortex of anaesthetized monkeys (Logothetis et al., 2001). In general, both the local field potentials, as well as the spiking activity correlated with the changes in the BOLD signal. At many recording sites, however, the local field potential correlated slightly better with the hemodynamic response. An even stronger dissociation between spiking activity and hemodynamic responses was demonstrated in the rat cerebellar cortex (Mathiesen et al., 1998; Mathiesen et al., 2000). Using different stimulation paradigms, these authors could demonstrate correlated increases of local field potentials and cerebral blood flow in the absence of spiking activity. Given that local field potentials measure local subthreshold potentials in dendrites and somata (Mitzdorf, 1985; Mitzdorf, 1987; Juergens et al., 1999), the picture emerging from these studies is that the fMRI signal represents more the input to a local region of the brain than the spiking activity within this region (Logothetis, 2003).

In the next section we extend the comparison of our results from the electrophysiology and fMRI experiments. Previous comparisons of fMRI and electrophysiology showed that local field potentials and spiking activity match differently to the BOLD responses and that this match varies strongly between individual recording sites (Logothetis et al. 2001). At some sites, the local field potential and the firing rates matched the BOLD signal equally. At other sites the match of BOLD and local field potential was better. Based on this observation we investigated the relation between BOLD and electrophysiology for individual recording sites and counted the number of sites at which different response measures agree.

IV.4 Comparing BOLD and electrophysiological measures of neuronal activity

The BOLD responses presented above were obtained by averaging the percent signal change of individual voxels across anatomical areas. The responses in the electrophysiology experiments, however, have been obtained from much smaller regions within the different areas. They were sampled along the electrode tracks within the representation of the central visual field in areas 17/18 on the posterior lateral gyrus and in area 21a on the suprasylvian gyrus. In order to base the comparison of the two methods in data from similar cortical reasons we defined two regions of interest (ROI) in the fMRI data following the electrode placement in the physiology experiment. One ROI was defined as the region on the posterior lateral gyrus in two consecutive slices with similar A-P Horsley-Clarke coordinates as in the electrophysiology recordings. Similar, a second ROI was defined as the region on the suprasylvian gyrus in two consecutive slices. We then computed the average BOLD response within these ROI’s. It turns out that the overall response strength within these ROI’s is different than within the entire areas but that the relative activity for different stimuli is similar. In detail we find that the following comparisons between stimuli hold in four out of five subjects:

- In areas 17/18:
  - Gratings elicit stronger responses than natural movies
- Gratings elicit stronger responses than the pixel noise
- The pixel noise elicits stronger responses than natural movies

- In area 21a:
  - Gratings elicit stronger responses than natural movies
  - Gratings elicit stronger responses than the pixel noise

In the following we count the number of recording sites in the physiology experiments at which the relative activity for the different stimuli obeys the same relations as in the fMRI. This is done separately for the different frequencies of the local field potential, for the analog-Mua and the firing rates. Note that applying a similar criterion to the fMRI data would not lead to a hundred percent match, since some of the differences between stimuli that form the basis of this comparison were only found in four out of five animals.

![Figure IV.4. Comparison of BOLD and electrophysiology. For each relation between stimuli extracted from the BOLD responses, the graph shows the percentage of recording sites at which the responses were in agreement with the respective relation. The comparison was done separately for the different frequencies of the local field potential, for the analog-Mua and the spiking activity.](image-url)

Counting the number of recording sites in areas 17/18 at which the response of the local field potential to gratings exceeds that to natural movies yields the curve shown in (A). The number of recording sites that are in agreement with the relations from the fMRI depends on the local field potential frequency. The best match is found in the gamma range (88 % at 42 Hz) and the percentage decreases quickly at low and high frequencies. Averaged across
frequencies this results in a match of 27 %. The analysis for the relation between gratings and pixel noise yields a similar result (B). In contrast, the comparison of pixel noise to natural movies is less dependent on the frequency (C). The curve reaches its maximum in the gamma range (68 % at 29 Hz) and yields a match of 51 % averaged across frequencies. In area 21a similar results were found. As a consequence, we note that a comparison of results from fMRI and local field potentials can strongly depend on two factors: the frequency band of the local field potential used and the types of stimuli involved.

Comparing the responses of the analog-Mua to the BOLD responses we find that the comparisons involving gratings yield a bad match (roughly 20%). In contrast, comparing the responses to the pixel noise to that to the natural movies, the analog-Mua agrees with the fMRI at 65 % in area 18 (C). Overall, the comparison of the analog-Mua to the fMRI yields a similar match as found for the local field potential.

Comparing the results from the spiking activity with the fMRI we find that concerning the comparisons involving gratings, the spiking activity yields a worse match with the fMRI than does the analog-Mua. At no site was the firing rate for the gratings higher than for the natural movies. Only at 20 % of the sites in areas 17/18 and 5 % in area 21a was the firing rate for the gratings higher than for the pixel noise. Comparing the activity for the two complex stimuli, the spiking activity matches the fMRI at 46 % of the sites in areas 17/18 and 50 % in area 21a. Thus, the comparison of the spiking activity to the fMRI for the complex stimuli yields worse match than does the local field potential or the analog-Mua.

These results show that a comparison of fMRI responses and the different measures of neuronal activity is dependent on the type of stimuli used. The comparisons involving gratings show a strong influence of the frequencies of the local field potential. However, when using only complex stimuli the details about frequency ranges used become less important. We hypothesized that this frequency dependence should become even weaker when only natural stimuli are used. From the BOLD activations for the three individual natural movies we extracted their relative activations. On average across animals the three natural movies induced different BOLD responses (1.15, 1.05 and 0.87 %) and the difference of the activity induced by these movies is consistent in four out of five animals. Figure (D) shows the number of recording sites in areas 17/18 at which the local field potential obeyed these relations. Indeed, the dependence of this curve on the frequency axis is small and yields an average of 56 %. A similar comparison with the activations from the analog-Mua yields a match of 68 %. In sharp contrast, a comparison with the spiking activity shows that only 25 % of the recording sites show the same relative activations for the different natural movies. Thus, when only natural stimuli are considered, a comparison of local field potentials and BOLD responses becomes insensitive to the range of frequencies used. Furthermore, the dissociation between activations in the local field potentials and in the analog-Mua on the one hand and the thresholded spiking activity on the other hand becomes stronger and the spiking responses show a weak concordance with the BOLD responses.
The above analysis used only single frequencies of the local field potential. Most studies investigating this signal however use frequency bands for the analysis. We thus performed a similar analysis above but using frequency bands instead of single frequency bands. The results support those from the single frequencies and show that these results show that independently of the stimuli used, the match of BOLD activations and local field potentials is strongest for frequency intervals roughly in the range of 20 Hz to 50 Hz.

IV.5 Summary of the results

Using the same set of stimuli as in the electrophysiology experiments, we conducted a series of fMRI scans measuring BOLD responses in different visual areas of anaesthetized cats. Averaged across all voxels within the primary visual cortex, gratings lead to the strongest responses. Similar to the electrophysiology results, there was no clear difference in the responses to the natural movies and the pixel noise stimuli. This latter result holds as well in the higher visual areas 19 and 21a. Thus the BOLD measurements support the electrophysiological findings that the large-scale neuronal activity in the primary visual areas is not sensitive to the higher order statistical structure of natural scenes and suggest that this holds at least in the next processing stage as well.

Based on the responses at different recording sites we performed a detailed comparison of the response properties of the BOLD signal on the one hand and the different electrophysiological measures of neuronal activity on the other hand. The responses measured in the BOLD signal were by an order of magnitude weaker than those measured in the analog-Mua or the lfp. The responses in the spiking activity, in contrast, were another order of magnitude larger. Thus, BOLD measurements are likely to underestimate the strength and reliability of neuronal responses. In general we found that a comparison of BOLD and electrophysiology depends on the details of the definition of the electrophysiological activity measure. For example the match between BOLD and lfp depends on the lfp frequency band considered. In general the BOLD signal matches the local field potential best within the gamma frequency range (20-50 Hz). Thus it might well be that the BOLD signal is most sensitive to this classical frequency range often studied in conjunction with visual processing. The spiking activity, in contrast, performed worse compared to the lfp or compared to the analog-Mua. Overall the match between BOLD and spiking activity was significantly weaker. Furthermore, depends the comparison on the type of stimuli used. Overall the best match between BOLD and lfp was obtained when the comparison was based only on complex stimuli. The dissociation between spiking activity and lfp was strongest when the comparison was based on natural stimuli only. These results support the notion that the lfp corresponds more to the input to a cortical region than to its output.
IV.6 Directions of future work & conclusion

Comparing the local field potential to the BOLD signal we found the best match in the gamma frequency band (20-50 Hz). The gamma band has often been used to investigate visual processing in the local field potential and is the frequency band in which correlates of higher order effects like feature binding or object selection have been reported (Singer and Gray, 1995). However, the precise origin of the strong gamma oscillations is unresolved and its significance with respect to coding global stimulus properties is hotly debated (Singer, 1999; Shadlen & Movshon, 1999). Most models addressing these experimental data assign a prominent role to inhibitory mechanisms (von der Malsburg, 1999). Inhibitory interactions between neurons do not increase the overall spike rate but consume energy and lead to an increased blood flow, as supported by experimental results obtained in the cerebellum (Mathiesen et al., 2000). As a result, fMRI measurements are sensitive to this type of process. Thus, the data presented here are compatible with the speculation, that by virtue of inhibitory processes, fMRI measurements are sensitive to activity in the gamma frequency range.

The results presented above support existing data showing that the BOLD signal in general corresponds more to the lfp than to the spiking activity. However, the mismatch between lfp and spiking activity depends on the recording site (Logothetis et al. 2001) and, even stronger, on the experimental paradigm (Mathiesen et al. 1999). Given that the BOLD signal is very slow and determined by large-scale processes (blood flow), it seems reasonable that is influenced by several sources of neuronal activity, local input and spiking activity within a region. Thus, there might be no final answer to the question what the electrophysiological correlate of the BOLD is, but the answer depends on many parameters of the experimental paradigm.

In order to investigate the relation between two signals, two different strategies seem suggestive: the use of stimuli that maximize the difference between the two signals or the use of stimuli leading to very similar responses. Here the present results add to the literature, as they indicate, at least for the visual system, which stimuli belong to which of these two classes.

The strongest limitation of the results presented above is the differential use of anaesthesia for the two populations of animals. The electrophysiology was obtained from awake animals while the fMRI data were obtained from anaesthetized animals. Ideally the comparison should be performed in the same animal and the different signal should be recorded simultaneously. However, this can only be achieved by a technical tour the force (Logothetis et al. 2001). There is no clear knowledge about the effect of anaesthetics on the response properties of cortical neurons. As already mentioned in chapter 3 anaesthetics influence the pattern of spontaneous activity (Erchova et al 2002) and also properties of receptive fields like their size (Roberston 1965; Lee 1970; Friedberg et al. 1999). Also, higher order effects like figure-ground segmentation seem to be influenced by the anaesthesia (Lamme et al. 1998). Thus, it is clear that anaesthetics do influence the response properties also in V1. However, the effect of
a given anaesthetics can be difficult to judge. Even anaesthetics of the same class, like halothane and isoflurane, have been reported to have differential effects on global measures of brain activity (Villeneuve et al., 2003). Interestingly, effects of sub-aesthetic doses of isoflurane, as measured by the fMRI technique, are noticeable in many cortical areas but seem to spare primary visual areas (Heinke and Schwarzbauer 2001). Along similar lines, the effect of ketamine on neuronal responsiveness as measured by the BOLD signal occurs only at high doses and shortly (<1 hour) after i.m. injection (Leopold et al., 2002). Thus, the effect of the anaesthetics might be more modulatory and might not change the relative responses to different stimuli. However, at the present time this is only speculation and further investigations are required to proper understand the effects of anaesthetics on the BOLD responses measured here.

**IV.7 Publications**

The results presented in this chapter are published in the following paper:

V. Optimality criteria linking properties of V1 cells and image statistics

The idea that the statistics of the sensory stimuli we encounter in our natural environment are important for perception and cognition has been around for a long time (Mach 1886; Helmholtz 1925; Barlow 1953). The hypothesis that sensory systems are adapted to the statistical properties of their every day input, however, has received increased interest only in the last years (c.f. chapter 1). Natural sensory stimuli usually are very complicated and high dimensional signals. However, their effective dimensionality is reduced by the many redundancies in these signals (Barlow 2001). In view of this, it was proposed that to process such input efficiently the brain should reduce this redundancy, but also retain the important information (Barlow 1961): “… sensory relays encode sensory messages so that their redundancy is reduced but comparatively little information is lost.” This hypothesis is known as the redundancy reduction hypothesis.

V.1 Redundancy reduction and sparse coding

The idea of redundancy reduction was motivated based on the work by Shannon (1951) on information transmission in communication channels and developed in a similar ways by Attneave (1954) and Barlow (1961). It turned out that coding schemes following this redundancy reduction lead to activity patterns with the following two properties: many neurons are activated at the same time and each neuron is activated by many stimuli. Such codes are very demanding in terms of energy consumption in the brain and furthermore sub-optimal for learning rules and probability extraction (Barlow 2001). To overcome these problems, a coding scheme that allows both, redundancy reduction and saving of energy was proposed: “The principle of recoding is to find what messages are expected on the basis of past experience and then allot outputs with few impulses to these expected inputs, reserving outputs with many impulses for the unusual and unexpended inputs” (Barlow 2001). This is known as the sparse coding hypothesis. Stated in other words it requires that a neurons activity should be zero or close to most of the time and reach a high value only at some rare events.
The link with the image statistics was made by the seminal studies of David Field. He explored the properties of natural images using filters with properties similar to those of the simple cells in primary visual cortex. It turned out the activities produced by such filters on natural images have a highly kurtotic and thus sparse distribution (Field 1987).

This lead to the idea, that one could obtain simple cell receptive fields from natural stimuli by requiring model cells to have a sparse activity distribution. Indeed Olshausen & Field (1996) showed that starting from a population of model cells with random receptive fields, a population with properties very similar to those of simple cells in V1 can be obtained by optimising a sparseness objective function. There results were confirmed by other researchers extending such models into the temporal dimension (van Harteren & Ruderman 1998). The conclusion following from these studies is that simple cells form an optimal sparse representation over natural images (Olshausen & Field 1996; Bell & Sejnowski 1997; van Harteren & van der Schaaf 1998; Hoyer & Hyvarinen 2000). Thus the sparse coding hypothesis forms a link between the statistical properties of natural scenes and the properties of neurons in the visual cortex.

Besides the properties of V1 receptive fields also the properties of cells at earlier stages of processing could be explained based on the redundancy reduction principle. The receptive fields of neurons in the LGN were shown to be optimal in the sense that they whiten the natural input (Atick & Redlich 1990). These theoretical predictions were later on confirmed in electrophysiological experiments (Dan et al., 1996). Sensory information taken up by the photo receptors is send to the visual cortex via the optic tract. This communication channel imposes a bottleneck onto the visual system, and it was shown that the receptive fields of the retinal ganglion cells optimally whiten the natural input in the presence of noise (Atick & Redlich, 1992). Similar advances could be made in the understanding of retinal colour coding (Atick et al. 1992). In the following we extend this approach and study how the response properties of neurons at higher processing stages can be modelled in networks based on the optimisation of responses to natural scenes.
V.2 Predictive and stable coding

The response properties of neurons at higher stages have two properties: First, they show increasing selectivity to higher order features. For example, complex cells in V1 are selective to orientation and spatio-temporal frequency while cells in IT are selective to complex shapes. Second, the responses are insensitive to a number of transformations of the stimulus like translation, reflection and rotation. A number of optimality criteria were proposed to explain how such response properties can be linked to the properties of natural scenes and how they can be modelled in artificial networks. Many of these build upon the idea that meaningful events or stimuli are extended in both space and time. For example, the presence of a human lasts usually at least several seconds and covers a certain region of space. Noise and other irrelevant signals, in contrast, vary on much shorter time scales.

![Figure V.2](image)

**Figure V.2.** The concept of stable coding. Top: Image sequence taken from a natural movie clip. Indicated is the presence of a human in the first three frames. Middle: Brightness value of a single pixel chosen from the middle of the images. Bottom: Signal indicating the presence of the human. The brightness of the pixel varies on much shorter timescales than the presence of the human.

The notion that a certain event is extended in space or time can be mathematically expressed in an objective function. Let’s assume a neuron is selective to a feature, which we want to extract. If this feature is present, it will also be present at some point later in time. Thus if the neuron is active at a time point $t$, it should also be active at a later time point $t + dx$. Here $dx$ is an appropriate time constant that depends on the variable and system of interest. Thus, the objective function should take on large values if the derivative of the neurons activity does not change:

$$
\psi_{time} = - \frac{\langle (A(t) - A(t - \Delta t))^2 \rangle_{\text{Stimuli}}}{\text{var}(A(t))_{\text{Stimuli}}}
$$
Here A is the neurons activity, $<>$ denotes the average across time and var is the variance of the neurons activity. The division by the variance does not directly contribute to the temporal coherence, but is necessary to prevent the neurons receptive field from becoming trivial; a receptive field that is zero at every point of course yields the most stable response.

Stable coding, also known as the principle of temporal coherence, was used to show that artificial networks can learn receptive fields that allow invariant responses. Foeldiak (1991) used moving bars as stimuli for a network maximizing temporal coherence to show that the receptive field properties of complex cells can be reproduced. In a similar way invariance response properties were obtained using a spatial coherence criterion (Becker & Hinton 1995). These seminal studies, however, used artificial stimuli such as uniformly moving bars or random dots. Whether such learning schemes perform similarly when operating on complex or natural stimuli is unclear. Furthermore, no detailed comparison of the resulting receptive field properties with those found in biology was made. Thus it is not clear whether such learning schemes could operate under real world conditions and whether the results show a good match to physiological findings.

**V.3 Decorrelation - a form of lateral interaction**

One problem common to many simulations of populations of model neurons is that different neurons tend to develop similar if not identical receptive fields. This problem is very prominent if a whole network is trained to optimise a certain energy function. To overcome this, a form of lateral inhibition between different neurons is required. One common choice of such an interaction is strongly related to the redundancy reduction approach: a decorrelation of the activity of different neurons.

$$\Psi_{decorr} = -\frac{1}{(N-1)^2}\left(\sum_i \sum_{i\neq j}(\sigma_j^2(t))\right)_{stimuli}$$

where

$$\sigma_j = \frac{(A_i-A_i\bar{A})|A_i-A_i\bar{A}|}{\sqrt{\text{var}(A_i)\text{var}(A_i)}}$$
denotes the coefficient of correlation and N, the number of neurons.

Most studies include such a lateral interaction only because it is necessary to overcome the technical problem of identical neurons. However, as shown by our results, adding such a decorrelation can actually have a profound effect on the results (see chapter V.4).

**V.4 Stable coding and simple cells**

Many biological and artificial neural networks require the parallel extraction of multiple features, and meet this requirement with distinct populations of neurons that are selective to
one property of the stimulus while being non-selective to another property. In this way, several populations can resolve a set of features independently of each other, and thus achieve a parallel mode of processing. This raises the question how an initially homogeneous population of neurons segregates into groups with distinct and complementary response properties.

We addressed this issue and investigated a population of neurons with simple linear receptive fields followed by a static non-linearity (see Einhäuser et al., 2003). The activity A of a neuron is given by:

$$A = \varphi \left( \sum_j W_j I_j \right)$$

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

We trained the population of neurons on a set of natural colour movies and optimised a combination of the stability and decorrelation objective functions. In the resulting population of cells two sub-groups can be identified. The cells of the first group have the properties that their receptive fields are chromatic. These cells have the property that they are optimal with respect to the stability objective and sub-optimal for the decorrelation. The cells of the second group are achromatic. These cells are sub-optimal with respect to stability. The important property of these two groups is that the chromatic cells have isotropic receptive fields while the achromatic cells are selective for spatial structures such as orientation and spatial frequency.
In addition, we showed that changing the relative weight of the two objective functions changes the relative fraction of chromatic and achromatic cells. The relative contribution of the decorrelation versus the stability objective determines the proportion of chromatic versus achromatic neurons and thus also the average chromaticity across the network. Therefore, a single parameter is sufficient to determine the proportion of chromatic versus achromatic neurons in our network. The proposed objective thus successfully leads to the segregation of neurons into complementary populations that are either selective for colour or orientation.

The two types of receptive fields developed by the model neurons match the experimental findings in the primary visual cortex. In monkey, both achromatic and chromatic simple cells have been reported. Furthermore, there is physiological evidence that sensitive-sensitive neurons tend to be non-oriented, whereas achromatic neurons tend to exhibit precise orientation tuning (Gouras, 1974; Lennie et al. 1990). This is further supported by psychophysical experiments showing that humans can resolve higher spatial frequencies for isochromatic patterns than for isoluminant patterns (Webster et al. 1990, Sekiguchi et al. 1993).

V.5 Stable coding and complex cells

Simple cells have an approximately linear input output relationship (Carandini et al. 1997). In contrast to this, complex cells are known to have a nonlinear behavior (Hubel & Wiesel 1962, Spitzer & Hochstein 1985). Their responses are invariant with respect to small translations and contrast polarity of the stimulus. As a consequence, a complex cell cannot be modelled using a single linear receptive field. The simplest model capable of representing a complex cell is the two-subunit energy detector (Adelson & Bergen 1985). As indicated by its name, it consists of two linear subunits, which are combined nonlinearly into the response of the model neuron. The activity of a neuron is defined as

$$A = \sqrt{\left(\sum_j W_{1,j}I_j\right)^2 + \left(\sum_j W_{2,j}I_j\right)^2}$$

where I is the input and $W_{1,j}$, $W_{2,j}$ are the weight vectors of the subunits.
Figure V.4. The two-subunit energy detector. Each neuron is described by two weight vectors $W_1$ and $W_2$. The input patch is multiplied with these two weight vectors yielding the activity values for the two subunits. Then, the geometric mean of these two is computed yielding the activity of the neuron.

In order that this cell behaves roughly like a complex cell several conditions must be met: both subunits must be Gabor like receptive fields with the same orientation and spatial frequency. Furthermore, the receptive fields must be shifted with respect to each other by 90 deg.

As motivated above, the temporal smoothness objective functions seems well suited to train a network of two sub-unit energy detectors to obtain complex like response properties. We investigated this thoroughly and compared the learned receptive fields to those measured in physiology (Körding et al. 2003). We found that optimising the stability criterion on a population of subunit energy detectors leads to cells with translation invariant responses. This holds not only for the two-subunit model but also for neurons with four or more subunits. Furthermore, the quantitative properties of the learned receptive fields are a good match to those found in biology.

Figure V.5. Examples of receptive fields after optimising a two-subunit energy detector to maximize the stability objective. All cells are selective to orientation and spatial frequency. For the three cells on the left, the two subunits are phase shifted with respect to each other. In contrast the subunits on the right are not shifted with respect to each other.
To test whether such results can also be achieved using the sparse coding hypothesis alone, we used a combination of sparse coding and decorrelation to optimise the model neurons. However, most cells did not acquire translation invariant response properties and also the match of the orientation selectivity and aspect ratio to the biology was worse. Our results show that the invariant properties of complex cells cannot be understood by an adaptation to a pure sparseness objective. If we however adapt the neurons to be optimally stable, they share the spatial properties of complex cells in the primary visual cortex.

What are the properties of natural stimuli that lead to the emergence of such responses when optimising stability? Within a small part of the visual field oriented structures are frequently observed. Structures with the same orientation often occur at subsequent instances albeit at translated positions (Kayser et al. 2003). A neuron’s response to those stimuli therefore changes less frequent if it is translation invariant. This biases the optimally stable neurons to become translation invariant.

V.6 Learning the non-linearity

Complex cells are just one example of neurons in sensory areas that exhibit nonlinear response properties. However, most studies modelling in sensory areas use a fixed cell model with fixed nonlinearity and only optimise parameters describing a linear receptive field; just as in the above chapters. These studies are successful in explaining the properties of some types of neurons. However, they incorporate a priori knowledge about the cell model and the type of nonlinearity. But to better understand and predict properties of less well-studied systems, where such prior knowledge is not available, models are needed that can adapt the non-linearity of the cell model.

To make a step in this direction, we experimented with a network of visual neurons in which both, a parameter controlling the nonlinearity and the linear receptive field kernels are learned simultaneously (Kayser et al. 2003). As in chapter V.5 we used a two-subunit energy model for the neuron. However, instead of a fixed squaring nonlinearity, we used a variable exponent $N$. The activity of a neuron $I$ is given by:

$$A(t) = \left( \sum_{j} W_{1,j} I_{j} \right)^{N} + \sum_{j} W_{2,j} I_{j}^{N} \right)^{1/N}$$

We then optimised a population of such neurons using the stability objective function. Both, the receptive fields and the exponent $N$ show a smooth learning during the optimisation process and converge towards nontrivial solutions.
The spatial characteristics of the receptive fields are similar to those of complex cells in the visual cortex. The cells are selective to spatial frequency and orientation. Looking at the distribution of the exponent after convergence we find that most cells acquire an exponent close to two. Thus many cells become similar to the classical two-subunit energy detector. However, there is also a population of cells that acquires a very large exponent. Such cells perform a max operation on their two subunits, which is another model that was proposed for complex cells (Riesenhuber & Poggio 1999).

Recent electrophysiological experiments investigated the nonlinearity of cortical cells quantitatively and determined the minimal number of subunits necessary to describe the response properties of complex cells (Lau et al. 2002). This study found that most complex cells could be adequately described using two to four subunits. Furthermore, the distribution of exponents describing the nonlinearity is rather broad with a peak around two. Thus this study provides experimental support for the cell model used in the above simulations and shows that the distribution of exponents in our simulations matches rather well with biology.

On which properties of natural scenes do these results rely? In control simulations we showed that the development of a Gabor like receptive field and a stable exponent requires natural movies. Learning on space-time pink noise yields a much broader distribution of exponents and receptive fields that are selective to spatial frequency but not clearly oriented. If learned from temporally shuffled stimuli, neither the receptive field acquires structure nor does the exponent converge.

V.7 Summary of the results
Receptive field properties of cortical cells can be related to the properties of natural scenes using different criteria for optimal coding. Optimising the responses of model neurons to natural scenes, we showed that complex cells can be understood as forming an optimal temporally stable code. Thus whereas simple cells in V1 and at lower processing stages seem
to be optimal with respect to the reduction of redundancy, complex cells can be obtained based on a different coding principle. Using combinations of different objective functions, we showed how groups of cells with distinct and complementary response properties can emerge from an initially homogenous population. This is crucial to develop networks that extract multiple features in parallel within the same network layer as is done in real biological networks. Last, we developed a learning scheme capable of optimising the neurons linear receptive field together with its output nonlinearity. Such networks are well suited to model the response properties of yet unexplored biological systems of which we have only a limited a priori knowledge which is necessary to specify some model parameters and properties in advance.

V.8 Directions of future work & conclusion

Coding principles like sparse-coding or temporal stability link the properties of natural scenes with properties of neurons in the visual system. Many properties of neurons in the retina, thalamus and of simple cells in V1 have been explained based on the sparseness principle. Here we showed that the response properties of complex cells in V1 can be modelled based on the assumption that the neurons’ responses are optimal with respect to the temporal stability objective. Furthermore, many properties of the learned receptive fields showed a good match with results from physiology. These results suggest that a criterion exploiting the temporal coherence, i.e. the degree of temporal correlation, of a given feature might be well suited to learn detectors selective to this but invariant to other features.

The neuronal response properties at higher processing stages, e.g. in the inferotemporal cortex are highly selective to complex objects and at the same time invariant with respect to a large class of transformations (Tanaka et al. 1991; Kobatake & Tanaka 1994; Logothetis et al. 1995). In order to explain or model such response properties networks with several layers are needed a number of models has been proposed (Fukushima 1988, Riesenhuber & Poggio 1999; Riesenhuber & Poggio 2001). However, few learning schemes exist, that successfully train a network to perform invariant object detection or recognition in an unsupervised way. The temporal stability objective seems well suited for the use in a multi-layer network developing several layers of neurons with increasing invariances.

A problem related to that of training a network to learn invariant responses is that of segregating an initially homogenous group of neurons into groups with complementary response properties. For example in the visual cortex of the monkey some neurons are selective to colour whereas other neurons are achromatic (Gouras, 1974; Lennie et al. 1990). Another example are neurons in area PMLS of the cat. While some neurons seem to extract the optical flow with respect to a coordinate system based on the fovea, other neurons do so with respect to the receptive field centre (Merabet et al. 2000; Brosseau-Lachaine et al. 2001). This is but few examples. However, the question arises how an initially homogenous
population of neurons, that all change their synaptic connectivity based on the same learning rule, can be split into several groups, each being sensitive to one of these features and possibly invariant to the other. Our results indicate that using a combination of several objective functions can achieve such separation. One of the objective functions leads the neurons to develop a specific selectivity, while a second objective function prevents some of the neurons from becoming optimal with respect to the first objective. The question which population then becomes selective to which feature then has its answer in the statistics of the natural stimuli. The temporal stability objective, as was used in these experiments, extracts the most slowly varying non-trivial feature from a set of stimuli. As suggested by ongoing analysis of natural movies, colour is a feature varying on very slow time scales. Thus a cell that is trained on natural colour movies will become selective to colour and invariant to features varying on faster scales, e.g. orientation or position. If the cell on the other hand is prevented from becoming truly optimal with respect to the stability, e.g. by a second objective function, it will not become selective to the colour, but to a feature varying on a faster time scale, i.e. the orientation. Thus, parallel extraction of complementary features using combinations of objective functions forms a close link between the statistical properties of natural scenes and the selectivity of visual cortical cells.

In order to model cortical receptive fields, even in a network modified according to some learning scheme, many parameters need to be specified in advance according to a priori knowledge. Taking the subunit energy detector models used above as an example, these parameters are: the number of subunits, their size, the transfer function applied to the activity of each subunit, and the way how the outputs of different subunits are combined. For some types of neurons, like the simple and complex cells in V1, we have a great deal of a priori knowledge that can be used to fix many of these parameters (Anzai et al. 1999; Lau et al. 2002). However, for many other sensory systems we lack this detailed knowledge although nonlinearities are ubiquitous in sensory systems. For example the integration of the information about different whiskers in the somatosensory thalamus is nonlinear and the degree of nonlinearity depends on the strength of the cortical feedback from S1 (Ghazanfar et al., 2001). The neurons in S1 in turn can also show nonlinear interactions of the signals from several whiskers depending on the timing of the movement of individual whiskers (Shimegi et al., 1999). In the auditory nuclei of the inferior colliculus, the responses of many neurons can be described based on the linear integration of spectro-temporal energy. However the responses of a distinct subpopulation fail to follow such linear integration (Escabi & Schreiner 2002). Another example is given by the rod bipolar cells in the retina. Their output shows a supralinear dependence on the stimulus intensity (Field & Rieke, 2002). In order to model less well understood systems, we either need to chose many of the parameters based on a trial and error scheme, or we need networks that require less many a priori’s. Our results on the learning of the nonlinearity of the transfer function make a step into this direction.

Our results describe how several new features of the processing in V1 can be replicated in artificial networks learning from natural stimuli. However, there are many further response
properties of visual cortical neurons that have not been taken into account. For example, contrast normalization has been often reported in V1 (Heeger 1992) and is one of the features incorporated in many models of visual neurons (Heeger 1993; Simoncelli & Olshausen 2001). Recent results show that this phenomenon can be linked to the principles of optimal coding and a model was devised in which the strength of the normalization terms was optimised during presentation of natural images Schwartz & Simoncelli 2001). Thus also the parameters of more complex and sophisticated models for visual neurons can be derived from the statistical properties of natural stimuli. Summarizing we find that the different optimality criteria like sparseness or stability offer interesting ways to understand and model the response properties of sensory neurons based on the statistical properties of their natural input.

V.9 Publications

The results presented in this chapter are published in the following papers:

References


Mach E (1886) The analysis of sensations and the relation of the physical to the psychical. (Translation of the German original) New York: Dover.


Acknowledgements

This work would not have been possible without the inspirations and the help of many people. Especially, I would like to thank Markus Siegel and Konrad Körding for introducing me to the field of neuroscience. Many thanks go to Mina Kim, Gudrun Möller and Rodrigo Salazar, who spent many hours with me in the lab collecting data, and to Wolfgang Einhäuser with whom I did parts of the simulation work. Not to forget all other people at INI providing help and ideas: Jörg Conradt, Nuno DaCosta, Jörg Hipp and Selim Onat. Next, I would like to thank the people at the Centre for Magnetic Resonance Research in Minneapolis for their hospitality and for the best espresso west of Mississippi! Last but not least, the following people provided constructive contributions in one or the other respect: Tansu Celikel, Matteo Carandini, Adrienne Fairhall, Jack Gallant, Jakob S. Horn, Bruno Olshausen, Rasmus Petersen, several anonymous referees and Dario Ringach.

This work arose from many discussions with my supervisor Peter König. I profited much from his way of heading his group and enjoyed the freedom I had. Furthermore, I’m grateful to Dae-Shik Kim for introducing me to the field of fMRI and for his supervision during this project.

Special thanks go to Mathew Diamond and Kevan Martin for being referees of this thesis!

My salary was provided by a PhD grant from the Neuroscience Center Zurich. I would like to thank them for their confidence and support. Both the experimental and the theoretical work were supported by a number of sources: The Swiss National Science Foundation, the EU Amouse project, the National Institutes of Health, the Keck Foundation and the Human Frontiers Science Program.
Curriculum vitae

Christoph Kayser
Luegislandstrasse 163, 8051 Zurich, Switzerland, kayser@ini.phys.ethz.ch

Born 18.01.1977, Ulm, Germany

Education
1995   University entrance qualification: Matura (Type B), Basel.

Research experience
2002/2003 Research visits at the Center for Magnetic Resonance Research, University of Minnesota, USA. Functional MRI experiments on the visual system.

Teaching experience

Academic training
2002:   Course on surgery and anesthesia on laboratory animals.
2001-2003 PhD training program in molecular, clinical and systems Neuroscience.
2001:   Course on handling and use of laboratory animals.
2001:   Summer school on the ‘Development of Intelligent systems’. Antonio Borsellino College on Neurophysics, Trieste, Italy.
2000:   Summer school of the “Schweizerische Studienstiftung” on ‘what is nature?’

Languages
German (native), English (fluent), French (basic conversation), Spanish (understanding)

Research Skills
• All aspects of in vivo electrophysiological experiments. Sterile surgery, anesthesia, chronic implantations, fabrication of microelectrodes, multi-channel recordings. Recordings in anaesthetized and in awake behaving subjects. Training of animals. Design of recording setup.
• Histology and common staining techniques.
• Using eye-tracking devices with human and animal subjects.
• Design of and conduct of fMRI experiments; fMRI data analysis and visualization.
• Statistical analysis of image and video data.
• Modeling of neural networks.
**IT Skills**
- Extensive experience using and maintaining all major operating systems (Windows, Linux, Mac).
- Development of routines for data management, analysis and visualization in different programming languages; Matlab, Mathematica, C/C++.
- Long-time experience with standard (office) software (Acrobat, Illustrator, Photoshop, Premiere; Canvas; Microsoft Office; HTML; Latex).

**Grants and financial support**
PhD Grant from the Center of Neuroscience (ZNZ) Zurich (2001-2004); Travel grant from the European Neural Network Society (2001).

**Memberships and Affiliations**
Member of the American Society of Neuroscience, the Swiss Society of Neuroscience, the European Neural Network Society and the Neuroscience Center Zurich.
Appendix – Publications
Publications discussed in chapter 2

Temporal correlations of orientations in natural scenes

Christoph Kayser*, Wolfgang Einhäuser, Peter König

Institute of Neuroinformatics, ETH/UNI Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Abstract

The visual system performs complicated operations such as visual grouping efficiently on its natural input. To study this adaptation to natural stimuli we measure spatio-temporal interactions of orientations in scenes with natural temporal structure recorded using a camera mounted to a cat’s head. We find long range spatial and long lasting temporal correlations of orientations with collinear interactions being most prevalent and preserved over time. The spatial extent of correlations corresponds to the length of horizontal cortical connections and the temporal duration of the interactions allows co-activation of lateral and bottom up input by the same visual event. © 2003 Elsevier Science B.V. All rights reserved.

Keywords: Natural scenes; Orientations; Gestalt principles; Image statistics; Temporal coherence

1. Introduction

In recent years processing of natural stimuli by the visual system received increased attention (cf. Ref. [12]). Indeed it was found that early stages of visual processing are specifically adapted to the structure of natural scenes [1,4]. Furthermore, laws for object perception and visual grouping, the Gestalt rules [8,14], can be linked to the statistics of natural scenes. As an example the law of good continuation, favouring collinear arrangements of orientations over parallel, was shown to have a counterpart in the interaction of orientations in still images [7,9,11]. Similar interactions of orientations are also found in contextual effects in psychophysical experiments [6,10], in surround interactions in V1 receptive fields [6] and in lateral connections in V1 [2,5]. Therefore it is of particular interest to link them to properties of natural scenes.

* Corresponding author. Tel.: +41-1-6353044; fax: +41-1-6353053.
E-mail address: kayser@ini.phys.ethz.ch (C. Kayser).
To our knowledge, however, up to now correlations in natural scenes have only been investigated in still images. This neglects the temporal structure and it remains unclear whether these correlations persist on time scales relevant for lateral interactions in the cortex. Given the possibly long delays for tangential connections, correlations must extend over substantial temporal periods in order to fully cover the spatial extent of long-range connections. Furthermore, some of the previous studies did not report filter or correlation scales in units of degrees of visual angle leaving possible links to anatomical scales uncertain. Finally, some of the previous studies used still images captured by humans, therefore introducing a possible artistic or anthropocentric bias.

Here we address these issues and study spatio-temporal interactions of orientations in a large database of natural movies captured by a camera mounted to a cat’s head.

2. Methods

We recorded movie sequences using a removable lightweight CCD-camera (Conrad electronics, Hirschau, Germany) mounted to the head of cats while taken for walks in different local environments like grassland, forest and the university campus. These videos contain a large variety of different speeds and accelerations as a result of the natural movements of the cat. Fig. 1 shows four sample images of our database. For this study a total of three animals was used and all procedures are in agreement with national and institutional guidelines for animal care.

Videos were recorded via a cable connected to the leash onto a standard VHS-VCR (Pal) carried by the human experimenter and digitised offline at a temporal resolution of 25 Hz, $320 \times 240$ pixels (1 pixels $\approx 12$ min of arc) and 16 bit color depth. For this study videos were converted to 8-bit gray scale and 12 sequences (about 40,000 frames total) were used. Before further processing the images were normalized to zero mean.

The image statistics was investigated using oriented wavelets. Single frames were convolved with pairs of circular Gabor wavelets of $90^\circ$ relative phase shift. Filters had an envelope of 20 pixel width and a spatial frequency of 7 (1/pixels). The amplitude of the orientation was computed by summing the squared amplitudes of two phase

Fig. 1. Four sample frames of our database are shown on the left. The amplitudes of the oriented energy detectors for the same frames is shown on the right. The bars indicate the orientation of the respective filters used.
shifted filters and subjecting the result to a square root, resembling a two subunit energy model. At each point the amplitudes of eight equally spaced orientations from $0^\circ$ (horizontal) to $157.5^\circ$ were computed. We define the ‘prominent’ orientation of each point by averaging the amplitude vectors (length = amplitude of filter response, orientation = orientation of the filter) of the eight filters. The resulting vector average has an orientation $\Theta$, defining the prominent orientation of the point, and a length $A(\Theta,x,t)$, specifying the magnitude of the local orientation strength. For computational convenience these orientations $\Theta$ were binned into 16 bin between $0^\circ$ and $180^\circ$. The second order statistics of these orientations was calculated assuming translation invariance of natural images. Thus correlations of two prominent orientations $\Theta_1$ and $\Theta_2$ were computed over all pairs of points with the same spatial separation $\Delta x$ and temporal separation $\Delta t$ (the mean $\langle \rangle$ runs over all points $(x,t)$ with prominent orientation $\Theta_1$).

$$C(\Theta_1, \Theta_2, \Delta x, \Delta t) = \frac{(A(\Theta_1,x,t) - \langle A(\Theta_1,x,t) \rangle) \cdot (A(\Theta_2,x+\Delta x,t+\Delta t) - \langle A(\Theta_2,x+\Delta x,t+\Delta t) \rangle)}{\sqrt{\langle (A(\Theta_1,x,t) - \langle A(\Theta_1,x,t) \rangle)^2 \rangle} \cdot \sqrt{\langle (A(\Theta_2,x+\Delta x,t+\Delta t) - \langle A(\Theta_2,x+\Delta x,t+\Delta t) \rangle)^2 \rangle}}.$$  

Correlations were computed for temporal lags from $\Delta t = 0$ to 30 frames (1.2 s) and on a spatial grid of points spaced about $2^\circ$ apart. Therefore the kernels overlapped only for the smallest spatial distance used. As a control we also computed correlations using the maximally active orientation at each point instead of the ‘prominent’ orientation yielding similar results as reported below.

### 3. Results

First we investigate temporal correlations at the same point in space. Fig. 2A demonstrates that if an orientation is present at one point in time then the amplitude of this orientation in the next frames at the same point is also likely to be high. Temporal correlations are strongest for the cardinal orientations, i.e. horizontal and vertical. For the other orientations correlations decay faster but are still significant over several hundreds of milliseconds (decay time constants for $0^\circ$: $> 1$ s, $45^\circ$: 490 ms, $90^\circ$: 900 ms, $135^\circ$: 360 ms). Thus the presence of an oriented segment gives a strong prediction for the orientation at the same point later in time.

Next we look at the two dimensional spatial distribution of correlations as well as correlations of different orientations. Fig. 2B shows the correlations between segments of 4 different orientations ($0^\circ$, $45^\circ$, $90^\circ$, $135^\circ$) situated at different relative locations in the same frame. Iso-orientation correlations (panels on the diagonal) are stronger than cross-orientation correlations. Furthermore the contour lines of the iso-orientation correlations are elongated along the direction of the particular orientation. This shows that collinear structures are more prevalent than parallel shifted contours. Also parallel contours occur more likely than T-junctions since the iso-orientation correlations are at all points stronger than the correlations of this orientation with the orthogonal. An example of how the spatial correlations decay independently of the spatial direction is
Fig. 2. (A) The correlation of orientation amplitude over time at the same pixel. Squares: 0°, stars: 90°, dashed: all other orientations (spaced 22.5°). (B) Correlations of different combinations of orientations and different spatial arrangements of the two points in the same frame. The orientations are (from top to bottom and left to right): 0°, 135°, 90°, 45°. (C) Correlations over time of points with prominent horizontal orientation but which are spatially separated by different distances independent of the relative orientation. Squares: 2.1 deg spatial distance, stars: 4.2°, dashed: 6.4°, diamonds: 8.4°. (D) Same as in B but here the two points are also separated by 400 ms in time.

shown in Fig. 2C for the horizontal (90°) orientation. Correlations decay fastest during the first 2° of spatial distance but extend well up to 8°.

Our data set allows analyzing how these spatial correlations evolve over time. Fig. 2D shows the same data as in Fig. 2B but for segments 400 ms apart in time. The spatial arrangement of correlations is the same as for zero time lag but the amplitudes decayed by a factor higher than 2. For the cardinal orientations again collinear interactions are prevalent. This is in agreement with Fig. 2A which shows that these orientations are very stable over time. Since the oblique orientations are less well correlated over time we would expect that collinearity will here be less prominent for larger time lags. Indeed the contour lines of the correlations for the oblique orientations are more circular symmetric. To quantify these changes over time we measure the aspect ratio (length/width) of the contour lines for the different time lags. Collinearity means a high aspect ratio and a loss of collinearity therefore is accompanied with a decrease
Fig. 3. (A) Relative change of the aspect ratio of the correlation contours in Fig. 2C as a function of time. Shows is the aspect ratio at each point in time divided by the aspect ratio at $t = 0$. Squares: 0°, stars: 90°, solid: 45°, dashed: 135°. (B) Areas of strong correlations. We defined spatio temporal separations with a correlation over 0.4 as strong. The figure shows these areas for the correlation diagram of Fig. 2C. (C) Shows the size of these areas relative to the total patch size over time. Lines are labeled as in A.

in aspect ratio. Using this measure, Fig. 3A shows that collinearity is preserved over long temporal lags and is strongest for the cardinal orientations.

To quantify the change in amplitude of the spatial correlations in a different way, we define areas of strong interactions by thresholding correlations. We chose a threshold of 0.4 to ensure that even for zero time-lag only iso-orientation correlations exceed this threshold (Fig. 3B). As expected from Fig. 2 the decay times are slowest for the cardinal orientations but independent of the orientation there exist points with strong correlations for at least 280 ms (Fig. 3C).

We performed controls to see how these results depend on the amount of data used. The above data were averaged over our whole database. Since one feature of our video sequences is their variety in terms of landscapes, etc. we look at the differences between different sequences. In Fig. 4 we show the correlations for one oblique orientation (135°). The mean and standard deviation over 12 video sequences is shown in Fig. 4A. The error is rather small compared to the correlation values. More importantly, the correlation surface plus minus the error (Fig. 4B) shows the same spatial structure as the mean. Also, the distinct pattern of correlations is visible in averages over shorter sequences (data not shown). Thus the distinct patterns of spatial correlations are not introduced by averaging over a large data set.

As a further control, we use filters of a different spatial scale and frequency to measure the orientation content. The filters used for Fig. 4C are twice as large as the ones used for the other experiments. The results are basically the same as with the lower frequency filters. Again collinearity is most prevalent. Therefore our results generalize over a wide range of filter parameters.
4. Discussion

We recorded natural image sequences from a camera mounted to a cat’s head closely matching the animal’s visual input. Thereby our database circumvents possible artistic or anthropocentric biases introduced in pictures and movies taken by humans. The database contains a large set of different environments, ranging from forest to grasslands and university campus. Furthermore the used sequences were recorded in different seasons and times of day providing a huge variety of lighting conditions. In respect to the temporal analysis it is worth noting that our video sequences contain natural movements of an animal, which might differ considerably from e.g. commercial movies filmed by humans.

In qualitative agreement with previous studies [7,9,11] we find spatial correlations corresponding to the Gestalt laws. For all orientations collinear contours are more prevalent than parallel contours and correlations between orthogonal orientations are weakest. However we find correlations over distances of up to 8 degrees of visual angle (Fig. 2). This is considerably larger than distances reported in previous studies. For example Kaschube et al. [7] find that already for small distances correlations are relatively weak (< 0.15 in a range from 1° to 4°). However, they do not indicate the size of their kernels in the same units. Sigman et al. [11] report similar correlations using filters of size smaller than 10 min of arc. Our higher correlation could be due to methodological differences to other studies besides the use of different and possibly larger kernels. We computed the ‘prominent’ orientation of a point by vector averaging the outputs of 8 oriented energy detectors. But correlations computed on these prominent orientations are very similar to correlations computed on the maximally active orientation (data not shown) a method used in Ref. [11].

The spatial distances of the correlations reported here fit well with anatomical data on long-range horizontal connections in primary visual cortex. In cat V1 8° of visual
angle correspond roughly to 8 mm [13]. This is also the extent of long-range connections which preferentially connect iso-orientation domains [5] and in some mammals preferentially mediate collinear interactions [2].

In the temporal domain we find long lasting correlations of orientations to extend several hundreds of milliseconds preserving their spatial structure i.e. collinearity. These persist sufficiently long to allow bottom up and long range lateral input to be coactive and driven by the same orientated structure even given the slow speeds of lateral connections reported in Ref. [3]. Therefore the spatio-temporal interactions of orientations seem to fully cover the range of tangential connections and provide a substrate that could also guide the development of orientation maps and long-range connections in primary visual cortex.

Acknowledgements

This work was financially supported by the Centre of Neuroscience Zurich, (ZNZ), Honda R& D Europe (Germany) and the Swiss national fund (SNF grant No. 31-65415.01).

References

Publications discussed in chapter 3


Responses to natural scenes in cat V1

Christoph Kayser, Rodrigo F. Salazar & Peter König
Institute of Neuroinformatics, University & ETH Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; email: {kayser, salazar, peterk}@ini.phys.ethz.ch

Correspondence should be addressed to C.K. {kayser@ini.phys.ethz.ch, Phone: +41-1-6353044, Fax: +41-1-6353053}

Running head: Processing of natural scenes in V1
Abstract

Studies on processing in primary visual areas often use artificial stimuli such as bars or gratings. As a result, little is known about the properties of activity patterns for the natural stimuli processed by the visual system on a daily basis. Furthermore, in the cat, a well-studied model system for visual processing, most results are obtained from anaesthetized subjects and little is known about neuronal activations in the alert animal. Addressing these issues, we measure local field potentials (lfp) and multi unit spikes in the primary visual cortex of awake cats. We compare changes in the lfp power spectra and multi-unit firing rates for natural movies, movies with modified spatio-temporal correlations as well as gratings. The activity patterns elicited by drifting gratings are qualitatively and quantitatively different from those elicited by natural stimuli and this difference arises from both spatial as well as temporal properties of the stimuli. Furthermore, both local field potentials and multi-unit firing rates are most sensitive to the second order statistics of the stimuli and not to their higher order properties. Last, responses to natural movies show a large variability over time due to activity fluctuations induced by rapid stimulus motion. We show that these fluctuations are not dependent on the detailed spatial properties of the stimuli but depend on their temporal jitter. These fluctuations are important characteristics of visual activity under natural conditions and impose limitations on the readout of possible differences in mean activity levels.
Introduction

How are sensory stimuli represented and processed in cortical networks? Which stimulus properties determine the neural activity at a given processing stage and which properties of an activity pattern are relevant for the representation of the stimulus? These questions have been central to neuroscience research for a long time. However, we do not have a conclusive answer at hand. In cat visual cortex, one of the best-studied sensory systems, most of our knowledge is based on the activity of single neurons recorded in anaesthetized subjects stimulate with artificial stimuli. This experimental paradigm, however, has several limitations.

First, most experiments recording in cat primary visual cortex (V1) use artificial stimuli such as bars, gratings or texture fields. While these are mathematically well defined and easy to generate, they contrast with the natural scenes encountered by a mammalian visual system on a daily basis. Natural stimuli are very complex and their statistical structure differs markedly from that of the stimuli used in most experiments. It is presently unclear whether results obtained using artificial stimuli can be extrapolated to the processing of natural scenes. Often, the stimuli are stationary or smoothly varying and their parameters, such as the orientation of a grating, are matched to the preferred properties of the neuron recorded. The response properties are then inferred from the steady state responses to these stimuli. Under natural conditions, however, stimuli are not optimal for most neurons and the visual system cannot wait for a steady state response, but has to use whatever activity pattern the stimuli elicit to build a representation of the visual scene.

Second, the anesthesia used in many experiments allows convenient electrophysiological experiments and prolongs the time available for recording. However, anesthetics can have profound influences on the response properties (Roberston 1965; Lee 1970; Lamme et al. 1998) and especially in the cat visual system little is known about activities in awake animals.

Third, most experiments record spikes from isolated single cells, usually one or a few cells at a time. For statistical analysis, data are not averaged across a large population of neurons, but over repeated trials with identical stimulation. This is in contrast to the
situation of a single neuron: it receives input from a large population and does not average over repeated trials, but enacts its input-output transformation continuously. The statistical properties of the activity in a population of neurons might, however, differ from those of the particular single unit under investigation.

The goal of the present study is to advance our understanding of the processing of natural scenes in awake animals. To do so we record population activities in the primary visual cortex of alert cats to natural movies. More specific, we measure stimulus induced changes in the power spectrum of local field potentials and firing rates of multi unit spikes. While the multi unit spikes reflect the activity of a small number of nearby neurons, the local field potential includes both pre and post synaptic potentials around the electrode tip. The natural movies used as stimuli were captured by a camera mounted on the head of a freely moving cat exploring different environments. These movies resemble the world as seen from a cats’ perspective and are a good approximation to the animal’s natural visual input during every day vision. To quantify the impact of different stimulus properties on the activity we compare responses to modified stimuli with altered statistical properties and altered motion. Last, we quantify the temporal fluctuations imposed by the natural movies onto the activity pattern.

**Materials and Methods**

*Surgical and recording procedures*

Recordings were performed in four adult female cats. In each animal a micro drive was implanted under aseptic conditions. The animals were initially anaesthetized using ketamine hydrochloride (Narketan, Chassot, Bern Switzerland; 20 mg/kg) and xylazine hydrochloride (Rompun, Bayer, Leverkusen, Germany; 1.1 mg/kg). They were intubated, ventilated (30 % O2 and 70 % NO2) and continuously anaesthetized with Isoflurane (0.4% – 1.5%). Body temperature, endexpiratory CO₂ and blood oxygenation were continuously monitored and kept in the desired physiological range (37-38 deg Celsius, 3%-4%, 90%-100% respectively). The animals were infused with ringer-lactate solution (40 ml/h) and received i.m. injections of steroids and analgesics. 7-8 titanium screws
were fixed in the skull to later hold the implant. Two small holes were drilled and reference and ground electrodes were placed between dura and bone. Two small craniotomies (roughly 4mm diameter) were made over areas 17/18 and 21 of one hemisphere according to stereotaxic coordinates (AP: -3, L: +2 and L: +8 respectively). After removing the dura the micro drive was positioned and fixed to the skull and the screws using dental acrylic (Stoelting Physiology, USA). The cavity was filled with Silicon oil. Nuts, later used to restrain the animal in the recording setup, were inserted into the implant and fixed in the acrylic. Recording sessions only begun after the animal had fully recovered, usually after four days.

Each micro drive contained 4 movable electrodes (500 – 1000 kOhm impedance), two of which were placed over the primary visual area 18 and two over area 21. The present project concentrates on the analysis of neuronal activity in primary visual cortex only. Signals were first passed through a 24-channel preamplifier (Neurotrack, 10x amplification) and finally amplified and digitized using a Synamps system (Neuroscan, El Paso, USA) at a resolution of 20’000 Hz.

Recordings were made at sites of different depths. The depth was estimated from the reading of the micromanipulator and the first and last site of a penetration where any visual stimulus evoked significant activity. The activity measures used in this project, i.e. spectrograms of local field potential power, yielded qualitatively similar results at recording sites of different depth. Furthermore, there was no qualitative change in the power spectrum of the local field potential between supra and infra granular sites (see also Results). Therefore most results were averaged across all recording sites within one subject.

For recordings the animals were placed in a sleeve equipped with adjustable velcro fasteners. This served the purpose to restrain the animal and to provide a comfortable position. This sleeve was placed in an acrylic tube allowing stable and accurate positioning of the animal in front of the monitor and within a Faraday cage. To ensure a stable visual stimulation the head of the animal was fixed using screws inserted into the chronic implant holding the micro drive. Each recording session lasted roughly 15 minutes and each animal performed one or two sessions a day. We regularly checked the state of alertness of the subject either by direct visual inspection or using an infrared
camera system. All procedures were in accordance with the national guidelines for use of experimental animals and conformed to the National Institutes of Health and Society for Neuroscience (U.S.) regulations.

**Visual stimuli**

To investigate activity pattern for natural time varying stimuli we used a set of movies closely resembling the visual input to the cats’ eye under natural conditions. To furthermore determine the effect of specific properties of natural movies such as motion or higher order statistics on the activity, we used a set of modified movies.

Natural scenes differ from classical lab stimuli in both their spatial as well as temporal properties. A uniformly drifting sine wave grating for example is characterized by a single spatial as well as a single temporal frequency. Time varying natural scenes on the other hand contain a wide range of spatial and temporal frequencies. The contrast of the different frequencies can be computed from the amplitudes of the stimulus’ Fourier spectrum. Natural scenes have a characteristic amplitude distribution of both, spatial and temporal frequencies. The properties determined by the amplitudes of the Fourier spectrum are also know as the second order structure, or second order correlations. The phases of the Fourier spectrum characterize the alignment of the different frequencies, and determine the higher order structure inherent to natural scenes. Artificial images constructed from the amplitude spectrum of a natural image but with random phases of the different frequencies have a quite different appearance compared to the original image from which the amplitude spectrum was taken. Such images, known as pink noise, have a foggy like appearance lacking any visible object due to their random higher order statistics.

We applied phase randomizations to natural movies in both the spatial and temporal domain simultaneously. The obtained stimuli have the same spatio-temporal power spectrum and thus the same spatial and temporal frequency distribution but lack the higher order structure of the original natural movies. The same principle, using an original stimulus and one with altered higher order structure, was also applied to natural movies filtered with Gabor wavelets. These stimuli, which are based on a reconstruction of a natural movie from a wavelet representation,
have a reduced content of spatial frequencies. The corresponding manipulated movies have the property, that the alignment of different wavelets defining the stimuli is changed. In this way the local contrast edges defined by the Gabor wavelets are left unchanged but their global alignment is randomized.

The following gives a detailed presentation of the stimuli used in the present study (Fig.2, left panel):

1) **Square and sine wave gratings (Fig. 2A).** Gratings were oriented either horizontally or vertically. The parameters of the sine wave grating (spatial frequency: 0.2 cycles/degree; temporal frequency: 4 Hz) were chosen to elicit strong responses in the recorded local field potentials. These parameters match the tuning properties of single units in area 18 (Movshon et al. 1978). The properties of the square-wave grating (1.2 degree width of a bar; drifting at 6.6 degrees/sec) were adapted to the statistical properties of the natural stimuli.

2) **Natural movies (Fig. 2B).** These were recorded from a camera mounted to a cats’ head while the animal was exploring different local environments such as forests and meadows (for details see Kayser et al. 2003). Thus these videos incorporate the specific body and head movements of a cat. However, they do not include the animals’ eye movements. Given that cats move their eyes infrequently and slower compared to primates (Crommelinck & Roucoux 1976; Evinger & Fuchs 1978; Möller et al. 2002) this does prevent these movies from closely reproducing the spatial as well as temporal structure of a natural visual stimulus.

3) **Pink pixel noise (Fig. 2C).** For each natural movie we created a stimulus with the same second order distribution of spatial and temporal frequencies but random higher order correlations. This was done by computing the space-time Fourier transform over all movie frames and replacing the phase at each frequency by a random value between 0 and $2\pi$. The inverse Fourier transform was applied to obtain the new stimulus. In total, three pixel noise stimuli were used, each constructed from of the three natural movies.

4) **Wavelet filtered movies (Fig. 2D).** These stimuli were constructed from a natural movie by applying a set of Gabor filters. The set of filters contained six equally spaced orientations, three spatial frequencies (0.6, 1.25, 2.4 cycles/degree) and a bandwidth of 1.1 octaves. For each frame of a video the amplitudes of all filters were computed at the
critical spatial resolution to allow reconstruction of the original frame (Mallat 1989). Then each frame was reconstructed from these amplitudes. Applying these Gabor filters effectively corresponds to a band pass filter in the spatial domain. As before, our stimulus set consisted of three wavelet-filtered movies, each obtained from one of the natural movies.

5) **Pink wavelet noise (Fig. 2E).** Similarly to the wavelet filtered movie, this stimulus was reconstructed from the wavelet amplitudes obtained from the natural movies. However, before reconstruction, the relative alignment of different wavelets in space and time was altered by eliminating their higher order correlations. The wavelet amplitudes computed from a natural movie form a three dimensional (two spatial and one temporal) matrix. In total there are 6 (orientations) * 3 (frequencies) wavelets and thus 18 such matrices for a movie. For each of these matrices the space-time Fourier transform was computed and the phases at each frequency were replaced by random numbers equally distributed between 0 and 2 pi. The matrices were then transformed back. The wavelet noise stimulus was obtained by reconstruction from these manipulated wavelet amplitudes.

Stimuli were presented in a block design. In each session one of three possible blocks was chosen randomly. Each block contained all 5 stimulus types listed above: The sinusoidal and the square wave grating of either horizontal or vertical orientation (chosen randomly), i.e. 2 stimuli of type 1; all three clips of natural movies, i.e. 3 stimuli of type 2; each modification based on one natural movie, i.e. one stimulus of types 3, 4 & 5 each. This resulted in a total of 8 stimuli within a block. Each stimulus lasted 2 seconds. The stimuli were separated by a uniform screen (blank) having the same mean luminance as the stimuli, and also lasting 2 seconds. Each block was repeated 30 times within one recording session. Stimuli were presented full screen on a 19” Hitachi CRT monitor (120 Hz refresh rate) 50 cm in front of the animal, thus covering 40 by 30 degrees of visual angle. The room was otherwise darkened. The color lookup table was manipulated to obtain a linear transformation between pixel intensity and luminance on the monitor. This was verified with a photometer (J1800 Luma Color, Tektronix Inc., Wilsonville, USA) under monitor settings with 30 cd/m2 mean luminance. For recordings, we used a somewhat lower mean luminance value and a monitor radiation shield (3M, Switzerland)
resulting in a mean luminance of 8 cd/m². However, we would like to point out that the effective gamma value is of no great importance, since the different stimuli have a similar distribution of pixel intensities. The stimulus presentation was controlled by a Macintosh computer (Apple, Cupertino, USA) running custom written MATLAB (Mathworks, Natick, USA) code based on the psychophysics toolbox extensions (Brainard 1997; Pelli 1997).

**Motion altered stimuli.** In a second series of recordings we used stimuli with altered motion properties. The flow field describing the motion in a stimulus movie was estimated as described below. This flow field was decomposed into two components. First, a linear drift obtained as the linear component of the flow field. Second, the residual, after the linear drift was removed from the flow field, was termed *jitter*. These two components of the flow field were used to generate a set of stimuli. A single frame taken from a natural movie and a grating patch were used as the base images. The motion pattern imposed on the two base images matched either the linear drift or the jitter component of the flow field. This gives a total of four motion-altered stimuli.

All stimuli were encoded as a series of 8 bit gray scale frames at a resolution of 640 by 480 pixels. The pixel values in each frame were additively scaled so that each frame had the same mean intensity and multiplicatively scaled so that each frame had the same root mean square contrast.

**Estimation of flow fields**

The flow field characterizing the motion was measured on a frame by frame basis using standard techniques (see e.g. Beauchemin and Barron 1995): The optical flow was measured at each point of a grid covering the central 20 degrees of the visual field. A patch of 30 by 30 pixels centered on the grid point was compared to patches at different positions in a range of 70 pixels in each direction in the next video frame. The comparison was based on the mean square difference after removal of the overall mean luminance of each patch. The best match defined the local optical flow. The global motion vector of the video was computed from this locally defined flow field as the arithmetic average.
Data analysis

Separate analysis was carried out for the local field potentials and multiunit activity, both implemented in MATLAB. The local field potential was extracted by low pass filtering the recorded signals below 500 Hz. Each recording session was cut into single trials, with a trial consisting of one stimulus plus a part of the blank before and after it. Sometimes the signals were contaminated by movement artifacts. If in any channel the maximal amplitude of the recorded signals exceeded five times the standard deviation of the signal, taken over the entire session, the trial was discarded. Usually less than 5% of the trials had to be discarded. The activity in the local field potential was quantified by computing the time-localized Fourier spectrum (spectrogram) using windows of 160 ms length, overlaid with a Hanning window and zero-padded to 256 ms. The overlap of neighboring windows was 152 ms, leading to a nominal temporal resolution of 8 ms.

To determine the activity induced by the visual stimulus, changes in the power spectrum locked to the stimulus onset were computed. The power at each frequency was normalized by the standard deviation of the power during 600 ms of the leading blank. This emphasizes changes in the power compared to the blank at each frequency separately. The resulting measure corresponds to a z score and is a measure of the reliability of neuronal activation by the respective stimulus (de Oliviera et al. 1997; Logothetis et al. 2001). Qualitatively similar results were obtained by computing the percent change of the signal amplitude at each frequency before the stimulus to during the stimulus (data not shown).

For the analysis of the multiunit activity, the signals were high pass filtered at 500 Hz. Spikes were identified by applying a threshold of 3 standard deviations of the signal and stored at a resolution of 1 ms. Visual inspection showed the reliability of this automated measure. Units that did not show a modulation for any stimulus of at least a factor of two compared to the blank (i.e. spontaneous activity) were discarded.
Results

The effectiveness of different types of stimuli

A total of 165 sessions were recorded in four animals (cat L: 55, cat S: 49, cat F: 42, cat M: 19 sessions). An example of raw data from the local field potential during presentation of a natural movie is shown in Fig. 1A. Evidently, the recorded signal was modulated by the stimulus. The spectrogram resulting from these data is shown in Fig. 1B. Following stimulus onset an increase in power at most frequencies is observed. To emphasize the activity elicited by the stimulus compared to the leading blank, each frequency axis was normalized by its standard deviation during the blank (Fig. 1C, see methods). This normalized spectrogram reveals that the strongest increase of power does not occur in the gamma range (30-80 Hz) but at frequencies above 80 Hz. A decrease is observed below frequencies of 20 Hz.

Local field potential activity for different stimuli

The following section compares the activity elicited by the different stimuli in one subject in detail.

Figure 2 summarizes the local field potential activity averaged across recording sites (n=48) at different depths. Common to all stimuli is evoked activity, characterized by an increase in power at all frequencies, just following the stimulus onset. In the following this evoked activity is ignored and the analysis concentrates on the ‘steady state’ activity after the onset response. Figure 2A shows the activity elicited by the gratings. It reaches a steady state after about 150 ms characterized by an increase in power in the range of 30-60 Hz, a moderate increase in the range of 80-130 Hz and decrease below 20 Hz. The response is stationary and can be well summarized by averaging the 2 dimensional spectrogram over time: We define a modulation curve by averaging the spectrogram over the interval of 400 - 1900 ms. This modulation curve shown on the right of the spectrogram, quantifies the relative contribution of different frequencies to the response. For gratings it reveals a prominent increase around 40 Hz and a decrease at low frequencies.
We now compare this response pattern to that elicited by natural movies. Figure 2B shows the spectrogram averaged across all presentations of the different natural movies. The response shows a phasic pattern, with large variations of activity over the time course of the stimulus. Again, by averaging over time a modulation curve is obtained showing a decrease for frequencies below 20 Hz, a strong increase in the gamma range, and an increase in activity at frequencies well beyond 80 Hz. Thus, the activity pattern for natural movies differs from that for gratings: strong activation at frequencies above 80 Hz and large fluctuations during the ‘steady state’ response.

The response to the pink pixel noise stimuli includes a large range of frequencies similar to the response to natural movies (Fig. 2C). Furthermore, pink noise also elicits an irregular activity pattern over time. For this stimulus the modulation curve is similar to that for natural movies (6.8% deviation over the whole frequency range). The pink pixel noise activates the same frequencies with similar amplitudes despite the different higher order statistics and the resulting very different appearances of the two stimuli. Thus the local field potential is most sensitive to the second order statistics of the pixel intensities, but not to their higher order correlations.

In contrast to natural movies and the pixel noise, the wavelet manipulations show a much weaker stimulus induced change in the local field potential (Figs. 2D, 2E). However, their modulation curves reveal a similar pattern as for natural movies: A decrease below 20 Hz, an increase in the gamma range and a second peak of activity above 80 Hz. Comparing the wavelet filtered movie with the wavelet noise, the difference between the modulation curves is small (4.7% deviation over the whole frequency range). Thus the response of the local field potential seems only influenced by the local contrast edges of the filters, but not by their global alignment.

**Magnitude of the response for the different stimuli**

Comparing the modulation curves in Fig. 2 shows that most of their shapes are similar. Figure 3 demonstrates this similarity more quantitatively and compares the total response magnitudes. In Fig. 3A all modulation curves from Fig. 2 are redrawn in one graph. Scaling each curve by the area under the curve makes the similarity of the activity for natural stimuli and all their modifications evident (Fig. 3B). The scaling reduces the mean absolute differences between all these curves considerably (13.1% vs. 4.5%,
before vs. after respectively). The grating stimuli, however, induce little activity beyond 80 Hz, resulting in a qualitatively different form of the modulation curve. For natural movies, pixel noise, the wavelet filtered and the wavelet noise stimuli the relative contribution of different frequencies to the response pattern is very similar, but differs qualitatively from that for gratings.

The only difference between the modulation curves for the natural stimuli and all their modifications relates to the mean amplitudes. These are a measure of the total response amplitude induced by the stimulus. The distribution of these amplitudes for the same subject is shown in Fig. 3E. The amplitudes for pixel noise and the natural movies are largest. In particular, the pixel noise lacking higher order structure is at least as efficient as the natural stimuli. A similar observation holds for the wavelet filtered and the wavelet noise stimuli.

**Comparison of different subjects**

Similar results as those described above were found in all subjects (n=4). For two additional subjects the averaged modulation curves are shown in Figs. 3C and 3D. The overall shape of the modulation curves differs between animals, an issue discussed further below. However, the activity patterns for natural movies and pink pixel noise are similar within all subjects. The same result holds for the comparison of the wavelet filtered stimulus and wavelet noise. The scaling behavior described in the previous section holds for all subjects as well. Figure 3F summarizes the distribution of modulation amplitudes across all four subjects. The natural movies and the pixel noise lead to modulations of the same strength (p>0.45, Wilcoxon test). The difference between the wavelet filtered stimulus and the wavelet noise is significant (p<0.05, Wilcoxon test). Also consistently across all subjects, the activity pattern for gratings differs in every respect from that for all other stimuli. Gratings elicit a steady state response and mostly activate frequencies in the gamma range with a similar peak of activity in different subjects (cat L: 43 Hz; F: 51Hz; S: 51Hz; M: 45Hz). Thus, although the local field potential activity elicited by one particular stimulus can differ between subjects, the relative activity of different stimuli is similar across subjects and the results described above for one subject hold in general.
In each subject we recorded at sites of different depths (see methods). The modulation curves for each subject were averaged across all recording sites from this subject. It is known that the shape of the evoked potential in the local field potential changes shape, amplitude and sign depending on the depth of the recording site (Freeman 1975). However, the spectrograms and the modulation curves are derived from the power spectrum of the signal. Therefore these measures are invariant with respect to a change in sign of the local field potential. The normalization of the spectrogram furthermore eliminates an overall scaling affecting the stimulus as well as the blank period. We separately analyzed recording sites that clearly were in supra and infra granular layers (data not shown). Within all subjects, the shapes of the averaged supra and infra granular modulation curves for a given stimulus were quite similar. Thus, the relative activity induced by different stimuli at supra and infra granular sites is similar. Additionally, for three subjects of which we had data from at least 5 supra and 5 infra granular sites, we compared these quantitatively: Averaged across all frequencies and stimuli, the modulation was stronger at the infra granular sites (4.2%, 5.2% and 5.8%; subject L, F and S respectively). For comparison, the difference of the modulation curves between subjects, averaged across stimuli and frequencies, is 13.6%. While we do not have an explanation for the difference in shape of the modulation curves between subjects, we can confidently exclude that this difference is the result of a sampling bias in the recording sites.

**Multi-unit activity**

As a complement to the local field potential, which measures the activity in a region around the electrode tip and emphasizes synchronous signals (Abeles 1982), we recorded spiking activity of multi-unit sites. The example shown in Fig. 4A illustrates the responses of one recording site to a natural movie. The response is clearly modulated by the presentation of the stimulus and varies over time. Figure 4B shows the averaged firing rates during the tonic part (300 ms after stimulus onset till stimulus offset) for the different stimuli and this recording site. Natural stimuli are clearly effective in driving neurons in primary visual cortex. Comparing the stimuli with different higher order
statistics shows that natural movies and the pixel noise elicit similar firing rates. The same holds for the comparison of wavelet-filtered movie vs. wavelet noise. To average across different recording sites one has to account for the differences in mean firing rate between sites. To do so, the mean rate over all stimuli was computed for each site and the rates for the individual stimuli were divided by this mean. The average across all sites (n=33, two subjects) is shown in Fig. 4C. The natural movies elicit slightly but significantly stronger firing rates than the pixel noise (1.28 and 1.20 times the mean rate respectively; p<0.05, Wilcoxon test). This quantitative difference (6.6%) roughly matches the result form the local field potential amplitudes (0.1 %, Fig. 3F) and is much smaller than the difference to the amount of activity induced by the other stimuli. Thus both the firing rate of multi units and the local field potential amplitudes are only weakly sensitive to the higher order statistical properties of the stimuli.

**Temporal structure of the activity pattern**

The above results show that natural movies elicit a variable and irregular temporal activity pattern (Fig. 2A, Fig. 4A). The following section investigates the cause of this temporal structure and compares it quantitatively to proposed modulations of the activity reported in previous studies on visual representations.

The natural movies used for stimulation contain many irregularities in their temporal structure such as short and rapid translations. To quantify their impact on cortical activity, cross correlations between the amplitude of the stimulus motion and the activity at different frequencies of the local field potential were computed. An approximation for the global stimulus motion was obtained from the average amplitude of the flow field (see methods). An example is shown in Fig. 5A together with the spectrogram of the local field potential. Visual inspection indicates a close correlation of neuronal activity and stimulus motion. The correlogram averaged over all recording sites in one subject is shown in Fig. 5B. As can be seen, the activity in the local field potential is correlated with the motion in the stimulus. This correlation is particularly prominent at frequencies beyond 80 Hz. The correlations are strongest at a time lag where the stimulus
leads the activity by 50-100 ms, which is in agreement with known latencies of visual responses. This finding is consistent across subjects and stimulus movies (Fig. 5C).

The strength of the temporal variations can be quantified by the coefficient of variation of the power in the local field potential computed over time. The coefficients averaged across frequencies lie between 0.23 and 0.25 for the different subjects. Given that our stimuli closely match the visual input to a cat’s eye under natural conditions these results highlight an important characteristic of activity patterns under natural conditions. Body and head motion lead to temporal changes in the stimulus bringing repeatedly new stimuli onto the retina. These continuous changes in the retinal image evoke a series of visual transients. Therefore, under real world conditions, we do not observe what classically would be called a steady state response, but response strengths are continuously changing.

In the following we analyze the magnitude of the evoked transients in the local field potential and the multi unit activity in a more quantitative way. A number of studies report modulations of the mean activity in the tonic part of neural responses. Examples of such modulations in primary visual cortex are for example contextual effects (Nothdurft et al. 2000) and figure ground modulations (Lamme 1995; Roelfsema et al. 1998), which both are especially relevant for the processing of natural scenes. These studies report an average modulation of single unit firing rates of the order of 25%. To quantify the impact of the fluctuations in our population responses we measure the length of the interval necessary to reliably detect a 25% modulation of the recorded mean activity.

Taking a limited sample of the response gives an estimate of the average activity with a certain precision. Using extended windows can increase this precision. The following analysis seeks for the length of the interval necessary to reliably (e.g. on the 5% level) detect a 25% modulation of the mean activity. This implies that our method should give a positive result in only 5% of the cases when no modulation is present. Simultaneously, the detection threshold should be that low that a 25% modulation is detected in half the cases. Spectrograms of local field potentials were computed as described above. From these the mean and the variance of the total energy in a window were determined as a function of the length of the window and the width of the frequency band. Using t-statistics we determined the minimum window length, where one sample is
enough to satisfy the requirements for detection (i.e. p<0.05) as explained above. Averaging over subjects and stimuli, the minimum window length in the gamma frequency domain (30-80 Hz) was 380 ms. At higher frequencies (80-250 Hz) the necessary window length resulted in roughly 380 ms as well. By pooling different frequency bands the sensitivity could not be increased, as the variations in power were correlated across different frequencies. Obviously, higher modulations (40%) are quicker to detect (230 ms, gamma frequency range) and lower degrees of modulation (12%) require more time for significant detection (590 ms, gamma frequency range).

The temporal structure of natural movies elicits not only a highly variable activity in the local field potential, but also induces fluctuations in the spiking response of units (Fig. 4A). For these we performed a similar analysis searching for the shortest window allowing the discrimination of a given mean firing rate and an in or decreased mean firing rate. The different percentages of offsets in mean firing rate resulted in similar window lengths as for the local field potential (12%: 460 ms; 25% 420 ms; 40% 400 ms).

The above analysis investigates the minimal temporal interval needed to reliably detect a modulation of the local field potential and multi unit activity. These signals do not necessarily match the input to a decoding neuron. Whereas the multi unit activity constitutes the activity of a local group of neurons and the local field potential reflects the neuronal activity in a region of a few hundreds of micrometers, neurons in cortex may integrate signals from a larger domain. However, the fluctuations in the neuronal activity observed are correlated over large distances, thus making it unlikely that these fluctuations average out in the input to a decoding neuron. Recent results, furthermore, strengthen the link between population activity and behavior (Supèr et al. 2003). Thus, fluctuations in population the activity are likely to affect the processing of incoming stimuli as well as behavioral responses. The intervals necessary to detect changes in the average activity reported above are long compared to visual latencies of cells in higher visual areas (Perrett et al. 1982; Oram and Perret 1992) and reaction times of experimental subjects (Thorpe et al. 1996, van Rullen & Thorpe 2001). Thus, the transient fluctuations of the activity evoked by the motion the natural stimuli impose constraints on theories of cortical representations relying on differences in mean activations.
Stimuli with altered motion patterns

The difference in local field potential activity between gratings and natural movies could have several reasons. On the one hand the stimuli contain different motion patterns, uniform and irregular respectively. On the other hand the stimuli have widely different spatial structures, simple and complex respectively. To investigate which of these aspects are dominant we constructed four artificial stimuli. The flow field inherent to a natural movie was decomposed into two parts and imposed on a grating and a single frame of a natural movie (see methods). The resulting stimuli are a uniform drifting grating, a uniform drifting natural image, a jittering grating and a jittering natural image.

The activity patterns for these stimuli were recorded in two animals (cat L: 7 sessions; cat S: 7 sessions) and are summarized in Fig. 6. The response to the uniformly drifting natural image is stationary over time as is the response to the drifting grating (Figs. 6A, 6C). In contrast, the response to the jittering stimuli is irregular in both cases (Figs. 6B, 6D). To quantify the temporal irregularity we compute the coefficient of variation of the power averaged over frequencies. For the grating the difference between uniform (CV=0.186) and jitter (CV=0.208) motion is significant in all animals recorded (p<0.05, Wilcoxon test). For the natural image the difference between uniform (CV=0.186) and jitter (CV=0.203) motion is significant in both subjects as well (p<0.1). Comparing the effect of the spatial structure there is no significant difference neither for the two uniformly moving stimuli (p>0.90), nor for the two jittering stimuli (p>0.50). Furthermore, the difference between the jittering natural image and the natural movie is not significant (p>0.30). Thus, the irregular motion pattern is the dominant factor of the high temporal variability of the power in the local field potential.

For the two jittering stimuli Figs. 6E, 6F show the cross correlation of the motion amplitude and the power of the local field potential at different frequencies. Both the gratings and the natural image lead to strong correlations extending over a large range of frequencies. Furthermore, for the more complex stimuli, the jittering natural image and
the natural video, the locking of the local field potential is most prominent at frequencies above 80 Hz (Figs. 5C, 6F).

However the different stimuli cause not only differences in the temporal response but also in the frequency range activated. The uniform grating leads to a decrease below 20 Hz and an increase in the gamma range. In contrast, the jittering grating leads to an increase above as well as below 20 Hz (Fig. 6B). Interestingly, the peaks of high activity elicited by the jitter also recruited frequencies above 80 Hz. However, this is not obvious in the modulation curve due to temporal averaging. In fact, the two modulation curves are significantly different at frequencies below 31 Hz, but not above (2-way Anova, p<0.05). The jittering natural image leads to a similar shape of the modulation curve as the real movie (Fig. 6D). The uniformly drifting natural image in contrast causes a stronger increase in power than the natural movie. The difference between the two modulation curves is significant at all frequencies above 70 Hz and most frequencies below 70 Hz. However, the basic shape of the modulation curve is the same.

Concluding, the above results show that temporally irregular stimuli induce locking of the local field potential to the stimulus motion independent of the detailed spatial structure of the stimulus. Thus temporal properties and irregularities of the responses reported above for the natural movies are not bound the spatial characteristics of natural images but occur for other reasonably complex stimuli. The detailed pattern of the response, like the frequencies of the induced local field potential, depends on both the spatial and the temporal properties of the stimulus in complex way.
Discussion

In this study we investigate processing of global structure of natural scenes in the primary visual cortex of alert cats. We measured local field potentials as well as multiunit firing rates for natural movies, modified movies with different statistical properties as well as gratings. The activity pattern for drifting gratings differs strongly from that for natural movies and this difference is contingent on both the spatial and the temporal properties of the natural movies. Furthermore, the activity elicited by pink noise stimuli and the activity elicited by natural stimuli does not differ significantly, indicating only a weak impact of the higher order stimulus statistics on the responses. Natural movies elicit a very irregular temporal activity pattern. We show that the temporal structure is dominated by the motion in the visual input and quantify this irregularity that strongly limits the concept of a steady state response.

Limitations of the present study

We recorded from awake animals that, although their head was fixed, were freely watching the stimuli and moving their eyes. However, instead of excluding some eye movements, e.g. saccades from analysis or performing separate analysis during fixations and eye movements, we averaged across all eye positions. Thus we cannot directly exclude whether eye movements as such, independent of the stimulus, are the cause of some of the response properties described. However, from other studies in the same recording setup it is known that cats move their eyes much more infrequent upon stimulation with natural movies than for example humans do (Möller et al. 2002). The average inter saccade interval (>3 s) is much longer than the window used for analysis in the present study. Furthermore we measured eye movements in a few of our recording sessions and found no relation of eye movements to the timing of the stimulus presentation.

A second potential shortcoming is our limited knowledge about the nature of the local field potential and its relation to the activity of single neurons. Supposedly the local
field potential is determined by electrical activity ~500 um around the electrode tip and is influenced not only by spiking response of neurons but also by somatic and dendritic potentials especially emphasizing synchronous components (Abeles 1982). However, since it is a population measure it is not susceptible to noise in the firing of single neurons and might give a better approximation to the relevant activity in a patch of cortex. Recently evidence was put forward showing a close relationship between the local field potential and measures from magnetic resonance imaging, especially the BOLD signal (Mathiesen et al 2000; Laurizten 2001; Logothetis et al. 2001). In view of the growing body of studies using fMRI especially with primate subjects one can expect a larger number of studies with results comparable to those presented here. Thus although the sources of the local field potential are not known in detail it seems a well suited measure for the study of population activity.

*Are the results compatible with previous studies?*

Several properties of the local field potential activity reported in this study are in agreement with results previously published. We find that stimulation with drifting gratings leads to strong activation in the classical gamma range. Using natural stimuli, however, we find a strong increase of activity at higher frequencies above 100 Hz as well. In a previous study, also recording from awake cats, the strongest increase in power upon stimulation with flashed gratings was also found in the gamma range (Siegel and König 2003). However, the optimally orientation tuned frequency band determined in the latter study extended beyond 100 Hz. Together, these results consistently point to the relevance of high frequency activity in the local field potential.

Recently, a study recording in V1 of anaesthetized cats reported locking of the local field potential activity to the time course of uniform moving and randomly accelerated gratings (Kruse and Eckhorn 1996). The present study extends this finding to awake animals and stimuli of different spatial structure. In particular we show that natural movies induce activity fluctuations locked to the stimulus motion. The study by Kruse and Eckhorn furthermore reports a decrease of induced oscillations with increasing irregularity of the stimulus motion. The induced oscillations, as measured with the power
spectrum of the local field potential, were most prominent for the uniform drifting grating. In the present study we do find this effect as well albeit only upon stimulation with natural stimuli.

Little is known about how activity patterns in the visual cortex depend on the global structure of the stimuli. So far only a small number of studies compared stimuli with different statistical properties. Lehky et al. (1992) compared firing rates in the primary visual cortex of anaesthetized monkeys to image patches of different complexity. They found that complex stimuli such as random textures or 3D surfaces elicit larger firing rates than simple stimuli such as gratings. This result is in good agreement with our data showing stronger activity for natural movies and pink noise than for gratings.

Baddeley et al. (1997) measured responses in V1 of anaesthetized cats to different natural movies and white noise stimuli. They reported larger firing rates for natural movies than for the white noise, however the differences were small (4 Hz vs. 2.5 Hz). Given these small differences in activity and the fact that they used a different type of noise, this difference is not a contradiction to the results reported here.

A further comparison of cortical activity to natural stimuli and pixel noise comes from functional imaging. Measuring BOLD responses in anaesthetized monkeys, Rainer et al. (Rainer et al. 2001) report significantly stronger V1 activation for natural images than for pink pixel noise.

Despite the relevance of natural stimuli to everyday life, most studies use highly simplified stimuli and it is presently unclear which results generalize to processing of natural scenes. However it seems clear that neurons in the primary visual cortex are well adapted to processing of natural scenes. In a seminal study Vinje and Gallant (Vinje and Gallant 2000, 2002) showed that V1 neurons in awake monkeys are well driven by natural movies that information rate and transmission increase with increasing stimulus size and that response sparseness is increased by stimulation of the receptive field surround. Another study reporting efficient coding of natural scenes in V1 comes from recordings in anaesthetized ferrets. Weliky and coworkers (Weliky et al. 2003) report high population and lifetime sparseness upon stimulation with natural images. Interestingly this study reports that single cells responses are not enough to provide a reliable estimate of the local contrast in natural images but shows how a population
encodes this information efficiently. Although a direct comparison of the results reported in these and the present study is impossible, the above studies provide support for the paradigm used here: First, to obtain response properties matching those under natural conditions large field stimulation with complex scenes is required. Second, a population measure might be more relevant than single unit activity.

**Implications for theories of object binding**

Natural scenes contain many structures a human observer would classify as objects. The question how the visual system achieves this segmentation of an image into distinct objects and what the underlying neuronal mechanisms are lead to a number of theories. Two prevailing mechanisms have been proposed to explain how the responses of neurons responding to different features of the same object are bound together. The first, binding by synchrony, proposes that the spikes of two such neurons have a fixed temporal relationship (Eckhorn 1994; Singer and Gray 1995). Such response properties have been demonstrated using artificial bars (Gray et al. 1989) and figures defined by texture differences (Castelo-Branco et al. 2000; Gail et al. 2000; Woelbern et al. 2002). Following the second mechanism, global visual structure is represented by the modulation of firing rates. Two neurons falling on parts of the same object show similar increases or decreases of activity. Evidence for this hypothesis comes from studies reporting differences in the tonic part of the response depending whether the receptive field lies on an object or the background (Lamme 1995; Zipser et al. 1996; Nothdurft et al. 2000; Roelfsema et al. 2002). However, both mechanisms are hotly debated and no final conclusion has been reached so far (Lamme and Spekreijse 1998; Singer 1999; Shadlen and Movshon 1999).

What are the predictions of these mechanisms for the responses to the stimuli used in this study? The noise stimuli have a random phase spectrum and do not contain structures a human observer would classify as an object whereas the natural movies do. Thus following either binding mechanism there should be a significant difference in the response properties for these two stimuli. Following the binding by synchrony hypothesis we should see an increased level of synchronization for the natural movies compared to the noise. Assuming that this synchronization affects nearby neurons within a few
hundred micrometers, such an increase in synchronization should affect the power of the local field potential. However, this is not what we observe. The modulation curves for pink pixel noise and natural movies are very similar. Following the binding by modulation hypothesis, one could expect a change in the overall level of activity between noise and natural stimulus. In the local field potential we do not see this, and the firing rates of the multiple units show an increase for natural movies only by a small amount. Furthermore, the variance of the activity over time during the “steady state” is large and severely limits a readout of changes in the mean activity. Our estimates of the necessary readout times to detect such differences are on the order of several hundreds of milliseconds. This questions whether such a mechanism can operate fast enough under real world conditions. These considerations, however, have to be taken with a grain of salt. Whereas previous key experiments supporting either hypothesis (Gray et al. 1989; Lamme 1995) have been conducted using artificial stimuli, neither hypothesis has been elaborated in sufficient detail to allow a quantitative prediction and a hard experimental test under natural stimulus conditions. For example, we do not know enough about the statistics of “objects” in natural stimuli, about their spatial and temporal scale, to be certain that the methods applied are actually sensitive enough to elucidate the responsible mechanisms in the brain.

Acknowledgements

We would like to thank Gudrun Möller for her help during the experiments, Markus Siegel for inspiring discussions and the referees for their prompt and helpful comments on the manuscript. This work was supported by the Center of Neuroscience Zurich (CK), the SNF (Grant Nr 31-65415.01; PK, RS) and the EU/BBW (IST-2000-28127, 01.0208-1; PK).
References


Lamme VAF, Zipser K, Spekreijse H. Figure ground activity in primary visual cortex is suppressed by anesthesia. *Proc Natl Acad Sci* 35: 3263-3268, 1998.


Roelfsema PR, Lamme VA, Spekreijse H, Bosch H. Figure-ground segregation in a recurrent network architecture. J Cogn Neurosci 14: 525-537, 2002.


Figure Captions

Figure 1. Example of the activity in the local field potential. A) Raw data from one electrode after low pass filtering (500 Hz). The dashed line indicates the stimulus presentation (2000 ms duration). B) Spectrogram obtained from the data in A. The
parameters for the Fourier analysis are: 160 ms data window; zero padded to 265 ms; 152 ms window overlap. Overall the power in the local field potential decays as 1/frequency. To allow better visualization each row of the spectrogram was multiplied by the corresponding frequency. To emphasize the activity pattern induced by the stimulus a normalized spectrogram was computed: Each frequency row is divided by its standard deviation during the blank computed in a 600 ms window before stimulus onset (see gray box in B). C) Normalized spectrogram obtained from B. A value of one indicates no modulation by the stimulus whereas values above one indicate strong modulation by the stimulus.

Figure 2. Local field potential activity for the different types of stimuli in one subject. The left panel shows example frames form the different stimuli. The colored square around the frames defines the color code used in the following figures to identify the different stimuli. The right panel shows the normalized spectrograms averaged across all recording sites from this subject (n=48). Stimulus presentation starts at 0 and lasts 2000 ms. On the right of each spectrogram the corresponding modulation curve is shown. This modulation curve was obtained by averaging the spectrogram over a temporal window ranging from 400 to 1900 ms. The different stimuli are: A) Gratings (black). B) Natural movies (red). C) Pink pixel noise (orange). D) Wavelet filtered movies (blue). E) Pink wavelet noise (cyan).

Figure 3. Scaling property of the modulation curve and total response strength. Results from all subjects. A) All the modulation curves from Fig. 2 in one graph. B) Each modulation curve from A was scaled by the area under the curve demonstrating a good overlap of the different curves except for gratings (black). C) Average modulation curves (not scaled) from a second subject (cat F, n=42 recording sites). D) Average modulation curves (not scaled) from a third subject (cat S, n=42 recording sites). E) Distribution of modulation amplitudes from one subject (cat L) over recording sessions. The modulation amplitude is the area under the modulation curve. The box indicates lower and upper quartile of the distribution, the middle horizontal bar the median. The whiskers indicate the range of outliers. All pair wise comparisons are significant (Wilcoxon test, p<0.05).
F) Distribution of mean modulation amplitudes across subjects and recording sessions (n=152). The difference between natural movies and pixel noise is negligible (p>0.45) and all other differences reach significance (p<0.05). See Fig. 2 for the color code identifying the different stimuli.

Figure 4. Multi unit activity. A) Spiking activity for a natural movie recorded in subject S. The peri-stimulus histogram (PSTH) (50ms bins) is shown together with the spike trains from 19 repeats of a natural movie. B) The average firing rate of this unit for the different stimuli. For each stimulus repeat the average firing rate was computed as the mean rate in a window starting 400 ms after stimulus onset till stimulus offset. Error bars denote the standard deviation. C) Distribution of relative firing rates for the different stimuli for all units recorded (n=33). The relative firing rate for a given unit and stimulus is defined as the firing rate for this particular stimulus divided by the mean firing rate of this unit across all stimuli. Error bars indicate the standard deviation. All pairwise differences are significant (Wilcoxon test, p<0.05). See Fig. 2 for the color code identifying the different stimuli.

Figure 5. Correlation of stimulus motion and local field potential activity. A) Upper part: Spectrogram of one recording site for a natural movie. Lower graph: Amplitude of the motion in this natural movie in units of deg/s. B) Cross correlation between the local field potential activity and the motion amplitude for one subject and one stimulus movie. For each frequency of the spectrogram the cross correlation of the power with the motion amplitude of the stimulus was computed in a range of ± 300ms (at negative time lags the stimulus leads the response). The first 300 ms of the stimulus presentation were excluded to avoid artifacts induced by the evoked onset response. The result was then averaged across sessions. To filter out insignificant correlations a baseline was computed by averaging cross correlations of the same spectrogram with the motion amplitudes of other stimuli than the one shown to obtain the spectrogram. Correlations less than one standard deviation away from this baseline were taken as insignificant and set to zero. C) Cross correlations averaged across all four subjects and three stimulus movies.
Figure 6. Control stimuli with different types of motion. Two sets of stimuli were used, one moving uniformly (A and C) and one following the jitter motion extracted from natural movies (B and D). The spatial profile of the stimuli was either a grating (A and B) or a natural image (C and D). The panels A-D show the spectrograms averaged across 7 recording sessions from one subject. The modulation curves on the right in panel B (for gratings) and D (for the natural image) were obtained as in Fig.2. The solid curve refers to the uniform moving stimulus and the dashed curve to the stimulus following the jitter. In D furthermore the modulation curve for natural movies is redrawn in gray (taken from Fig. 2B). For the stimuli following the jitter motion the cross correlation of the local field potential and the stimulus motion is shown in E) for the grating and in F) for the natural image. Each correlogram was averaged across two subjects and 7 recording sites each and threshold as explained in Fig. 5.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Figure 6
Stimulus locking and feature selectivity prevail in complementary frequency ranges of V1 local field potentials

Christoph Kayser & Peter König
Institute of Neuroinformatics, University & ETH Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Address correspondence to:
C. Kayser, phone: +41-1-6353044, fax: +41-1-6353054, email: kayser@ini.phys.ethz.ch

Keywords: cat, natural stimuli, locking, tuning
Running title: Stimulus locking and feature selectivity of the lfp

Number of pages in word file: 12
Number of figures: 3
Number of tables: 0
Number of equations: 0
Number of words in manuscript: 4210
Number of words in abstract: 230
Number of words in introduction: 361
Abstract
The local field potential (lfp) is a population measure of neuronal activity complementary to spike trains. While the response properties of the spiking activity in the visual cortex have been extensively characterized, those of the lfp are not well explored. No coherent picture exists about which frequency ranges exhibit feature tuning or show stimulus locked activity. Addressing this, we record lfp in the primary visual cortex of alert cats and calculate the tuning indices for orientation, spatial and temporal frequency. Further, we quantify the locking of the power in different lfp frequency bands to the velocity profile of artificial and natural stimuli. We find that the lfp in alert animals is well tuned with similar specificity to orientation, spatial frequency and temporal frequency. Tuning to these features is most prominent in two frequency bands (8-23 Hz and 39-109 Hz). In two complementary frequency bands (23-39 Hz and above 109 Hz) the dynamics of the lfp power is tightly locked to the temporal structure of the stimulus. This locking is furthermore independent of the spatial structure of the stimulus. Together these four frequency bands cover the whole frequency range investigated. In contrast to previous studies, which often reported correlates of visual processing only in a limited frequency range of the lfp, the present results suggest that the entire frequency range of the lfp can be assigned a role in visual processing.

Introduction
Local field potentials (lfp) are a measure of neuronal activity complementary to spike trains. The lfp is influenced by currents originating from axons, somata and dendrites around the electrode and thus mainly reflects the input to a local brain region (Freeman 1975; Mitzdorf, 1985, 1987; Logothetis, 2003). As a continuous signal the lfp is well suited to investigate the temporal structure of neuronal activity. Indeed, the temporal structure of the lfp was shown to be directly related to the processing of sensory stimuli in a variety of paradigms (Gray et al. 1989; Eckhorn et al., 1993; König et al., 1995;
König & Engel, 1995; Rols et al., 2001; Woelbern et al., 2002; Fries et al., 2002; Brosch et al., 2002; Barth, 2003). In analogy to seminal work on the tuning properties of spiking activity in visual cortex (Hubel & Wiesel, 1962) the LFP has been shown to be selective to the orientation of a bar or a grating in the anaesthetized and awake preparation (Gray & Singer, 1989; Fries et al., 2000; Siegel & König, 2003). According to these studies, only selected frequency bands of the LFP are tuned to orientation. In contrast, tuning to other stimulus features, such as spatial and temporal frequency is poorly explored in the awake preparation. Furthermore, few studies exist which investigate tuning to several features in the same animal.

The phase of the LFP can lock to the temporal structure of a stimulus. This has been demonstrated using either flickering full field stimuli (Rager & Singer, 1998) or randomly moving gratings (Kruse & Eckhorn, 1996). Furthermore, the modulation of the LFP amplitude is locked to dynamic stimuli (Kayser et al., 2003). This can occur upon stimulation with gratings as well as with natural movies and thus is a property of everyday sensory processing. However, different frequency ranges have been reported to exhibit locking, but no systematic investigation exists.

Here we address the two phenomena of feature tuning and stimulus locking simultaneously in the awake cat. We study in which frequency ranges the LFP shows tuning to the features of visual stimuli and in which frequency ranges amplitude locking to the stimulus motion is observed.

**Materials and Methods**

*Subjects and recording procedures*

Data were obtained from 8 adult cats using chronically implanted electrodes. In 5 animals a micro-drive was implanted (for details see Kayser et al., 2003), in 3 animals floating electrodes were used (for details see Siegel & König, 2003). For recording the animals were head fixed and stimuli were presented on a 19-inch Hitachi monitor (120 Hz refresh, 8-bit grey scale, 57 cm in front of the animal). In part of the recording sessions, eye movements were measured using a Dual Purkinje Imaging system (Fourward Optical Technologies Inc., Clute, TX, USA; see Körding et al. 2001 for further details). The signals of the electrodes were passed through a preamplifier (Neurotrack, Ltd.) and
digitised at 20'000 Hz using a Synamp system (Neuroscan Ltd., El Paso, USA) using a 5 Hz analogue high-pass filter. During recording the state of alertness of the animal was continuously checked using an infrared camera and by examining the lfp online. All procedures were approved by the local ethics committee and conformed to NIH guidelines.

**Stimuli**

Tuning properties were measured using drifting sine-wave gratings (sf = spatial frequency, tf = temporal frequency). For spatial frequency tuning we used horizontal gratings including both directions of drift, tf=2 Hz and sf= 0.1, 0.2, 0.4, 0.8, 1.6, 3.2 cpd. For temporal frequency tuning we used horizontal gratings including both directions of drift, tf= 0.8, 1, 1.6, 2, 2.5, 4 Hz and sf=0.2 cpd. For a subset of the recording sites we used also tf=8 and 12 Hz. For orientation tuning we used 12 equally spaced directions of drift between 0 and 360 deg, sf=0.2 cpd and tf=2 Hz. Stimuli were shown in a pseudorandom sequence and repeated 20 times.

Locking of the power in a given frequency band of the lfp to the stimulus motion has two consequences: first, changes of the lfp power are correlated with changes of the stimulus velocity and second, changes of the lfp power are correlated across trials when the same stimulus is repeated. Following this, we first use two types of stimuli with known irregular motion to directly demonstrate locking of the lfp to the stimulus motion: (i) natural movie clips recorded using a camera mounted to the head of a freely moving cat and (ii) jittering gratings (sf = 0.2 cpd) moving according to the velocity profile extracted from a natural movie. Then we compute cross correlations across trials upon stimulation with (i) the same natural movie clips and (ii) pink pixel noise that has the same spatial and temporal frequency content as a natural movie but lacks the higher order structure. The latter stimulus is used to demonstrate that locking occurs also for stimuli with smoothly changing structure. The velocity profiles of the natural movies were estimated using standard techniques described elsewhere (see Kayser et al., 2003).

All stimuli had the same mean luminance and r.m.s. contrast. Stimuli lasted 2 seconds and were separated by uniform blank screens of the same length and same mean luminance.
Data analysis

The lfp was extracted by low-pass filtering the raw data at 500 Hz using a 3rd order Butterworth filter and was resampled at 1000 Hz. To quantify the activity in different frequency bands, a time localized Fourier spectrum was computed using data windows of 160 ms, overlaid with a Hanning window and zero padded to 256 ms. This results in a resolution of 3.9 Hz and 8 ms in the frequency and temporal domain respectively. Data from the first 200 ms following stimulus onset were excluded to avoid artefacts from the evoked potential.

Tuning curves were computed for each recording site and lfp frequency separately. First, the power at a given frequency was averaged over the time course of the stimulus presentation. To obtain the tuning curve, the power was further averaged across repeats of the same stimulus and across the two opposite drift directions. The selectivity of the tuning curve was quantified using a tuning index: (max-min)/(max+min); max and min denote the maximal and minimal value of the tuning curve. For each recording site this results in a curve describing the selectivity index as a function of the lfp frequency. The results in Figure 1B were obtained by averaging these curves across recording sites.

The significance of each tuning curve was assessed using a Wilcoxon rank test. The test compared the power of individual repeats of the stimulus at the optimal orientation with the power at the orientation with the lowest average response (similar for spatial and temporal frequency). This yielded one p-value for each lfp frequency. For a given recording site, we thus obtained two curves: The selectivity index and the p-value, both as a function of the lfp frequency. A high selectivity index implies a low p-value and vice versa. Tuning was taken to be significant if the p-value was below 0.05. By averaging the selectivity indices of those lfp frequencies where the p-value curve crossed this threshold, we obtained the average selectivity index corresponding to a threshold of p=0.05. These thresholds were used to determine frequency ranges with significant tuning and are indicated in Figure 1B. To visualize the selectivity of the lfp, we also computed tuning curves from the average power in the frequency range from 30 Hz to 100 Hz (Fig. 1A, upper panels).
The correlation between stimulus motion and LFP activity was estimated for each repeat of a stimulus and LFP frequency separately: The cross-correlation between the time course of the stimulus velocity and the time course of the LFP power at this LFP frequency was computed. These cross correlations were then averaged across stimulus repeats and recording sites (see Fig. 2A for an example). The significance of these correlations was estimated using a bootstrap technique (Press et al., 1992). We computed 1000 cross-correlations between a velocity profile from one stimulus and the LFP power recorded using a different stimulus. From this population the 95% confidence levels were obtained. These confidence intervals are indicated in Figure 2. The correlation of LFP power across trials and animals for a given stimulus was computed as follows. First, random trials from two different recording sites in the same or different animals were chosen. Then, for each LFP frequency, the cross correlation of the time course describing the LFP power in the two trials was computed. These correlations were averaged over 3000 random pairs. The significance of these correlations was estimated using a bootstrap as above. 1000 cross-correlations using trials obtained with two different stimuli were computed and from these the 95% confidence levels were obtained.

**Results**

The LFP power in a broad frequency range is clearly tuned to orientation, spatial frequency and temporal frequency (Fig. 1A, upper panel). From the tuning curves calculated at individual LFP frequencies and recording sites we compute the average tuning-index characterizing the strength of the tuning as a function of LFP frequency (Fig. 1A, lower panel). Tuning to orientation is prominent between 8 and 16 Hz and in a broad range of frequencies between 43 and 102 Hz. Tuning to spatial frequency is most prominent between 8 and 20 Hz and between 39 and 59 Hz. Temporal frequency tuning is most prominent between 39 and 63 Hz. Concluding, the LFP power is tuned to all three features to a similar degree and within strongly overlapping frequency bands.

These results were obtained by averaging across all different directions of gaze of the animal. To determine whether eye movements could confound these results, we measured the animals’ eye movements in a number of recording sessions during presentation of the stimuli used to assess temporal frequency tuning. From these data
saccadic eye movements were extracted and grouped according to three different conditions: blank screen, slowly drifting gratings (below 3 Hz) and fast drifting gratings (above 3 Hz). In each case the saccadic eye movements scatter in all directions of the visual field (Fig. 1B). However, the median of the horizontal component is only shifted from zero by a small fraction of the average amplitude of the eye movements (8.1%, 5.7% and 6.5% for blank screen, slowly and fast drifting gratings). In all three cases this shift is not significantly different from zero (p>0.7 in all three cases, Wilcoxon rank test). This indicates that the animals did not systematically follow the drifting stimuli.

Figure 2A shows an example of the curve describing the velocity in a natural movie clip together with the power in the lfp at 200 Hz. These curves form the basis of the cross correlation between stimulus velocity and lfp power. An example of such cross correlation is shown in Figure 2B. Strong correlations occur in two frequency bands of the lfp: around 25 Hz and well above 100 Hz. The time lag between neuronal response and stimulus velocity that leads to the strongest correlation is different for these two frequency regimes. For the first range the stimulus leads the response by roughly 100 ms, for the second by 50 ms. This latency difference is in agreement with known temporal lags between beta and gamma activity (Bekisz & Wrobel, 1999). On average we find significant locking in two frequency ranges: between 20 and 35 Hz and above 102 Hz (Fig. 2C, black curve). To control whether these frequency bands depend on the spatial structure of the stimuli, we computed the same cross correlation using randomly jittering gratings (Fig. 2C, grey curve). The frequency ranges with significant locking are comparable to those obtained using natural movies, except that now locking occurs also between 35 and 43 Hz. This locking leads to significant correlations of lfp power across trials (Fig. 2D). For natural movies the average inter-trial correlations are significant between 16 and 43 Hz and above 133 Hz, in rough agreement with the above. For the pink pixel noise significant correlations occur in a similar. The fact that locking occurs also for broadband stimuli within a limited range of lfp frequencies indicates that it is indeed a property of the brain. Furthermore, this locking phenomenon is not restricted to natural stimuli or stimuli with highly irregular motion, but occurs in the same frequency ranges also for stimuli with smoothly varying structure.
Discussion

The results of the present study show that the frequency axis of the lfp can be divided into four regimes (Fig. 3). In two of these, modulations of the power are tightly locked to the temporal structure of the stimulus, roughly between 23 and 36 Hz and above 109 Hz. In the other two, the lfp power is tuned to structural features of the stimulus, roughly between 8 and 23 Hz and between 36 and 109 Hz. These four frequency ranges are complementary and together cover the range of frequencies investigated.

Stimulus locking occurs in different forms. In a previous study locking of the phase of the local field potential to the temporal structure of the visual stimulus was studied (Rager & Singer, 1998). This relates directly to the well known evoked potential. In contrast, here we investigate modulation of the power in different frequency bands of the lfp. The modulation of the power of a given lfp frequency can take place on much slower time scales than the oscillation of this lfp frequency. For example, the power of the 100 Hz frequency component of the lfp can be modulated by a 4 Hz signal in the visual stimulus. As a consequence, the details of visual stimulation, such as the monitor refresh rate and the flicker fusion frequency limit only the signal in the stimulus that could cause the locking, i.e. the 4 Hz signal. The lfp frequency at which locking is observed, i.e. 100 Hz in this example, is only limited by the Nyquist frequency of the sampling of the electrode signals.

Working with awake animals introduces a number of complications. First, the level of alertness of the subject is difficult to define even when continuously monitoring the animals. The little power in the low frequency bands of our data supports the view that epochs of drowsiness or sleep are a small contamination at most and quantitatively not relevant. Secondly, the animals can and do freely move their eyes. This introduces a difference between the visual projection of the stimulus, and the retinal image. Stimuli similar to those used for the tuning experiments here are often used to induce an optokinetic nystagmus. This results in a regular alternation of pursuit and saccadic eye movements (Goldberg, 2000). Under the conditions of the present study, however, we found no evidence for a systematic relation of fast eye movements and drift direction. In agreement with this, a previous study found that the gain of pursuit movements in cats is low (Möller et al. 2002). Furthermore, the gain drops with increasing stimulus velocity,
suggesting that cats do not frequently follow drifting gratings with their eyes. Concluding, we do not have any indication eye movements confound the results presented here.

The tuning curves of the lfp in the awake cat reported here are compatible with previous results (Bonds, 1982; Gray & Singer, 1989) and comparable to those obtained form the firing rates of single cells (e.g. DeAngelis et al., 1993). Thus, the lfp qualitatively exhibits similar selectivity as individual neurons, although the difference between activity induced by optimal and non-optimal stimuli is smaller. Furthermore, the frequency bands for orientation tuning reported here are largely overlapping with those of optimal orientation tuning reported earlier in monkey (Frien et al., 2000) and cat (Siegel & König, 2003).

The present results indicate that the frequency axis of the lfp can be split into complementary regimes exhibiting correlates of different aspects of visual processing. These frequency bands have little overlap and together cover the whole frequencies axis. We do not want to create the impression that the stated frequency bands serve a single purpose. From numerous studies, either recording lfp or EEG, it is well known that state changes of the subject modulate low frequency activity in very narrow bands (McCormick and Bal, 1997; Destexhe et al., 1999). Similarly, attention and memory have profound effects on the activity recorded in EEG studies (Tallon-Baudry et al., 1998; Herrmann & Knight, 2001; Jensen et al., 2002). Although our results indicate a division of the frequency scale in feature tuned and stimulus locked parts, phenomena like attention and memory can modulate similar frequency ranges as well.

Acknowledgments

We would like to thank Gudrun Möller and Rodrigo Salazar for their help during the experiments, Markus Siegel for inspiring discussions and the referees for their prompt and helpful comments. This work was supported by the Centre for Neuroscience Zurich (C.K.) and the Swiss National Science Foundation (SNF grant Nr. 31-65415.01, P.K.).
Abbreviations

lfp, Local field potential; EEG, electro-encephalogram

References


**Figure Legends**

Figure 1. Feature tuning in the lfp. A) Upper row: Examples of tuning curves obtained from the average power between 30 and 100 Hz based on 20 repeats of the stimulus. Errorbars denote s.e.m. The magnitude of the response at each feature value is expressed in units of % difference compared to the response at the feature value yielding the lowest response. The response at this null feature was normalized to zero. Lower row: Frequency dependence of the tuning-index averaged across recording sites (n=37, 45, 41 from 8 animals, for orientation, spatial and temporal frequency respectively). The solid lines denote the mean, the dashed line the s.e.m. The dashed horizontal lines indicate the confidence level of 95 %. B) Distribution of saccadic eye movements during the presentation of the blank screen (left, 380 saccades), gratings drifting below 3 Hz (middle, 154 saccades) and gratings drifting faster than 3 Hz (right, 184 saccades). The stimulus was drifting in the direction of 0°. The inner circle denotes 2 standard deviations of the entire region covered by all eye movements (corresponding roughly to 15 degrees) the outer circle denotes 4 standard deviations (corresponding roughly to 30 degrees).

Figure 2. Locking of the lfp activity to the stimulus motion. A) Examples of stimulus velocity and the power in the lfp at 200 Hz during the presentation of this stimulus. The power was averaged over 50 repeats of the stimulus. B) Example of the cross correlation between lfp power and stimulus velocity for one natural movie and one recording site. At negative time lags the stimulus leads the lfp activity. C) Grand average of the cross correlation of lfp power and stimulus velocity at the optimal lag for each lfp frequency for natural movies (5 animals, 152 recording sessions, 3 natural movies) and jittering gratings (3 animals, 21 sessions). The dashed horizontal line indicates the 95% confidence interval of a bootstrap estimate. D) Cross correlation of lfp power across trials for natural movies and pink pixel noise stimuli (5 animals, 152 recording sessions). The dashed horizontal line indicates the 95% confidence interval of a bootstrap estimate.
Figure 3. Feature tuning and stimulus locking prevail in complementary frequency ranges. Solid line: The three curves describing the tuning indices from Fig. 1A have each been cut off at their respective significance threshold and averaged. Dashed line: The two curves describing the correlation between LFP power and stimulus velocity from Fig. 2B have each been cut off at the corresponding significance threshold and averaged as well. For visualization, the dashed and solid lines have been scaled to the same maximum.
Figure 1
Figure 2
Figure 3
Publications discussed in chapter 4

A comparison of hemodynamic and neural responses in cat visual cortex using complex stimuli

Christoph Kayser*, Mina Kim§, Kamil Ugurbil§, Dae-Shik Kim§ and Peter König*

*Institute of Neuroinformatics, University & ETH Zurich, 8057 Zurich, Switzerland
§Center for Magnetic Resonance Research, University of Minneapolis, Minneapolis, MN 55455, USA

Address correspondence to C. Kayser, Institute of Neuroinformatics, University & ETH Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
Email: kayser@ini.phys.ethz.ch

Running title: Hemodynamic and neural responses in visual cortex

Keywords: BOLD, firing rate, local field potential, natural stimuli, V1, Area 21
Abstract

We compare fMRI-BOLD responses in anaesthetized cats with local field potentials (LFP), aggregate high-frequency responses (analog-Mua) and spiking activity recorded in primary and higher visual cortex of alert animals. The similarity of the activations in these electrophysiological signals to those in the BOLD is quantified by counting recording sites where different stimuli elicit the same relative activation as in the imaging experiments. Using artificial stimuli, a comparison of BOLD and LFP strongly depends on the frequency range used. Stimulating with complex or natural stimuli reduces this frequency dependence and yields a good match of LFP and BOLD. In general, this match is best between 20 and 50 Hz. The measures of high-frequency activity behave qualitatively different: The responses of the analog-Mua match those of the LFP; the spiking activity shows a low concordance with the BOLD signal. This dissociation of BOLD and spiking activity is most prominent upon stimulation with natural stimuli.
Introduction

Current neuroscience employs a number of different methods to study brain function. On the one hand, there is a growing body of research using functional magnetic resonance imaging (fMRI), especially the blood oxygenation level-dependent (BOLD) contrast (Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1992). On the other hand, there exists a large body of results accumulated over the last decades using electrophysiological methods, ranging from extra cranial surface recordings to recordings of action potentials of isolated neurons. Ultimately, these two lines of research should converge to form a coherent understanding of the neuronal basis of sensory processing. However, for a number of reasons, a comparison of results obtained with these different methods is still difficult and the direct relation of activities measured by electrophysiological methods and fMRI is still debated (c.f. reviews in J. Neurosci. 23(10), 2003, pp. 3959-4011).

First, the different methods are often used in conjunction with subjects from different species. Due to its non-invasive nature, fMRI dominates studies with human subjects while electrophysiology dominates research conducted with animals. Furthermore, given the technical difficulties of simultaneously recording electrical activity and fMRI, usually only one of the two methods is used in a single subject and thus in the whole study (but see below). Classical model systems of sensory processing such as the visual system of cats and monkeys have only recently been studied using fMRI (Logothetis et al., 1999; Kim et al., 2000; Duong et al., 2001; Rainer et al. 2001).

Second, the stimuli used in an experiment are usually either specific to the question at hand or are tailored to yield strong responses with the particular method applied. Single unit recordings in the primary visual cortex, for example, often stimulate with gratings whose properties are optimized to elicit strong responses of the neuron recorded. Only few studies exist that allow comparing results obtained using fMRI and electrophysiology with the same stimulation paradigm. Furthermore, as a result of tailoring the stimuli, most studies use stimuli that have a completely different structure than the every day input processed by the system under investigation. However, recent results indicate that sensory processing is adapted to natural stimuli and that results
obtained using artificial stimuli might not directly generalize to the processing of every
day stimuli (Creutzfeldt and Nothdurft, 1978; Dan et al., 1996; Baddeley et al., 1997;
Vinje and Gallant, 2001; Vinje and Gallant, 2002; Smyth et al., 2003; Weliky et al.,
2003). Little is known about the relation between results from fMRI and
electrophysiology upon stimulation with natural stimuli, and whether the relation depends
on the type of stimulus used.

Only recently, a number of studies brought progress to a comparison of fMRI and
electrophysiology. For example, combinations of fMRI with both simultaneous and
separately recorded electroencephalographic signals in humans and animals (Menon et
al., 1997; Bonmassar et al., 1999; Brinker et al., 1999; Ogawa et al., 2000) showed
qualitative and quantitative agreement between these two methods. Recording a direct
measure of neuronal activity, the firing rate of individual cells, another group of studies
compared spiking activity and BOLD signals in the visual system. Two studies recorded
firing rates of neurons in the motion selective area MT of monkeys and BOLD responses
from the MT homologue in humans. Heeger and coworkers (Heeger et al., 1999) found
similar levels of motion opponency in both experiments, indicating a qualitative
agreement. Changing the coherence of stimulus motion in a parameterized way, Rees and
colleagues (Rees et al., 2000) found that BOLD and firing rates were directly
proportional. A similar result was obtained in a different experiment comparing contrast
response curves of BOLD and spikes in the primary visual cortex (Heeger et al., 2000).
Summarizing, these studies suggest that the BOLD signal is a good representative of
spiking activity and that fMRI can be used to infer properties of spiking activity in a non-
invasive manner.

However, there is increasing evidence that the match of fMRI and spiking activity
might not hold in general and that other measures of neuronal activity than firing rates
might better correspond to the BOLD signal. Evidence for this comes from simultaneous
recordings of local field potentials, multi unit activity and BOLD signals in the primary
visual cortex of anaesthetized monkeys (Logothetis et al., 2001). In general, both the
local field potentials and the spiking activity correlated with the changes in the BOLD
signal. At many recording sites, however, the local field potential correlated slightly
better with the hemodynamic response. An even stronger dissociation between spiking
activity and hemodynamic responses was demonstrated in the rat cerebellar cortex (Mathiesen et al., 1998; Mathiesen et al., 2000). Using different stimulation paradigms, these authors demonstrated a correlated increase of local field potentials and cerebral blood flow in the absence of spiking activity. Given that local field potentials measure local subthreshold potentials in dendrites and somata (Mitzdorf, 1985; Mitzdorf, 1987; Juergens et al., 1999), these studies suggest that the fMRI signal represents more the input to a local region of the brain than the spiking activity within this region (Logothetis, 2003).

In the present study, we compare fMRI-BOLD activations with three electrophysiological measures of neuronal activity in the visual system of the cat for a set of stimuli with different global structure. More specifically, BOLD responses in the primary and higher visual areas of anaesthetized cats to gratings, natural movies and noise patterns were measured. Using the same set of stimuli, local field potentials over a wide range of frequencies, aggregate high frequency activity (analog-Mua) and spiking activity were recorded in alert animals. We ask how well these three measures of neuronal activity correspond to the BOLD signal. First, the magnitudes of the stimulus driven responses in each of the four signals are compared. Second, we obtain relations between the activations elicited by the different stimuli in the fMRI experiments. We then calculate the number of recording sites in the physiology experiments at which the relative activity for the different stimuli obeys the same relations. This is done separately for different frequencies and frequency ranges of the local field potential, the analog-Mua and the spiking activity.

**Materials and Methods**

*Visual stimuli*

Given that stimuli with different global structure, such as drifting gratings and complex noise patterns, can elicit markedly different activity patterns (Lehky et al., 1992; Reid et al., 1992; Kruse and Eckhorn, 1996; Baddeley et al., 1997), and given that a comparison of electrophysiology and fMRI might depend on the stimulus used, we chose the following set of stimuli to compare these two methods (Figure 1A): *Sinewave gratings.*
The drifting sinewave gratings had a spatial frequency of 0.2 cycles/degree and a temporal frequency of 4 Hz. These frequency parameters are in a range yielding strong activations in the primary visual areas (Movshon et al., 1978). The orientation of the grating was chosen randomly at each presentation from one of three equally spaced orientations (0°, 60°, 120°). Each presentation of the grating included both directions of drift perpendicular to the orientation of the grating. Natural movies. These movie clips were recorded from a camera mounted to a cats’ head while the animal was exploring different environments such as forests and meadows. These videos incorporate the specific body and head movements of a cat and are an approximation to the natural visual input of the cat (for details see Kayser et al., 2003a). In total three such movie clips were used. Pink pixel noise. For each natural movie a stimulus was created which has the same second order correlations of pixel intensities but random higher order structure. These stimuli were constructed by computing the space-time Fourier transform over all movie frames and replacing the phase at each frequency by a random value between 0 and 2π. The inverse Fourier transform was applied to obtain the new stimulus. These movie clips have the same spatio-temporal frequency contrast as the original movies but lack the higher order structure that leads to (moving) objects in the original movies. In total three pixel noise stimuli were used, each constructed from one of the three natural movies.

fMRI - animal preparation

The fMRI experiments were conducted on anaesthetized and paralyzed cats. For preparation the animals were anaesthetized with Ketamine (20 mg/kg, Phoenix Inc., St. Joseph, MO) and Xylazine (1.1 mg/kg, Phoenix Inc., St. Joseph, MO). They were orally intubated and ventilated with O₂ and N₂O (3:7). During the experiment they were continuously anaesthetized with 0.8-1.3 % Isoflurane and paralyzed with Pancronium (0.4 mg/kg/hr). The eyes were dilated with Atropin sulfate, and Phenylepherine hydrochloride was applied to retract the nictitating membranes and paralyze accommodation. The eyes were fit with appropriate contact lenses to focus the stimulus onto the retina. The animals were placed in a custom build stereotaxic frame made out of plexiglas, which allows accurate positioning within the magnet. Endtidal CO₂ and body temperature were monitored continuously and kept in the desired range (3.5-4 %, 37°-38°).
Celsius respectively). All procedures were approved by the Institutional Animal Care and Use Committee (IACUC).

fMRI - data acquisition
Experiments were performed on a 9.4 Tesla horizontal magnet (Magnex, UK) with a 31 cm bore and equipped with a unity INOVA console (Varian, Ca, USA) and a 30 Gauss/cm magnetic field gradient insert (ID=11 cm, 300 µs rise time, Magnex, UK). A quadrature surface coil (4 cm diameter) was used and attached to the stereotaxic frame. In each session both functional and anatomical data were acquired. Functional data were acquired using a gradient-echo Echo-Planar-Imaging (GE-EPI) sequence (TR = 150 ms, TE = 15 ms), 13 coronal slices (2 mm thickness) and a native resolution of 64x64 voxels on a field of view of 3x3 cm. For analysis, the slices were zero padded to a size of 128x128 voxels yielding a nominal resolution of 230x230 µm. The anatomical data were obtained using a Turbo-FLASH multi-slice sequence. The visual stimuli were rear projected onto a screen (10 cm diameter) 8 cm in front of the animal. The stimulus presentation was controlled using custom written software based on the psychophysics toolbox extensions (Brainard, 1997; Pelli, 1997) and triggered from the MR scanner. Stimuli were shown in a block design with each block containing three repeats of each stimulus in a pseudorandom order. Each stimulus lasted 12 seconds and the stimuli were separated by a uniform gray screen of the same mean luminance as the stimuli but lasting 24 seconds. In each block one of the three natural movie clips, the corresponding pixel noise and one of the three orientations of the grating were used. For each animal we acquired functional data from at least 9 stimulation blocks.

fMRI - regions of interest (ROI)
The ROI’s were chosen based on the anatomical scans. These were manually aligned with the functional data correcting for the different field of views and pixel resolutions. Two sets of ROI’s were used in the present study. First, ROI’s containing regions similar to those from which the electrophysiological data were obtained: the representation of the central visual field in areas17/18 on the posterior lateral gyrus and in area 21a on the suprasylvian gyrus. Thus one ROI was defined as the region on the posterior lateral gyrus.
in two consecutive slices with similar Horsley-Clarke coordinates as in the electrophysiology recordings. Similar, a second ROI was defined as the region on the suprasylvian gyrus in two consecutive slices. The second set of ROI’s consisted of the entire visual areas 17,18,19 and 21a, and thus comprised the first set of ROI’s. These visual areas were identified on the anatomy scans using standard literature (Palmer et al., 1978; Tusa et al., 1978; Tusa et al., 1979; Tusa and Palmer, 1980). For each animal all ROI’s were defined on both hemispheres. Examples of the ROI’s are shown in Figure 1.

**fMRI - data analysis**

Functional data were analyzed using Matlab (Mathworks Inc., Natick, MA, USA). First, the time series of each voxel was preprocessed: Data points from the first six seconds of the time series were discarded to avoid transient effects of magnetic saturation and possible linear trends were removed. Functional activations were determined for each region of interest either using all voxels within the ROI or using only a subset of active voxels. These active voxels were determined using a correlation analysis. The correlation of the time series with a boxcar function representing the stimulus paradigm was computed. This boxcar function was shifted according to the hemodynamic delay, which was estimated for each animal separately. This correlation map was thresholded at a correlation value of 0.25 and a cluster criterion eliminating voxels with less than four neighbors exceeding this threshold was applied. As a control correlation thresholds of 0.1 and 0 were used as well.

The response to a particular stimulus was computed as the average percent change in the BOLD signal between stimulus and blank screen. For a given voxel and stimulus this number was computed by averaging the time series of the three presentations of this stimulus within a block and computing the percent change of the signal between stimulus and blank screen. The signal value during the blank was defined as the average of two windows, one before the stimulus (11 till 2 seconds before stimulus onset) and one after stimulus offset (10 till 19 seconds after stimulus offset). The value during the stimulus presentation was defined as the average of a 9 second window starting 3 seconds after stimulus onset. Finally, the response within an ROI was defined as the average of the percent change of either all voxels or all active voxels within this area. From the data of
each stimulus block one activity value for each ROI and stimulus was obtained. The average activity for each animal and stimulus was then computed as the average across stimulation blocks.

**Electrophysiology - recording sites**

Electrophysiological signals were recorded in five alert cats using chronically implanted electrodes (for details of implantation see Kayser *et al.*, 2003b). In each animal a small microdrive containing four bundles of movable electrodes was implanted. Two bundles were placed over areas 17/18 and two over area 21a at locations of the central visual representation. In detail, the two bundles in areas 17/18 were implanted at stereotaxic coordinates P: -3, L: +2 and L: +4; the two bundles in area 21a were implanted at P: -3, L +7 and L +9. Recordings were performed along these four penetrations at sites of different depths in supra granular, granular and infra granular layers. We did not systematically study the receptive field location of the different recording sites but they were well within the central 5° degrees of the visual field. Furthermore, measurements of spatial frequency tuning, carried out for a subset of the recording sites, showed that the spatial frequency chosen for the grating stimulus was close to optimal for these sites. Given that the penetrations were roughly orthogonal, receptive field location and spatial frequency should not vary dramatically between sites along each penetration (Hubel and Wiesel, 1962). Using standard histological techniques, we verified in three animals that the electrodes were positioned in the desired locations of the visual areas and that the electrode tracks passed through all cortical layers.

**Electrophysiology - recording procedures**

For recording the animals were restrained and their head was fixed to allow a stable visual stimulation. The stimuli were identical to those in the fMRI experiments and presented in a pseudo-random block design together with other stimuli not reported here. Stimuli were presented on a 19-inch Monitor (Hitachi, 120 Hz refresh rate, placed 50 cm from the animal) and lasted 2 seconds. Consecutive stimuli were separated by blank screens of the same length. During a recording session each stimulus was presented roughly 40 times. Broadband signals between 5 and 5’000 Hz were recorded using a 24-
Channel preamplifier (Neurotrack) and finally amplified and digitized at 20,000 Hz using a Synamp system (Neuroscan, El Paso, TX, USA). Further details of the recording procedures are reported elsewhere (Kayser et al., 2003b). All procedures were in accordance with the national guidelines for use of experimental animals and conformed to the National Institutes of Health and Society for Neuroscience (U.S.) regulations.

**Electrophysiology - data analysis**

Three measures of neuronal activity were extracted from the recorded broadband signals: local field potentials, a measure of aggregate high-frequency activity (analog-Mua) and spiking activity. The activity in the local field potential was obtained by low-pass filtering the recorded signals below 500 Hz and applying a time localized Fourier analysis (spectrogram). The Fourier amplitudes were computed for windows of 160 ms length overlaid with a Hanning window, zero-padded to 256 ms and using an overlap of neighboring windows of 152 ms; leading to a nominal temporal resolution of 8 ms. To isolate stimulus locked changes in power, the spectrogram was normalized: the percent change of the power at each frequency was computed with respect to a window during the blank preceding the stimulus (800 ms till 100 ms before stimulus onset). From each spectrogram a modulation curve characterizing the average activity at different frequencies was computed: The normalized spectrogram was averaged over a window starting 200 ms after stimulus onset till stimulus offset.

The analog-Mua was extracted by high-pass filtering the recorded signal above 500 Hz, rectifying and low-pass filtering at 200 Hz. Similar as for the local field potential, the stimulus induced activity was computed as the percent change from blank to stimulus. The average activity of the analog-Mua was computed as the average percent signal change in the same window as the modulation curve of the local field potential was computed (200 ms after stimulus onset till stimulus offset).

Multi unit spiking activity was extracted by high-pass filtering the recorded signal above 500 Hz and applying a threshold of three standard deviations of the signal. Units that did not show a modulation for any stimulus of at least a factor of two compared to the blank (i.e. spontaneous activity) were discarded. On average the spontaneous activity
of this multi unit signal was $11.5 \pm 4.1$ spikes/sec. Visual inspection showed the reliability of this automated measure.

**Results**

*fMRI measurements in the cat visual system*

The fMRI experiments were performed on anaesthetized animals in a 9.4 Tesla horizontal magnet using a quadrature surface coil placed over the posterior part of the skull. The field of view covered a large part of the visual cortex, most notably the primary and secondary visual areas. For the present analysis we used only data from animals that showed reliable hemodynamic responses during the entire experiment. This was the case in five out of eight animals. In the remaining three experiments either no hemodynamic response could be obtained at all or responses repeatedly vanished and appeared again. This would happen on the timescale of roughly half an hour. In these experiments no stable state could be reached. In contrast, from the animals included in this study we obtained data from at least 9 stimulation blocks each and the response strength was stable over several hours.

*Figure 1 about here*

Voxels showing significant stimulus driven responses were identified using correlation analysis and clustering. Reliable activity was observed in the primary and secondary visual areas, in the posteromedial lateral suprasylvian area and in the thalamus (Figure 1B). The response strength of a voxel for a given stimulus was computed as the percent change in the BOLD signal compared to the blank period. Figure 1C shows an example of the signal from a region of interest (ROI) on the posterior lateral gyrus. Clearly the signal is modulated by the visual stimulation and the strength of the signal increase depends on the stimulus. Together with the functional data, high-resolution anatomy scans were obtained. These were used to identify the visual areas 17,18,19 and 21a as indicated on the left hemisphere in Figure 1D. In addition, two ROI’s based on the
location of the recording sites in the physiology experiments were used. Examples of these ROI’s are shown on the right hemisphere in Figure 1D.

**fMRI results**

The following paragraph compares the hemodynamic responses to the different stimuli in the electrophysiology-based ROI’s. In order to have an estimate of the visually evoked BOLD response within a ROI, the percent signal change was averaged over all active voxels in that region. Figure 2 summarizes the average hemodynamic response for all animals, stimuli and the two areas of interest. In the primary visual cortex, the natural movies evoked the smallest responses (0.56 % averaged across animals, c.f. Table 1) whereas gratings are an effective stimulus and evoked the strongest responses (1.1 %). The noise stimulus lead to stronger responses than the natural movies, but weaker than the gratings (0.66 %). Area 21a showed the same tendencies, however, the overall response strength was higher than in the primary area (1.48 %, 0.74 % and 1.21 % for gratings, natural movies and pixel noise respectively).

**Figure 2 about here**

There is a difference in the average level of activation between animals. To compare the responses to different stimuli irrespectively of this variation, we compute the relative activation for different stimuli within each animal. Differences between stimuli consistent across animals are extracted by counting which relation between stimuli occurred in at least four out of the five animals. With this method one finds that in the primary areas gratings elicited stronger responses than natural movies and than the pixel noise. Furthermore, the pixel noise caused stronger activations than the natural movies. Similarly, in area 21a were the responses to gratings stronger than those for natural movies and the pixel noise. However, there was no consistent difference between natural movies and pixel noise in this ROI.

In the above analysis there are two arbitrary choices. First, the ROI’s were restricted to regions similar to those that were used to sample the electrophysiological data. However, since the two experiments were not conducted in the same group of
animals, the border of these regions is arbitrary. Second, a fixed correlation threshold of 0.25 was used to identify active voxels. However, this threshold sets a limit for the average activity and thus could also influence the relative activity between stimuli. We made two controls to evaluate the influence of these choices.

The first control used the entire visual areas as ROI’s to compute the average activities for the different stimuli. Figure 3A shows the average hemodynamic response to the different stimuli in areas 17/18 and 21a for one animal. Although the magnitude of the response differs compared to the ROI’s based on the physiology experiments, the relative activity for different stimuli was similar. The same result is found in all animals. Thus, changing the definition of the ROI changes the average magnitude of the hemodynamic responses, but the relations between the activities for the different stimuli found above hold also the when entire anatomical areas are used as ROI.

The second control used different correlation thresholds to determine the set of voxels included in the average. Besides a threshold of 0.25 as above, thresholds of 0.1 and 0 together with the clustering criterion and a threshold of -1 without clustering were used. In the last case all voxels within an ROI were included in the average. As expected, the average activity decreases with decreasing threshold: 0.98 %, 0.71 %, 0.41 % and 0.07 % for thresholds of 0.25, 0.1, 0 and -1 respectively (numbers averaged across all animals, stimuli and ROI’s). Figure 3B compares the relative activities for the different stimuli and correlation thresholds taking this difference in the overall response strength into account. Although the magnitude of the relative response to different stimuli depends on the correlation threshold, the three relations between the stimuli reported above hold for all values of the correlation threshold: For all thresholds is the response to gratings stronger than that to the natural movies and the pixel noise; furthermore is the response to the pixel noise stronger than that to the natural movies. A similar effect of the choice of the correlation threshold is found in all animals. This indicates that the results reported above do not depend on the specific choice of which voxels are included in the ROI’s. But the relative responses to the different stimuli are a property of the visual areas and not of a particular small population of voxels.
Electrophysiology experiments
In a different group of animals we recorded broadband signals in areas 17/18 and area 21a of alert cats. From these recordings three measures of neuronal activity were extracted, each capturing a different aspect of the underlying processes. First, local field potentials (LFP) which were defined as the low frequency component (5-250 Hz) of the recorded signal and are a measure of local aggregate dendro-somatic potentials (Freeman, 1975; Mitzdorf, 1985; Mitzdorf, 1987; Juergens et al., 1999; Logothetis, 2003). Second, a measure of aggregate high frequency activity (analog-Mua; 0.5-3 kHz), which is influenced by the spiking activity in a larger region around the electrode tip (Buchwald and Grover, 1970; Gail et al., 2000; Super et al., 2003). This measure was obtained as a low-pass filtered version of the high frequency signal. Last, spiking activity of single or small populations of units (Adrian and Zotterman, 1929; Hubel and Wiesel, 1962). To allow a comparison with the fMRI data, the activations in these three measures are expressed in units of percent change from blank to stimulus.

Example data from one recording session are shown in Figure 4A. The activity in the local field potential was quantified using a time localized Fourier analysis. To quantify stimulus induced changes in the power, the spectrogram is normalized at each frequency with respect to the blank period. The example of such a normalized spectrogram in Figure 4A (top panel) shows a transient response after stimulus onset followed by sustained activity that selectively activates different frequency bands. Averaging the activity pattern across time yields a modulation curve quantifying the average response at individual frequencies for a given stimulus. Examples of such modulation curves from two animals, one from areas 17/18 and one from area 21a, are shown in Figures 4B and 4C (top panels). The amplitude of the response depends strongly on the frequency of the local field potential. Whereas low frequencies often show a decrease in power, frequencies in the gamma range show increases in the order of 50%. Although the shape of the modulation curves varies between animals, the relative
activities for different stimuli show a consistent pattern across animals (Kayser et al., 2003b). Comparing the total response strength across frequencies and animals (Table 1) reveals that gratings lead to weaker activations than both the natural movies and the pixel noise and that the latter lead to responses of similar strength.

An example of the high frequency activity measured in the analog-Mua is shown in Figure 4A (middle trace). It is characterized by a short transient response followed by a tonic response pattern similar to the local field potential. Figures 4B and 4C (middle panel) show the average of this tonic response across recording sites and animals. The responses are stronger in areas 17/18 compared to area 21a by a factor of two. Furthermore, gratings elicit roughly half the response as the other two stimuli, which were similar in response strength (Table 1). An example peri stimulus-time histogram of the spiking activity is shown in Figure 4C (bottom panel). Compared to the other measures of neuronal activity, the firing rates show the strongest activations (Figures 4B and 4C, bottom panel).

Table 1 about here

Comparison of average BOLD and neural response strength

The activations in the BOLD signal and the neuronal activities described above are all quantified in units of percent signal change and can thus be compared directly (Table 1). The amplitudes of the BOLD activations averaged across animals and stimuli are 0.77 % for the primary areas and 1.14 % for the higher area. This relatively low amplitude is in agreement with other fMRI studies using anesthetized felines (Kim et al., 2000). In contrast, the modulations of the local field potential averaged across frequencies and stimuli are stronger by an order of magnitude: 14.3 % in the primary area and 12.4 % in the higher area. However, as seen above the modulation depends strongly on the frequency and can reach 60 % in the gamma range while being small at high frequencies. In fact, the analog-Mua, which is directly constructed from the high frequency components of the recorded signal, shows a much weaker activation magnitude than the average local field potential: 3.7 and 1.7 % in the primary and higher area respectively. This is on the same order of magnitude as the BOLD responses. In contrast
to this are the modulations of the thresholded spiking activity much stronger: 195 % in the primary areas and 147 % in the higher area. These results agree well with findings from simultaneous recordings of fMRI and electrophysiology in the monkey visual system that showed a much stronger neuronal than hemodynamic response (Logothetis et al., 2001). Together, these results point to an underestimation of the reliability and strength of the neuronal activations by BOLD measurements.

Comparison of BOLD and neural responses across recording sites

From the BOLD activations we derived three relations between the different stimuli that were consistent across animals: gratings elicit stronger responses than the pixel noise; the pixel noise elicits stronger responses than the natural movies and the gratings elicit stronger responses than the natural movies. Looking at the average responses measured in the physiology experiments (Table 1) reveals that on average the local field potential is not in agreement with the first two relations but with the last one.

The following paragraphs extend this comparison to individual recording sites and count the number of recording sites in the physiology experiments at which the relative activity for the different stimuli obeys the same relations as in the fMRI. This is done separately for the different frequencies of the local field potential, for the analog-Mua and the firing rates. Note that applying a similar criterion to the fMRI data would not lead to a hundred percent match, since some of the differences between stimuli that form the basis of this comparison were only found in four out of five animals.

Figure 5 about here

Counting the number of recording sites in areas 17/18 at which the response of the local field potential to gratings exceeds that to natural movies yields the curve shown in Figure 5A. The number of recording sites that are in agreement with the relations from the fMRI depends on the local field potential frequency. The best match is found in the gamma range (88 % at 42 Hz) and the percentage decreases quickly at low and high frequencies. Averaged across frequencies this results in a match of 27 %. Figure 5B shows the same analysis for relation between the activities for gratings and pixel noise. Again, the number of recording sites that are in agreement with the fMRI depends on the
frequency (peak of 88 % at 39 Hz) and the average match is 29 %. In contrast, the comparison of pixel noise to natural movies is less dependent on the frequency (Figure 5C; maximum of 68 % at 29 Hz; 51 % averaged across frequencies). A similar result is found when the same analysis is performed on the recordings from area 21a (data not shown). The comparisons involving gratings show strong frequency dependence with similar peaks (89 % at 39 Hz for gratings > natural movies and 87 % at 46 Hz for gratings > noise). The comparison of the two complex stimuli shows a weaker frequency dependency (maximum 72 % at 25 Hz). Averaged across frequencies this yields a match of 30 %, 29 % and 57 % for the three relations. As a consequence, we note that a comparison of results from fMRI and local field potentials can strongly depend on two factors: the frequency band of the local field potential used and the types of stimuli involved.

Next we compare the responses of the analog-Mua to the different stimuli. Comparing the responses to gratings to that to the pixel noise, the analog-Mua matches the fMRI at roughly 20 % of the recording sites in areas 17/18 (Figure 5A); in area 21a the same percentage is reached. Comparing the responses for gratings to that for natural movies yields the same result. In contrast, comparing the responses to the pixel noise to that to the natural movies, the analog-Mua agrees with the fMRI at 65 % in area 18 (Figure 5C); in area 21a the match is 57 %. Overall, the comparison of the analog-Mua to the fMRI yields a similar match as found for the local field potential.

Last, we compare the results from the spiking activity with the fMRI. Looking at the comparisons involving gratings, the spiking activity yields a worse match with the fMRI than does the analog-Mua. At no site was the firing rate for the gratings higher than for the natural movies. Only at 20 % of the sites in areas 17/18 and 5 % in area 21a was the firing rate for the gratings higher than for the pixel noise. Comparing the activity for the two complex stimuli, the spiking activity matches the fMRI at 46 % of the sites in areas 17/18 and 50 % in area 21a. Thus, the comparison of the spiking activity to the fMRI for the complex stimuli yields a worse match than does the local field potential or the analog-Mua.

The outcome of the above comparison of fMRI responses and the different measures of neuronal activity is dependent on the type of stimuli used. The comparisons
involving gratings show a strong influence of the frequencies of the local field potential. However, when using only complex stimuli the details about frequency ranges used become less important. We hypothesized that this frequency dependence should become even weaker when only natural stimuli are used. From the BOLD activations for the three individual natural movies we extracted their relative activations. On average across animals the three natural movies induced different BOLD responses (1.15, 1.05 and 0.87 %) and the difference of the activity induced by these movies is consistent in four out of five animals. The curve in Figure 5D shows the number of recording sites in areas 17/18 at which the local field potential obeyed these relations. Indeed, the dependence of this curve on the frequency axis is small and yields an average of 56 %. A similar comparison with the activations from the analog-Mua yields a match of 68 %. In sharp contrast, a comparison with the spiking activity shows that only 25 % of the recording sites show the same relative activations for the different natural movies. The results obtained from area 21a are qualitatively and quantitatively similar (53 %, 62 % and 30 % for the local field potential, the analog-Mua and the spiking activity). Thus, when only natural stimuli are considered, a comparison of local field potentials and BOLD responses becomes insensitive to the range of frequencies used. Furthermore, the dissociation between activations in the local field potentials and in the analog-Mua on the one hand and the thresholded spiking activity on the other hand becomes stronger and the spiking responses show a weak concordance with the BOLD responses.

In the above comparisons single frequencies of the local field potential were used. However, studies analyzing local field potentials often use the average power in a given frequency band as the signal characterizing the activity in the local field potential. We made a similar comparison between local field potentials and the results from the fMRI as above, using the average power in different frequency bands instead of single frequencies. The results from these comparisons (Figure 6) are in good agreement with those from single frequencies. Comparing gratings to natural movies or pixel noise shows a strong dependence on the frequency interval used, with the peak occurring for the interval from 35 Hz to 54 Hz. The comparison of the two classes of complex stimuli, natural movies and pixel noise, yields a maximum for the interval of 19 Hz to 31 Hz and a minimum in the range of 50 Hz to 100 Hz. Similar, the comparison between the
different natural movies yields a good match with a maximum in the range of 15 Hz to 31 Hz and a minimum for intervals between 50 Hz and 100 Hz. Altogether, these results show that independently of the stimuli used, the match of BOLD activations and local field potentials is strongest for frequency intervals roughly in the range of 20 Hz to 50 Hz.

**Discussion**

The present study compares neuronal activities measured using electrophysiological methods to the fMRI-BOLD signal. The local field potential gave a good agreement with the BOLD signal only in a limited frequency range when different types of stimuli were mixed, e.g. gratings and complex stimuli. When comparing the activities for complex or natural stimuli, the frequency dependence of the match between local field potential and BOLD response was much reduced. Under these conditions the analog-Mua achieved a comparable match to the BOLD signal as the local field potential. The classical localized measure of neuronal activity, the thresholded spiking activity, showed a different behavior. Irrespective of the stimuli used for the comparison, the firing rates and BOLD activations for different stimuli never agreed at more than 50 % of the recording sites and the spiking activity performed poor at predicting the BOLD responses. Summarizing, the comparison of electrophysiology and fMRI data depends on two factors. First, the definition of the neural activity extracted from electrical signals recorded using microelectrodes. And second, the type of stimulation used, which in the present case is the type of visual stimulus.

Overall, the best match of local field potentials and fMRI-BOLD signals was found in the gamma frequency range of the local field potential. Neuronal activity in this frequency range has been observed in the visual cortex under a variety of stimulation paradigms (Singer and Gray, 1995). The precise origin of this phenomenon is still not resolved and its significance with respect to coding global stimulus properties is hotly debated (Singer, 1999; Shadlen & Movshon, 1999). Most models addressing these experimental data assign a prominent role to inhibitory mechanisms (von der Malsburg,
Inhibitory interactions between neurons do not increase the overall spike rate but consume energy and lead to an increased blood flow, as supported by experimental results obtained in the cerebellum (Mathiesen et al., 2000). As a result, fMRI measurements are sensitive to this type of process. Thus, the data presented here are compatible with the speculation, that by virtue of inhibitory processes, fMRI measurements are sensitive to activity in the gamma frequency range.

An important point requiring discussion is the fact that the animals in the fMRI experiments were anaesthetized. In general, the use of anesthesia is very prominent in animal research because it eases the experiment and allows prolonged time for data acquisition. Some experiments, like the fMRI scans with feline subjects in the present study, were unfeasible without anesthesia, due to the gradient noise during functional scans utilizing the fast Echo-planar imaging (EPI) sequences. The effects of anesthetics on the responses can be variable. Even anesthetics of the same class, like halothane and isoflurane, have been reported to have differential effects on global measures of brain activity (Villeneuve et al., 2003). However, anesthetics can alter the response properties of neurons. For example the spatio-temporal dynamics of spontaneous and stimulus evoked responses is altered by urethane anesthesia in the rat (Erchova et al. 2002). Furthermore, receptive field properties can be altered both in the visual cortex (Roberston 1965; Lee 1970) and the thalamus (Friedberg et al. 1999). However, the BOLD signal can be assumed to be relatively insensitive to such changes of temporal response dynamics as it measures a highly "smoothed" version of the underlying neuronal responses at a time scale of hundreds of milliseconds (rather than the millisecond scale of the underlying neuronal responses). Furthermore, effects of anesthetics seem to be more prominent in paradigms requiring visual attention as indicated by the disappearance of contextual effects like figure ground modulation during anesthesia (Lamme et al. 1998). Such an effect might be visible also in fMRI experiments as the BOLD signal is more sensitive to attentional effects than firing rates (Heeger & Ress 2002). However such effects of the extra classical receptive field play a minor role in the present stimulation paradigm. Modulations of responses by anesthetics have been assessed in fMRI experiments as well. Effects of sub-anesthetic doses of Isoflurane are noticeable in many cortical areas but seem to spare primary visual areas (Heinke and Schwarzbauer 2001).
Along similar lines, the effect of ketamine on neuronal responsiveness as measured by the BOLD signal occurs only at high doses and shortly (<1 hour) after injection (Leopold et al., 2002). Thus, anesthesia has little effects on the primary visual areas and especially on paradigms that do not require selective visual attention. Based on this, we conclude that the use of anesthetics supposedly has a limited effect on the comparison of the fMRI-BOLD responses from anaesthetized animals with electrophysiological signals recorded in the alert animal presented here.

Despite its wide use in neuroscience research, the neural origin of the BOLD signal is still debated. Recently, evidence was put forward that the BOLD signal corresponds well to firing rates of single units measured in the same visual area. Indirect comparisons of fMRI measurements in humans and recordings of firing rates in monkeys showed a qualitative (Heeger et al., 1999) and also a quantitative (Rees et al., 2000; Heeger et al., 2000) agreement. However, simultaneous recordings of local field potentials, spiking activity and hemodynamic responses in monkey visual cortex (Logothetis et al., 2001) and rat cerebellum (Mathiesen et al., 1998; Mathiesen et al., 2000) provided strong evidence that the signals measured using fMRI are stronger coupled to local field potential activity than to firing rates. Since the local field potentials reflect synchronized components of dendro-somatic potentials and subthreshold fluctuations (Mitzdorf, 1985; Mitzdorf 1987; Juergens et al., 1999) this led to the conclusion that the fMRI is more sensitive to the input to a local region of the brain than to its spiking output (Logothetis et al., 2001, Logothetis, 2003). The results of the present study support this view that the BOLD signal reflects local subthreshold processes more than spiking activity.

**Acknowledgements**

We would like to thank Gudrun Möller and Rodrigo Salazar for helping with the electrophysiological experiments and Itamar Ronen for helping with the fMRI experiments. Furthermore we thank the referees for their constructive and helpful comments. This work was financially supported by the Center for Neuroscience Zurich, the Swiss National Science Foundation (Grant Nr. 31-65415.01), the EU/BBW (IST-
2000-28127, 01.0208-1), the National Institutes of Health (RR08079, MH61937), the Keck Foundation and the Human Frontiers Science Program.
References


Figure legends

Figure 1. Stimulus paradigm and example activations.
A) Example frames from the three types of stimuli: Sinewave gratings, natural movies and pixel noise (from left to right). B) Activation maps color-coded on the anatomy scans. Shown is the correlation of the BOLD signal with the stimulus paradigm. Slices range from anterior to posterior (top left to bottom right). Active voxels are most prominent in the primary visual areas and in the thalamus (top right slice). C) Stimulation paradigm: stimuli (12 sec, gray bars) were separated by uniform blank screens (24 sec, white area). The black trace shows the BOLD signal obtained from a region of interest on the posterior lateral gyrus. D) Example from an anatomy scan together with the different ROI’s. On the left hemisphere the ROI’s representing the different visual areas are shown: dark green: A17; light green: A18; dark blue: A19; light blue: A21a. On the right hemisphere the ROI’s based on the recording sites in the electrophysiology experiments are indicated: red: ROI in areas17/18, orange: ROI in area 21a.

Figure 2. Average BOLD responses.
Shown is the average percent change in the BOLD signal for the different stimuli, animals and ROI’s. The average was computed across all active voxels (correlation threshold of 0.25) and the ROI’s based on the physiology experiments. The bars denote the average and the errorbars denote the s.e.m. over stimulation blocks. The different types of stimuli are color-coded (black: gratings; dark gray: natural movies; light gray: pixel noise).

Figure 3. BOLD responses averaged over entire visual areas and using different correlation thresholds.
A) Percent change in the BOLD signal averaged over all active voxels in the ROI’s comprising the entire visual areas 17/18 and 21a for one animal (animal Nr. 3). As in Figure 2 the bars denote the average and the errorbars denote the s.e.m. over stimulation
blocks; stimuli are color-coded. B) Control using different correlation thresholds to compute the average BOLD response in the two physiology-based ROI’s (again for animal Nr. 3). The results from the different thresholds are color-coded: 0.25, 0.1, 0 and -1 from black to increasing light gray. The different curves were shifted vertically to allow a better comparison of different thresholds despite differences in the overall level of activity (see text).

Figure 4. Electrophysiology experiments.
A) Example activity in the local field potential, analog-Mua and spiking activity during the presentation of a natural movie clip. Top panel: normalized spectrogram of the local field potential; Middle panel: analog-Mua; Bottom panel: PSTH of multi unit spikes, all averaged over 40 repeats of the stimulus. All responses are expressed in units of signal change from blank to stimulus. The dashed line below the spectrogram indicates tonic part of the response over which the modulation curves were computed. B, C). Average response strength in the primary (B) and the higher visual areas (C). Top: modulation curves of the local field potential for the different stimuli and one animal each (n=49 sites for areas 17/18; n=9 sites for area 21a). Middle: response in the analog-Mua (n=42 sites; 5 animals). Bottom: response in the spiking activity (n=15 and 18 sites respectively; 3 animals). The different types of stimuli are color-coded (black: gratings; dark gray: natural movies; light gray: pixel noise).

Figure 5. Neural activity consistent with fMRI results.
For each of the three measures of neuronal activity recorded in the physiology experiments the figure shows the percentage of recording sites at which the activity for the different stimuli obeys the relations that were obtained from the BOLD responses: A) Sine-wave gratings > natural movies; B) Sine-wave gratings > pixel noise; C) Pixel noise > natural movies; D) Relative activity for the different natural movies. The data include 167 local field potential recording sites (5 animals), 42 analog-Mua sites (5 animals) and 33 multi unit sites (3 animals).

Figure 6. Frequency ranges of the local field potential consistent with the fMRI results.
For each possible frequency range of the local field potential the graph shows the percentage of recording sites at which the activity for the different stimuli obeys the relations that were obtained from the BOLD responses (similar analysis as in Figure 5 for the local field potential, but using frequency intervals instead of single frequencies). The comparison of the activity for different stimuli is based on the average power in frequency bands whose lower border is given by the frequency on the x-axis and whose upper border is given by the frequency on the y-axis. The relations between stimuli are: A) Sine-wave gratings > natural movies; B) Sine-wave gratings > pixel noise; C) Pixel noise > natural movies; D) Relative activity for different natural movies.

**Tables**

<table>
<thead>
<tr>
<th>Stimuli / Signal</th>
<th>LFP</th>
<th>Analog-Mua</th>
<th>Spikes</th>
<th>BOLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gratings / Area</td>
<td>A 17/18</td>
<td>A 21a</td>
<td>A 17/18</td>
<td>A 21a</td>
</tr>
<tr>
<td>Gratings</td>
<td>10.0</td>
<td>9.8</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Natural movies</td>
<td>16.1</td>
<td>13.5</td>
<td>4.5</td>
<td>2.3</td>
</tr>
<tr>
<td>Pixel noise</td>
<td>16.8</td>
<td>14.0</td>
<td>4.6</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 1. Average response strength

The table lists the response strength averaged across animals separately for the different stimuli and areas. The numbers for the three electrophysiological measures of activity represent the average across all recording sites in the respective area, the BOLD responses were averaged across all active voxels in the physiology based ROI's. The response of the local field potential (LFP) was averaged over the frequency range from 4-250 Hz.
Figure 1
Figure 2
Figure 3

A

Areas 17/18

Area 21a

ROI in areas 17/18

ROI in area 21a

B

Grating Natural movie Pixel noise

Grating Natural movie Pixel noise

Grating Natural movie Pixel noise

Grating Natural movie Pixel noise

Corr. threshold: 0.25 0.1 0 -1

ROI in areas 17/18

ROI in area 21a

0

0.7

1.4

signal change [%]

Grating Natural movie Pixel noise

Grating Natural movie Pixel noise

Grating Natural movie Pixel noise

Grating Natural movie Pixel noise
Figure 4
Figure 5
Figure 6
Publications discussed in chapter 5


Learning the Nonlinearity of Neurons from Natural Visual Stimuli

Christoph Kayser
kayser@ini.phys.ethz.ch
Konrad P. Körding
koerding@ini.phys.ethz.ch
Peter König
peterk@ini.phys.ethz.ch
Institute of Neuroinformatics, University / ETH Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Learning in neural networks is usually applied to parameters related to linear kernels and keeps the nonlinearity of the model fixed. Thus, for successful models, properties and parameters of the nonlinearity have to be specified using a priori knowledge, which often is missing. Here, we investigate adapting the nonlinearity simultaneously with the linear kernel. We use natural visual stimuli for training a simple model of the visual system. Many of the neurons converge to an energy detector matching existing models of complex cells. The overall distribution of the parameter describing the nonlinearity well matches recent physiological results. Controls with randomly shuffled natural stimuli and pink noise demonstrate that the match of simulation and experimental results depends on the higher-order statistical properties of natural stimuli.

1 Introduction

One important application of artificial neural networks is to model real cortical systems. For example, the receptive fields of sensory cells have been studied using such models, most prominently in the visual system. An important reason for the choice of this system is our good knowledge of natural visual stimuli (Ruderman, 1994; Dong & Atick, 1995; Kayser, Einhäuser, & König, in press). A number of recent studies train networks with natural images, modeling cells in the primary visual cortex (see Simoncelli & Olshausen, 2001, and references therein). In these studies, a nonlinear model for a cell was fixed, and the linear kernel was optimized to maximize a given objective function. The resulting neurons share many properties of either simple or complex cells as found in the visual system. However, these studies incorporate prior knowledge about the neuron model to quantify the nonlinearity of the neuron in advance. But to better understand and predict properties of less-well-studied systems, where such prior knowledge is
not available, models are needed that can adapt the nonlinearity of the cell model.

Here we make a step in this direction and present a network of visual neurons in which both a parameter controlling the nonlinearity and the linear receptive field kernels are learned simultaneously. The network optimizes a stability objective (Földiák, 1991; Kayser, Einhäuser, Dümmer, König, & Körding, 2001; Wiskott & Sejnowski, 2002; Einhäuser, Kayser, König, & Körding, 2002) on natural movies, and the transfer functions learned are comparable to those found in physiology as well as those used in other established models for visual neurons.

2 Methods

As a neuron model, we consider a generalization of the two subunit energy detector (Adelson & Bergen, 1985). The activity $A$ for a given stimulus $I$ is given by

$$A(t) = (|I \cdot W_1|^N + |I \cdot W_2|^N)^{\frac{1}{N}}. \tag{2.1}$$

Each neuron is characterized by two linear filters $W_{1,2}$, as well as the exponent of the transfer function $N$. For $N = 1$, the transfer function has unit gain, and except for the absolute value, the neuron performs a linear operation. For $N = 2$, this corresponds to a classical energy detector. All of these parameters are subject to an optimization of an objective function. The objective used here is temporal coherence, a measure of the stability of the neuron’s output plus a decorrelation term:

$$\Psi = \Psi_{time} + \Psi_{Decorr} = -\frac{\langle (A(t) - A(t - \Delta t))^2 \rangle_{\text{Stimuli}}}{\text{var}(A(t))_{\text{Stimuli}}} + \sum_{\text{Neurons}i,j} -\langle A_i A_j \rangle_{\text{Stimuli}}^2. \tag{2.2}$$

A neuron is optimal with respect to the first term of the objective function if its activity changes slowly on a timescale given by $\Delta t$. The second term avoids a trivial solution with identical receptive fields of all neurons. Implementing optimization by gradient ascent, the formula for the change of parameters can be obtained by differentiation of equation 2.2 with respect to the parameters describing the neurons.

The subunits’ receptive fields $W$ are initialized with values varying uniformly between 0 and 1 and the exponents with values between 0.1 and 6. During optimization, in a few cases, the exponent of a neuron would diverge to infinity. In this case, the transfer function implements a maximum operation on the subunits, and the precise value of the exponent is no longer relevant. To avoid the associated numerical problems, we constrain the exponent to be smaller than 15.
The input stimuli are taken from natural movies recorded by a lightweight camera mounted to the head of a cat exploring the local park (for details, see Einhäuser et al., 2002). As controls, two further sets of stimuli are used. First, temporally “white” stimuli are constructed by randomly shuffling the frames of the natural video. Second, the natural movie is transformed into spatiotemporal “pink” noise by assigning random phases to the space-time Fourier transform. This yields a movie with the same second-order statistics as the original but without higher-order correlations. From two consecutive video frames, one corresponding to $t - \Delta t$, the other to $t$, we cut pairs of patches of size $30 \times 30$ pixels ($\approx 6$ degrees). The square patches are multiplied by a circular gaussian window in order to ensure isotropic input. The training set consists of 40,000 such pairs. The dimension of stimulus space is reduced by a principal component analysis in order to ease the computational load. Excluding the first component corresponding to the mean intensity, we keep the first 120 components out of a total of 1800, carrying more than 95% of the variance.

We quantify the properties of the receptive fields in an efficient way by convolving the subunits with a large circular grating covering frequencies ranging from 0.5 cycles per patch to 15 cycles per patch and all orientations (see Figure 1C). The convolutions of the subunits with this patch are summed pointwise according to the cell model (see equation 2.1) resulting in a two-dimensional activity diagram.

3 Results

We optimize the receptive fields and exponents of a population of 80 cells. In the following, we first discuss one example in detail before reporting results of the whole population.

The development of the subunits of the example neuron is shown in Figure 1A. It is selective to orientation and spatial frequency. This selectivity develops after about 50 iterations simultaneously in both subunits, and the spatial profile is very similar to a Gabor filter. Indeed the best-fitted Gabor filter explains 81% and 74% of the variance of the two subunits’ receptive fields. Simultaneously with their emergence, the oriented structures in the two subunits are shifted relative to each other by 87 degrees. The evolution of the exponent of this neuron during optimization is shown in Figure 1B. It converges to a value of 2.35, which is close to the exponent of 2.0 of the classical energy detector. The response profile characterizing the spatial receptive field (see methods) is displayed in Figure 1C. It shows that the neuron is selective to orientation (oblique) and spatial frequency (11 pixel wavelength). The response to a grating shows a constant increase toward higher spatial frequencies, with a small-frequency doubled modulation on top. The response to a drifting grating (data not shown) shows a similar effect, which is caused by the absolute value in the formula for the activity. However, it is weak compared to the constant increase of the response.
The neuron’s small modulation ratio for drifting gratings \( F_1/F_0: 0.26 \) is equivalent to a translation-invariant response. Therefore, this neuron shares the properties of complex cells found in primary visual cortex (Skottun et al., 1991).

Looking at the spatial tuning properties of the cells in the entire population, we find that all neurons are selective to orientation (mean width at half height 35 degrees and the ratio of the response strength at preferred orientation to the response at orthogonal orientation is 11.2). Furthermore, the neurons are selective to spatial frequency with a median spatial frequency selectivity index of 59 (for a definition, see Schiller, Finlay, & Volman, 1976).

In the population, independent of the initial value, most exponents converge quickly toward a value close to two. Then, for most cells during the next iterations, the exponent changes only little. Sometimes, however, the exponent suddenly explodes and grows until it reaches our constraint of \( N = 15 \). Over the whole population, most cells acquire an exponent close to 2 (see Figure 1D right, red), although the distribution is fairly broad. This shows that the classical energy detector is an optimal solution for the given objective. Furthermore, in comparison to recent physiological results obtained in cat primary visual cortex, the distribution of exponents is similar (see Figure 1D right, gray, reproduced from Lau, Stanley, & Dan, 2002). An overview of the relation between the exponent of a neuron and its response properties for the whole population is given in Figure 1F. There, we plot the exponent versus the modulation ratio for drifting gratings. Note that cells with a low modulation ratio would be classified as complex cells. Responses of cells with a high modulation ratio are specific also to the position of a stimulus, as are simple cells in visual cortex. Cells that can be classified as complex cells are found only in the region of the exponent close to two. Cells

Figure 1: Facing page. (A) Subunits’ receptive fields of an example neuron as a function of the number of iterations. (B) The exponent \( N \) and the similarity index of the subunits for the example neuron during the optimization. The similarity index is the scalar product of the final subunit (after 100 iterations) with the subunit at a given iteration divided by the lengths of these vectors. The tick marks on the \( x \)-axis indicate 35, 55, and 95 iterations in correspondence to panel A. (C) Left: The circular grating patch used for quantifying the neurons response properties. Right: Response diagram of the example neuron on a color scale from bright red (high activity) to dark blue (no activity). (D) Left: The evolution of the exponents for the population. The tick marks on the \( x \)-axis indicate 35, 55, and 95 iterations in correspondence to panel A. Right: The histogram of the exponents after 100 iterations over the population is shown in red. Physiological data taken from Lau et al. (2002) are shown in gray. (E) Histograms of the final values for the exponents in the control simulations. Light gray: temporally white stimulus; black: spatiotemporal pink noise. (F) Exponent vs. modulation ratio for drifting gratings. The response diagrams of seven neurons are shown at the upper border. On the right border, all neurons are located with an exponent larger than 5.
with large modulation ratios, on the other hand, are found for all values of the exponent. Furthermore, the preferred orientation and spatial frequency are distributed homogenously.

To probe which properties of natural scenes are important for the stable development of the exponent and the spatial receptive fields, we perform two controls. Since the objective function is based on temporal properties of the neurons’ activity, we first use a stimulus without temporal correlations. This temporal white stimulus is constructed by shuffling the frames of the original video (see section 2), but preserves its spatial structure. After a few iterations, the exponents of all neurons diverge to either the upper or lower limit imposed (see Figure 1E, gray). The spatial receptive fields appear noisy and display weak tuning to orientation (median of the ratio of response at optimal orientation to response at orthogonal orientation is 3.7). This shows that the distribution of exponents obtained in the first simulation is not a property inherent in our network. Rather, a smooth temporal structure of the input is necessary for the development of both the exponent and the spatial receptive fields.

As a second control, we train the network on spatiotemporal pink noise (see section 2) having the same second-order statistics as the natural movie. While this stimulus lacks the higher-order structure of natural scenes, it has a smooth, temporal structure compared to the first control. During optimization, the exponents converge as in the original simulation. However, the histogram of final values (see Figure 1E, black) differs markedly from that in Figure 1D. Most neurons acquire an exponent slightly above one, with many neurons having a linear slope of their transfer function. Furthermore, most neurons resemble low-pass filters, and their orientation tuning is weak (median of the response at preferred orientation to response at the orthogonal orientation is 2.9). Thus, the second-order correlations of natural movies are sufficient to allow stable learning of the exponent, as well as organized spatial receptive fields. However, to achieve a good match to physiological results in terms of both spatial receptive fields and the exponent of the transfer function, the higher-order structure of the natural scenes is decisive.

4 Discussion

We report here a network simultaneously optimizing the linear kernel of the given neuron model and the exponent of the neurons’ transfer functions with respect to the same objective function.

The results of the study are interesting for the following reasons. First, by changing parameters controlling the transfer function, neurons can represent a much larger class of nonlinear functions and, given a fixed network size, are able to solve a larger class of problems. Second, nonlinearities are ubiquitous in sensory systems (Ghazanfar, Krupa, & Nicolelis, 2001; Shimegi, Ichikawa, Akasaki, & Sato, 1999; Anzai, Ohzawa, & Freeman, 1999;
Lau et al., 2002; Escabi & Schreiner, 2002; Field & Rieke, 2002) and a wide variety of effects contributes to nonlinear effects within a neuron (Softky & Koch, 1993; Mel, 1994; Reyes, 2002). Yet in most cases, we simply lack the knowledge for a quantitative specification. Therefore, to model these systems, networks are required that can adjust all parameters and do not require detailed a priori knowledge.

While most learning schemes still concentrate on linear kernels, a small number of studies address optimizing the neurons nonlinearity. Bell and Sejnowski (1995) start from an Infomax principle and derive a two-step learning scheme, which allows optimizing the shape of a (sigmoidal) transfer function to match the input probability density. This directly relates to the concept of independent component analysis, where the adaptation of the transfer function adjusts the nonlinearity of the system (Nadal & Parga, 1994; Everson & Roberts, 1998; Obadovic & Deco, 1998; Pearlmutter & Parra, 1997). Recently, this approach was applied to natural auditory stimuli comparing the resulting receptive field structure to physiological results (Lewicki, 2002). However, some of these theoretical derivations require an invertible transfer function, which is not the case in our network and might not apply in general for neurons that discard information, such as neurons having invariance properties. Extending this previous research, we optimize the weight matrix and the nonlinearity simultaneously at each iteration and compare the resulting distribution of the parameters to recent experimental data.

In our network, we find many cells that are selective to orientation and spatial frequency, have an exponent around 2, and have translation-invariant responses. Their properties match the classical energy detector model for complex cells (Movshon, Thompson, & Tolhurst, 1978; Adelson & Bergen, 1985). Thus, our results show how this classical model of complex neurons can develop in a generalized nonlinear model. Furthermore, in a recent study of the cat visual system, the nonlinearity of the computational properties of cortical neurons has been investigated (Lau et al., 2002). The measured exponent of the nonlinear transfer function shows a large variation. An exponent around 2 best explains the response of many neurons, while in other cells it ranges between 0 and high values. The overall distribution of exponents found in the study examined here matches these experimental results surprisingly well.

Acknowledgments

We thank Laurenz Wiskott for introducing the evaluation of receptive fields using the circular grating patch to us, and Mathew Diamond and Rasmus Petersen for valuable discussion of yet unpublished results. This work was supported by the Centre of Neuroscience Zürich (C.K.), the SNF (Grant No. 31-65415.01, P.K.) and the EU IST-2000-28127/BBW 01.0208-1 (K.P.K., P.K.).


Received September 4, 2002; accepted January 30, 2003.
Learning distinct and complementary feature-selectivities from Natural Colour Videos

Wolfgang Einhäuser¹, Christoph Kayser, Konrad P. Körding & Peter König

Institute of Neuroinformatics (UNI/ETH Zürich) Winterthurerstrasse 190, 8057 Zürich, Switzerland; Phone: +41-1-6353044; Fax.: +41-1-6353053

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

¹Corresponding author

Keywords: learning, colour, visual cortex, natural stimuli, temporal coherence

Abstract

Many biological and artificial neural networks require the parallel extraction of multiple features, and meet this requirement with distinct populations of neurons that are selective to one property of the stimulus while being non-selective to another property. In this way, several populations can resolve a set of features independently of each other, and thus achieve a parallel mode of processing. This raises the question how an initially homogeneous population of neurons segregates into groups with distinct and complementary response properties. Using a colour image sequence recorded from a camera mounted to the head of a freely behaving cat, we train a network of neurons to achieve optimally stable responses, that is, responses that change minimally over time. This objective leads to the development of colour selective neurons. Adding a second objective, de-correlating activity within the network, a subpopulation of neurons develops with achromatic response properties. Colour selective neurons tend to be non-oriented while achromatic neurons are orientation-tuned. The proposed objective thus successfully leads to the segregation of neurons into complementary populations that are either selective for colour or orientation.
**Introduction**

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

Recently, several learning schemes have been proposed that link the response properties of visual neurons to statistical properties of the visual input. The principle of sparse coding, for example, can predict how visual experience gives rise to the receptive field properties of simple (Olshausen & Field 1996, Bell & Sejnowski 1997, van Hateren & van der Schaaf 1998) and of complex (Hyvärinen & Hoyer 2000) neurons in primary visual cortex. At
the highest level of the visual system, neurons in inferotemporal cortex also appear to form sparse representations (Rolls & Tovee 1995). Thus, sparse representations may be ubiquitous in the visual system.

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

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

**Methods**

**Stimuli**

We recorded natural image sequences using a removable lightweight CCD-camera attached to the head of a freely behaving cat, as described in detail elsewhere (Einhäuser et al. 2002). The cat is accompanied by an animal handler, but is otherwise free to explore an outdoor
environment as it chooses. All procedures are in compliance with Institutional and National guidelines for experimental animal care.

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

**Neuron models**

The activity of each neuron is computed as

\[ A_j = \varphi \left( \sum_j W_{ij} I_j \right) \]

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

**Objective functions**

The objective function used in this study is known as temporal coherence, which we implement by minimising the squared temporal derivative of the neurons’ activity. To avoid
the trivial solution of a neuron whose weights are all equal to zero, this squared derivative is normalized by the variance of a neuron’s output over time, yielding:

$$
\Psi_{\text{stable}} = \sum_{i} \psi^{\text{stable}}_i = \sum_{i} -\left( \frac{d}{dt} A_i(t) \right)^2 \text{var}_i(A_i(t)) \tag{1}
$$

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

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

$$
\Psi_{\text{decorr}} = -\frac{1}{(N-1)^2} \sum_{i,j} \sigma_{ij}(t) \tag{2}
$$

\(\sigma_{ij} = \frac{(A_i - \text{mean}(A_i))(A_j - \text{mean}(A_j))}{\text{var}(A_i) \text{var}(A_j)}\) denotes the coefficient of correlation and \(N\), the number of neurons (\(N=200\)).

The total objective to be optimised is then defined as

$$
\Psi_{\text{total}} = \Psi_{\text{stable}} + \beta \Psi_{\text{decorr}}
$$

where \(\beta\) scales the contribution of the de-correlation term.
Optimization

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

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

Analysis

We analyse the properties of the receptive fields in the network after convergence. First, the neuron’s receptive field representation in input space is obtained by inverting the PCA on the weight matrix, $W$. Then, each receptive field is scaled individually such that the values of each pixel fall between 0 and 1. The scaling is the same for each colour channel and thus does not bias either the chromatic or the spatial properties of the optimised receptive field. To analyse colour and spatial content independently, the receptive fields are transformed from a colour channel representation (RGB format) into a representation separating hue, saturation and brightness channels (HSV format), achieved using a standard function in MATLAB (‘rgb2hsv.m’).
We characterise the chromatic selectivity of a receptive field by calculating the mean saturation across pixels. Neurons with a mean saturation greater than 0.2 are classed as chromatic neurons, while the remaining neurons are classed as achromatic. Spatial properties of the receptive field are assessed by measuring the anisotropy of the receptive field using standard methods (Jähne 1997). Briefly, the tensor of inertia is computed on the values from the brightness channel. Anisotropy is defined as the ratio of the difference between the tensor’s long and short principal axis, divided by their sum. This measure is 0 for an isotropic (non-oriented) receptive field and approaches 1 for a perfectly oriented receptive field. Neurons with an anisotropy value less than 0.2 were considered to be non-oriented.

Results

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

The stability objective ($\Psi^{stable}$) does not include interactions between neurons in the network. The de-correlation term ($\Psi^{decorr}$) force neurons to acquire different receptive fields. As stability and de-correlation are competing mechanisms being added into a single objective function, neurons with sub-optimal stability emerge as well as neurons with suboptimal de-correlation. As one expects, activities of neurons with high stability values $\psi_{i}^{stable}$, tend to be more correlated than those with low stability values (Figure 2c).
Visual inspection of the neurons in figure 2b already suggests that chromatic neurons tend to have higher values of $\psi_{\text{stable}}$ than achromatic neurons. Quantitative analysis reveals that chromatic neurons indeed have higher stability (mean $\psi_{\text{stable}}$, -0.0041) than achromatic neurons (mean $\psi_{\text{stable}}$, -0.0053). There is a pronounced relation between a neuron’s chromaticity and its individual stability (correlation coefficient: $r=0.70$, Figure 3a). This demonstrates that chromatic neurons have optimally stable responses to natural stimuli, while the achromatic stimuli are a consequence of sub-optimal stability due the de-correlation objective.

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

Our results show that achromatic neurons emerge as a consequence of adding a de-correlation term ($\Psi_{\text{decorr}}$) to our objective function. The de-correlation objective permits interactions between neurons in the network, and results in receptive fields that are sub-optimal with respect to the stability objective alone. We went on to test whether increases in the relative contribution of the de-correlation objective (using the term $\beta$, see Methods) would further increase the fraction of achromatic neurons in the network. The limit case of $\beta \to \infty$ is simulated by omitting stability from the objective function altogether. For the simulations
using the linear-threshold model, the population average of mean saturation (chromaticity) is reduced from 0.50 where $\beta=0$ (no contribution from $\Psi^{\text{decorr}}$) to 0.19 where $\beta \rightarrow \infty$ (no contribution from $\Psi^{\text{stable}}$). The results from the simulations using the full-wave rectifying neurons are qualitatively similar, with a reduction in mean saturation from 0.5 ($\beta=0$) to 0.25 ($\beta \rightarrow \infty$). For both neuron types a 50% drop in mean saturation is achieved for values of $\beta=20$ (Figure 4a).

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

Next, we investigate the robustness of our model to reductions in the dimensionality of the training input. We determine the change in mean saturation when as few as 25 principle components were used, roughly 1/8th of the principle components used in the main simulations (see Methods). The change in mean saturation across both linear and full-wave rectifying models, and for values of $\beta$ between 5 and 10, was not larger than 53% (Figure 4b). This minimal dependence on the number of input dimensions contrasts with a study by Hoyer & Hyvärinen (2000) using independent component analysis, which reported a strong correlation between input dimensionality and the proportion of colour selective neurons. Increasing input dimension by a factor of 2.5 yielded an increase in the number of chromatic neurons of approximately 290% (estimated from Figure 10 in Hoyer & Hyvärinen 2000).
Doubling the number of independent components (from 100 to 200 dimensions) yielded a 77% increase in chromatic neurons, while in our simulations, which use PCA, the same increase in dimensionality produces an increase of just 11% (Figure 4c). This indicates that in simulations using ICA, the emergence of colour selective neurons depends on dimensionality of the input, while in simulations using the stability objective, colour selectivity is determined by the relative strength of the de-correlation objective.

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

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

We also determined the effect of shuffling the stimulus pairs on the emergence of colour selectivity in the network. Mean pixel saturation was 0.19 in the network trained on the shuffled stimuli, half the value that is obtained when the natural temporal order is intact (0.38). This resulted in the proportion of chromatic neurons in the network falling from 65%
in the naturally ordered condition to just 19% in the shuffled condition. Furthermore, most
eurons in the simulation using the linear-threshold model are close to the
chromatic/achromatic threshold of 0.2 (96% fall between 0.1 and 0.3 mean pixel saturation),
whereas for naturally ordered condition, mean saturation shows a wider distribution (43% between 0.1 and 0.3). While the change in mean saturation is less pronounced for the full-
wave rectifying model (0.24 compared to 0.29), a similar narrowing in chromaticity is
observed (63% compared to 30% of neurons fall between 0.1 and 0.3 mean saturation, Figure 4d). This indicates that in the shuffled condition, most neurons cannot be clearly identified as either chromatic or achromatic.

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

**Discussion**

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

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

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

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

Earlier studies using the temporal coherence objective to learn invariant responses delivered a clear proof of concept using artificial (Földiak 1991) as well as artificially
transformed natural stimuli (Stone 1996; Wallis & Rolls 1997). This studies were motivated by the idea that, if different transformations of the visual input naturally happen on different timescales, selectivity and invariance towards these transformations can be extracted using solely this fact. For example, the invariance to position and selectivity to orientation results from local orientations being correlated over longer time scales than positions (Kayser et al. 2003). As the temporal structure of the stimulus is therefore the decisive property exploited by this type of objective function, it is crucial that we demonstrate that these results are confirmed when using stimuli that preserve temporal structure in the natural input. Our unique stimuli, derived from recordings made from a camera mounted on a freely behaving cat, provide image sequences that preserve the natural temporal structure. Note however, that we found it technically infeasible to use more realistic colour representations, as the current sampling rates of devices capable of recording hyperspectral images are several orders of magnitude below the correlation time constants observed in our natural videos (Kayser et al. 2003). Thus, we were not able to capture this aspect of the natural input. Note that the results of the ICA studies using hyperspectral images (Wachtler et al. 2001) are not qualitatively different to those in which the standard RGB-representation was used (Tailor et al. 2000, Hoyer & Hyvärinen 2000). Thus, we can be confident that the RGB format is adequate for our purposes. Furthermore, the present ‘Catcam’ videos were not recorded at sufficient temporal resolution to model the temporal properties of receptive fields. Modelling chromatic spatio-temporal receptive fields with the stability objective thus remains an interesting issue for future research.

There is physiological evidence that colour-sensitive neurons in primate V1 tend to be non-oriented, whereas achromatic neurons tend to exhibit the precise orientation tuning seen to be characteristic of V1 simple and complex neurons (Gouras, 1974; Lennie et al. 1990). This is further supported by psychophysical experiments that show that humans can resolve higher spatial frequencies for isochromatic patterns than for isoluminant patterns (Webster et
This raises the issue as to how our particular objective function might be implemented physiologically. A possible mechanism for the de-correlation objective is provided by inhibitory lateral connections in primary visual cortex. Physiologically, this process does not necessarily require synaptic changes at the tangential connections, but can also exploit modifications to afferent synapses that are in turn influenced by lateral interactions (Körding & König 2000). This mechanism utilizes action potentials propagating retrogradely through the dendritic tree (Stuart & Sakmann 1994). Building on the same physiological mechanism, a temporally asymmetric learning rule (Makram et al. 1997, Larkum et al. 1999) could plausibly subserve the function of the stability objective (Einhäuser et al. 2002).

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

**Acknowledgements**

This work was financially supported by Honda Research Institute Europe, the Neuroscience Center Zurich, the Swiss National Science Foundation (31-61415.01) and EU/BBW (IST-2000-28127 / #01.0208-1). We thank T.C.B. Freeman for comments on previous versions of this manuscript.
References


Captions

Figure 1
Three examples of the two consecutive images that are selected from the natural image sequence. Each pair of images forms the basis of the input to our model. The white squares indicate three of the twenty 30x30 pixel patches that are sampled from the same location in each frame pair.

Figure 2
(a) Development of objective function $\Psi^{\text{total}}$ in the course of the optimisation process for a simulation of full-wave-rectifying neurons with $\beta=5$.
(b) Receptive fields produced following 60 iterations of the simulation shown in (a). The receptive fields are sorted according their individual stability value, with low stability values at the top of the figure and high stability values at the bottom.
(c) Coefficient of correlation between all the activities over one stimulus presentation of the neurons of panel b.

Figure 3
(a) Dependence of neuron chromaticity on individual objective value $\psi_i^{\text{stable}}$. Chromaticity was assessed by calculating mean saturation for each receptive field, averaged across pixels.
(b) Chromaticity versus anisotropy for $\beta=5$ and full-wave rectifying neurons.
(c) Chromaticity versus anisotropy for $\beta=5$, linear-threshold model.

Figure 4
(a) Mean chromaticity versus $\beta$ for the full-wave rectifying model (stars) and linear-threshold model (circles).
(b) Dependence of mean neuron chromaticity on PCA dimension for different values of $\beta$ and the two neuron models.
(c) Percentage of chromatic neurons (mean saturation > 0.2) of (a) compared to results from ICA, redrawn from Hoyer & Hyvärinen (2000).

(d) Distribution of mean pixel saturation for the control condition of shuffled stimuli (left) compared to stimuli that preserve the natural temporal structure (right). Top row shows results from the linear-threshold model, bottom row, the full-wave rectifying model.
Einhäuser et al. Figure 1
Einhäuser et al. Figure 2
Einhäuser et al. Figure 3
(a) $\varphi(x) = |x|$, $\varphi(x) = \max(x,0)$

(b) $\varphi(x) = \max(x,0)$, $\beta = 5$

(c) $\varphi(x) = |x|$, $\beta = 5$

(d) $\varphi(x) = |x|$

Einhäuser et al. Figure 4
How are complex cell properties adapted to the statistics of natural stimuli?

Konrad P. Körding*12, Christoph Kayser2, Wolfgang Einhäuser2 & Peter König2

(1) Institute of Neurology, UCL London, Queen Square, London, WC1N 3BG, United Kingdom.
(2) Institute of Neuroinformatics, University of Zürich and ETH Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland
email: {koerding, kayser, weinhaeu, peterk}@ini.phys.ethz.ch

Correspondence should be addressed to K.P.K (koerding@ini.phys.ethz.ch, Tel +44 (0)20 7837 3611 x4404, Fax +44 (0)20 7813 3107)

Running head: How are Complex Cells adapted to natural scenes?

Word count: 26,409 characters
Abstract:

Sensory areas should be adapted to the properties of their natural stimuli. What are the underlying rules that match the properties of complex cells in primary visual cortex to their natural stimuli? To address this issue we sampled movies from a camera carried by a freely moving cat, capturing the dynamics of image motion as the animal explores an outdoor environment. We use these movie sequences as input to simulated neurons. Following the intuition that many meaningful high-level variables, e.g. identities of visible objects, do not change rapidly in natural visual stimuli, we adapt the neurons to exhibit firing rates that are stable over time. We find that simulated neurons, which have optimally stable activity, display many properties that are observed for cortical complex cells. Their response is invariant with respect to stimulus translation and reversal of contrast polarity. Furthermore, spatial frequency selectivity and the aspect ratio of the receptive field quantitatively match the experimentally observed characteristics of complex cells. Hence, the population of complex cells in the primary visual cortex can be described as forming an optimally stable representation of natural stimuli.

Keywords: Primary Visual Cortex, Cat, Receptive fields, Unsupervised Learning, Sparseness, Stability,
Introduction

Most neurons in the primary visual cortex can be classified into one of two generic cell types. The simple cells respond selectively to bars and gratings presented at a specific position, orientation, spatial frequency and contrast polarity (Hubel and Wiesel. 1962; Schiller et al. 1976b). The neurons of the other type, complex cells, also respond to bars or gratings of adequate orientation and spatial frequency. They, however, respond equally well regardless of the contrast polarity of the stimulus and its precise location within the region of the receptive field (Hubel and Wiesel. 1962; Kjaer et al. 1997).

The properties of sensory neurons, including the complex cells, can be expected to be well adapted to the statistics of the stimuli they are exposed to under natural conditions.

The most prominent hypothesis of how neural properties should be adapted to the statistics of natural scenes is called “sparse coding”. It states that sensory neurons should be selective to specific features, only responding strongly to a small subset of stimuli, but otherwise showing low activities (Barlow 1961; Fyfe and Baddeley 1995; Olshausen and Field 1996). This theory could well explain the properties of simple cells in primary visual cortex (Bell and Sejnowski 1997; Olshausen and Field 1996; Van Hateren and van der Schaaf 1998).

Under what assumption about the objective of adaptation do simulated neurons develop the same properties as complex cells? To derive such an objective we start with the insight that it is one of the tasks of the brain to extract relevant sensory features (Barlow 1961). Relevant variables, such as the description of a visual scene in terms of objects, change on a slower time scale than low level features, such as luminance in a small spatial region. If we, for example, see an animal such as a tiger, it usually stays around for some time. However, the position of the image of its stripes on the retina changes on a shorter timescale. Such insight has lead to the development of criteria that measure the stability or temporal coherence of the responses of simulated neurons (Becker 1999; Einhäuser et al. 2002; Földiak 1991; Kayser et al. 2001; Klopf 1982; Stone and Harper 1999; Sutton and Barto 1981; Wallis and Rolls 1997; Wiskott and Sejnowski 2002). These studies have successfully applied this criterion to the representations of artificial stimuli such as moving bars to establish that such a mechanism could lead to complex-type neurons (Földiak 1991; Wiskott and Sejnowski 2002). However, by using such simple stimuli the population of neurons does not obtain a rich enough distribution to be thoroughly compared to physiology.
Here we apply a similar stability criterion to the representations of natural stimuli. We then compare the resulting neuronal response properties, i.e. their selectivity to orientation and spatial frequency as well as their response modulation and aspect ratio, to those of complex cells in primary visual cortex.

Methods

Stimuli

We study the response properties of simulated neurons after adaptation to image sequences of natural scenes. A freely moving cat explores the forest located next to the campus in Zürich while carrying a miniature CCD camera (for details see Einhäuser et al. 2002) on its head that samples the natural visual input. This procedure is carried out in accordance with institutional and national guidelines of animal care. A video of 3000 frames, recorded at 25 frames per second, digitized at a resolution of 4.5 pixel/deg and converted to grayscale using the MATLAB rgb2gray function, is used for this study. Ideally we would like to take a single long sequence from the central region of the video. Such a sequence however would need to be prohibitively long to uniformly sample the stimulus material. That is why we instead take pairs of patches measuring 30 by 30 pixels from randomly selected, but matching locations within two subsequent frames in the movie. Temporal coherence is evaluated between the patches of the same pair, approximating the optimal sampling process. The patches are first multiplied pointwise with a Gaussian kernel centered over the patch whose standard deviation (width) was 10 pixels. This procedure has a limited effect on the amount of information available in the input stream, but avoids edge effects and the anisotropy inherent in square patches. Repeating the simulations below without this windowing leads to qualitatively similar results (data not shown). The receptive field obtained in such simulations are localized, do not cover the full patch and are approximately round too. The resulting patches are decomposed into their principal components. The first component, representing the mean patch brightness, is removed. Components 2 through 100 carry more than 95% of the variance and define a vector $I$, which defines the input to the optimization algorithm. As the activity of each subunit linearly depends on the input, the preprocessing of the input by a principal component analysis, which is also linear transformation, has no influence on the optimization process. Discarding the higher order components, however, does have an effect. As these components carry only a small part of the total variance we do not expect an influence of this step on the results obtained. Indeed, this assumption is supported by the results of a recent study (Kayser et al. 2001). On the positive side, as the number of
dimensions of the optimization problem is reduced by a factor of 9 a significant increase in computational efficiency is achieved.

**Simulated Neurons**

Complex cells, in contrast to simple cells, display several strong nonlinear properties (Chance et al. 1999; Movshon et al. 1978; Ohzawa et al. 1997; Spitzer and Hochstein 1988). Hence, it is not possible to describe them adequately by linear models, and we have to consider nonlinear model neurons. Identical to the choice in a number of other studies (e.g. Hyvärinen and Hoyer 2000) we chose the the two subunit energy model (Adelson and Bergen 1985; Hyvärinen and Hoyer 2000).

Each such model neuron consists of two subunits (Figure 1A). Each of the subunits computes the scalar product of the same input patch (I) with a weight vector ($W_{1,i}$, $W_{2,i}$ respectively). Hence each neuron is characterized by two linear receptive fields. Both outputs are subsequently squared and summed to define the neurons activity:

$$A_i = \sqrt{(W_{1,i}I)^2 + (W_{2,i}I)^2}.$$

These simulated neurons can, given appropriate weights, exhibit a large variety of response properties. Most of these properties are never observed for real neurons. The simulated neurons can however also act like a complex cell if both subunits have Gabor-wavelet like receptive fields with identical orientation and spatial frequency, and the two wavelets have a relative phaseshift of 90° (Figure 1B). If such a neuron is excited by a visual stimulus in form of a bar that is moved over it’s receptive field, each subunit has an activity that depends on the bar’s position. As the bar is shifted the subunits alternate in having large squared activity. Thus, the neurons activity, the sum of the squared subunits activities, changes only little as the bar is moved within the receptive field. Given the large number of parameters (twice the length of the weight vector) involved in determining the response properties of these model neurons, such complex cell like properties are only one among many other conceivable outcomes.

**Optimization**

The input consists of image patches that are extracted from successive frames of the movies. To simulate the adaptation process we optimize the parameters of a population of 100 neurons so that their responses are maximally coherent over time while being decorrelated from one another. This is done by maximizing the following objective function:
Here, <> denotes the average over all stimuli, and thus over time; $\bar{A}_i$ is the activity of neuron $i$ at time $t$, minus its mean over all times. $\Psi_{\text{stable}}$ takes on large negative values if the output activities change fast. It thus punishes fast temporal variations. The 40 ms lag between two successive time points used in that objective function is well within the range of strong correlations of orientations in natural stimuli (Einhäuser et al. 2002). $\Psi_{\text{decor}}$ on the other hand takes on large negative values in the case of correlated activities of different neurons and thus punishes such correlations. The average squared value of each subunit’s activity is multiplicatively normalized to be one each iteration of the algorithm.

The parameters of the model neurons are optimized by scaled gradient descent. For $\Psi_{\text{stable}}$ this leads to a local Hebb-type learning rule. The weight change is local to the synapse and depends only on pre- and postsynaptic activities at two subsequent points in time.

We furthermore compare our results to the work of Hyvärinen and Hoyer (2000). In this work they simulate a set of optimally sparse neurons that are modeled as 4-subunit energy models. All subunits are constrained to have uncorrelated output thus effectively enforcing a phase shift of 90 degrees. We repeat their simulations, using their code with our data as input. In this simulation 24 energy detector neurons with four subunits are used. We also perform a number of control simulations where we substitute $\Psi_{\text{stable}}$ with one of a number of alternative definitions of sparseness.

**Data analysis**

In analogy to physiological experiments we characterize the response properties of the model neurons by several indices. The orientation tuning width is calculated as the range of orientations for which the response to a bar of optimal position is above $1/\sqrt{2}$ of the maximal activity. The best orientation $\varphi$ is defined as the stimulus orientation that leads to maximal responses. The selectivity for spatial frequency is defined via the range of spatial frequencies to which the response exceeds $1/\sqrt{2}$ of the maximal level (Schiller et al. 1976b). The difference between the lower and upper bound of this range is then multiplied by 100. We measure the responses of neurons to drifting sinusoidal gratings of optimal orientation and spatial frequency. The neurons AC/DC-ratio is the maximum minus the minimum divided by the mean of the resulting activity.
The models that are used for the modeling of complex cells, such as the two subunit energy model used here always respond to moving gratings with twice the temporal frequency of the moving grating as they respond equally well to bright and dark edges. This implies that the simulated neurons have a vanishing first harmonic (F1) while the second harmonic (F2) does not vanish. Real complex cells however show such frequency doubling only to a limited degree, and both components are small (Spitzer and Hochstein 1985; Heeger 1992). How should the AC/DC ratios of such simulated neurons be compared to the relative modulation of real neurons? Either we could compare the AC/DC ratio to the F2/F0 ratio of real neurons, assuming that the frequency doubling is just an artifact of the simulation method. Alternatively we could compare the AC/DC ratio of the simulated neurons to the F1 of the real neurons, which is the preferable method to distinguish complex cells from simple cells. In this scenario followed in this paper the simulated neurons should have small AC/DC ratio, compared to the relative modulation of real neurons.

The envelope of the receptive field is defined as:

\[ E_i(x, y) = W_{i,1}(x, y)^2 + W_{i,2}(x, y)^2 \]

The length \( L_i \) and width \( V_i \) (defined via the standard deviations) of the receptive field is calculated (using the abbreviation \([.] = \max(., 0)\)):

\[ L_i = \sqrt{\sum_{x,y} (x \sin(\theta) + y \cos(\theta))^2 \left[ E_i(x, y) - 0.5 \text{std}(E_i) \right]} \]

\[ V_i = \sqrt{\sum_{x,y} (x \cos(\theta) - y \sin(\theta))^2 \left[ E_i(x, y) - 0.5 \text{std}(E_i) \right]} \]

Where \( x \) and \( y \) are the positions relative to the center of gravity of the receptive field. The aspect ratio is defined as \( L_i / V_i \). The subtraction and rectification prevents points with low values, lying far from the receptive field, from strongly influencing the aspect ratio. This is comparable to removing values below the noise level in physiological experiments. Histograms are compared using a one-sided Kolmogorov-Smirnov (KS) test yielding the probability of both histograms being drawn from the same distribution.

**Parametric studies**

In parametric studies we characterize the dependence of \( \Psi_{\text{stable}} \) on the receptive field properties. To elucidate why sparse coding alone is not expected to result in complex cell type responses we also measure the dependence of a specific definitions of sparseness on the receptive field properties:
We repeat this simulation with the objective function derived from the Cauchy prior and the standard deviation obtaining essentially the same results. We use the same two-subunit model as in the optimization procedures above albeit with simplified receptive fields. Since the optimization methods result in Gabor type receptive fields and neuronal receptive fields are well approximated by these, we choose the subunits to be Gabor wavelets of fixed orientation and spatial frequency. The phase and aspect ratio of each subunit, however, remain free parameters:

\[ G(a, s, s_x, s_y, x, y) = \sin(180x/a + s) \exp(-x^2/(as_x)^2 - y^2/(as_y)^2), \]

where \(a\), which is fixed to a value of 5 pixels, is the size of the Gabor, \(s\) is the relative shift between the subunits, \(s_x\) and \(s_y\) are the relative length and width and \(x\) and \(y\) the relative positions of the pixels. For Figure 4B and C we choose identical shapes: \(W_1 = G(5, 0, 1, 1)\), \(W_2 = G(5, s, 1, 1)\) and vary the shift, \(s\), between the subunits. For Figure 4C we choose a fixed shift of 90°: \(W_1 = G(5, 0, \lambda, w)\), \(W_2 = G(5, 90°, \lambda, w)\) and vary length, \(\lambda\), and width, \(w\), between 0.5 and 4 in steps of 0.1. Aspect ratios are binned in steps of 0.2 between 0.2 and 5.

**Results**

We simulate neurons and adapt them to display optimally stable activity over time. The resulting response properties are characterized by the receptive fields of their two subunits (Figure 2A). Most of the subunits exhibit a receptive field that is well described by a Gabor wavelet. They thus have receptive fields that are localized in the visual space and that are selective to orientation and spatial frequency. Most neurons exhibit a phase shift between the Gabor wavelets representing the receptive fields of each of its subunits that is close to a quarter cycle (90 degrees). This suggests that the response properties of the simulated neurons exhibit some translation invariance (sketched in Figure 1B), a key property of complex cells. The neurons are furthermore tuned to orientation and spatial frequency (Figure 2B,C see also Webster and De Valois 1985).

In the following we quantitatively compare the simulated neurons’ responses to bars and gratings to those of real neurons. First we investigate the orientation specificity. In response to a bar of optimal width the population of optimized neurons displays a narrow orientation tuning (38° width, Figure 2A). This specificity is somewhat tighter than the tuning width of real complex cells.
The simulated neurons also exhibit a tight tuning (index of 51.9) to spatial frequency comparable to the tuning index of cortical neurons (average index of 46.9, Schiller et al. 1976b), although the small difference is significant (\(p<0.01\) KS-test).

Next we compare real and simulated neurons on the basis of their response to moving gratings. In primary visual cortex, a bimodal distribution of relative modulation strengths is observed (Skottun et al. 1991) (Figure 3C). Complex cells are defined as having a relative modulation below 1.0, while simple cells are defined by larger values of the modulation ratio. In our simulations a wide bimodal distribution of AC/DC values is also observed. The AC/DC ratios of the optimally adapted complex cells have a mean (0.41) that is not significantly larger than the experimentally observed relative modulations (0.40, \(p>0.3\) KS-test).

Last we compare the aspect ratios of the receptive fields, defined as the ratio of its width relative to its length. Real complex cells have an aspect ratio of 1.02 ± 0.2 (Ohzawa and Freeman 1997) (Figure 3D). The optimally adapted neurons have an aspect ratio of 1.09±0.3, closely matching the experimental values (\(p>0.3\), t-test).

AC/DC ratio and aspect ratio define the invariant processing performed by complex cells. Thus, the simulated neurons with optimally stable activity result in good fits to the measured properties of complex cells in the primary visual cortex.

It has been proposed that combining sparse coding with appropriate boundary conditions also leads to complex cells (Hyvarinen and Hoyer 2000). We repeat that simulation using our stimulus database. This simulation yields neurons with an orientation selectivity of 37° and a spatial frequency selectivity of 40.5, both well in the range of the physiological values (56°, 46.9 respectively) and comparable to optimizing a stability objective (38°, 51.9 respectively). For the AC/DC ratio this simulation, however, results in a value of 0.65 that is far larger than the physiological value (0.40) and the result of optimizing a stability objective (0.41) (\(p<0.001\) KS-test). Thus, combining a sparseness objective with additional boundary conditions does not result in sufficiently translation invariant neurons. Furthermore, the aspect ratio of 1.73 is far larger than the one observed for real complex cells (1.02, \(p<0.001\) t-test). Similar results and equally significant deviations are found if we exchange \(\psi_{\text{stable}}\) in our simulations by the objective function derived from a Cauchy prior as used by Olshausen and Field (1996) or the Kurtosis. This suggests that only the objective of stability adequately explains the properties of complex cells.

The head mounted camera does not register changes in gaze associated with movements of the eyes. However, recent results indicate that under the conditions the stimuli were recorded eye movements contribute little to stabilizing
the retinal image (Möller et al. 2003). To control for possible residual stabilizing effects of eye-movements we perform two experiments: (1) We simulate eye movements that randomly stabilize 50% of the patches. (2) We randomly shuffle 10% of the patches. The resulting receptive field properties are essentially unchanged in both cases. In particular in both cases they are translation invariant and have AC/DC ratios close to the relative modulation of physiological data (p>0.3 for both controls, KS-test). Therefore we do not expect major changes of the reported results if eye-movements of the cats under free viewing conditions were taken into account.

To investigate if the results generalize to a more general nonlinear model or if the results are due to the way we constructed our model neurons we perform an additional simulation (Figure 4A). Simulated neurons consisting of 8 half-squaring subunits are modeled. The neural properties resulting from optimizing $\Psi_{stable}$ are similar to those found for the two-subunit energy model described above. Importantly, the AC/DC ratio distribution is not significantly larger than the relative modulations of real complex cells (p>0.3, KS-test). Thus, the results do not critically depend on the constraints on the model neurons’ nonlinear properties defined by the two-subunit energy model. The type of the nonlinearity is set in our simulations. For the neurons to exhibit complex cell properties however the subunits need to obtain identical orientation and spatial frequency as well as the right phase shift. This simulation thus shows that these properties can be obtained from natural scenes even for varied neuron models.

To better understand the above results we proceed to characterize some important nonlinear statistical properties of videos natural scenes. To do so we measure the objective values of simulated neurons in response to the videos of natural scenes. We choose the subunits of the same model as above to be Gabor wavelets of fixed orientation and spatial frequency, leaving the aspect ratio and the relative phase as free parameters. With this more restricted set of subunit receptive fields, we can systematically analyze the influence of the receptive field properties on various objective functions. Varying the relative phase of the subunits reveals that $\Psi_{stable}$ is maximal if the simulated neuron is translation invariant and the wavelets have a relative phase of 90 deg (Figure 4B). Neurons then represent localized oriented energy detectors and are translation invariant, as are real complex cells. We furthermore analyze the influence of the aspect ratio on the objective functions (Figure 4C). $\Psi_{stable}$ reaches its highest value for spherical receptive fields with an aspect ratio of about 1 similar to the value of real complex cells (Ohzawa and Freeman 1997). For comparison with other studies we also plot sparseness as a function of phase and aspect ratio which peaks at values that are far from those found in physiology. It thus seems that stability is a good candidate
for an adaptation criterion that links complex cells with the statistics of natural scenes.

**Discussion**

We have show that adaptation to a stability objective leads to simulated neurons sharing important spatial properties of complex cells in the primary visual cortex. Sparseness can be derived from several ideas such as minimizing energy consumption, optimal channel coding or searching for a meaningful representation of data. Stability can also be derived from various ideas: High level variables such as object identities are stable, stable variables can be transmitted through channels with lower bandwidth and learning is easier in a system where variables change slowly.

Recently Hurri and Hyvarinen (2003) have proposed that optimizing stability of linear neurons in response to natural stimuli leads to receptive fields like those of simple cells. The stability of linear neurons however is always considerably lower than the stability of the nonlinear complex cells in our study. The authors furthermore use a slightly different objective that biases the neurons to be both stable and sparse. These results might still indicate that both simple and complex cell responses could be understood in a coherent framework derived from the idea of stability.

In our simulations each neuron only saw the input stimulus windowed by a Gaussian. Parts of the properties of the neurons, in particular the aspect ratio could thus be affected by this preprocessing. Some of the simulated neurons however do have receptive fields that are smaller than the size of the Gaussian. There is a tendency for neurons to obtain localized receptive fields. It would be interesting for future studies to analyze if the distribution of receptive field sizes can be obtained exclusively from optimizing stability. Such studies would however need very large numbers of simulated neurons as they would need to jointly encode the retinal space in addition to the orientation and spatial frequency space.

Do neurons found in primary visual cortex exhibit sparse or stable or maybe both types of response properties? Both objectives seem useful for processing in the nervous system. The question of which objective links the properties of natural scenes to the properties of complex cells is experimentally accessible. On one hand, for these analyses recordings from neurons in response to natural scenes would need to be compared to response to artificial stimuli such as bars or gratings. With respect to sparseness some experiments started to
address this issue (Baddeley et al. 1997; Vinje and Gallant 2000). If a large set of natural visual patterns is presented in sequence most of these are not effectively stimulating the recorded neuron. A small subset of stimuli, however, can activate the neuron strongly and elicit very high firing rates. Similar experiments could address how stable neural responses are.

The fact that complex cells of adult animals are well described as an adaptation to a stability objective raises the question, whether this adaptation occurs on onto- or phylogenetic timescales. If there is an ontogenetic component to the development of complex cells, it allows the following experimental test of the stability hypothesis. Changing the environment during an animal’s critical period (e.g. by strobe rearing) would impair the development of complex cell type receptive fields. In particular there should be a range of strobe rates in which complex cells are severely affected, while simple cells are not. From measurements of correlation times in natural videos (Kayser et al. 2003) this rate is expected to be of the order of 10Hz.

If simple cells optimize a sparseness criterion and complex cells optimize a stability criterion, it is tempting to speculate, whether such a division of labor is repeated in higher areas. Indeed in a widely used architecture for invariant object recognition, the Neocognitron (Fukushima 1980), a hierarchical network with an alternation of simple and complex type cells is used. Hence it is interesting to build larger systems consisting of several layers, each optimizing an adequate objective. This could result in a hierarchical system allowing to predict the response properties of neurons in higher cortical areas and to relate the response properties of such neurons to the statistics of the real world.

Acknowledgements

This work was supported by the Boehringer Ingelheim Fund and Collegium Helveticum (KPK), the Neuroscience Center Zurich (CK), Honda Research Institute Europe (WE), the Swiss National Science Foundation (PK, 31-65415.01) and the EU/BBW (IST-2000-28127 / 01.0208). We are grateful to T.C.B. Freeman, P. Dayan and B.A. Olshausen for comments on previous versions of this manuscript.
**Figure captions**

**Fig. 1.**
(A) Patches of size 30 by 30 are extracted from each frame, windowed by a circular kernel and fed into a network of neurons modeled as two subunit energy detectors. Each subunit sums up the common input weighted by its weightvector. The outputs are squared and added to obtain the output of the neuron. (B) The way such a model neuron can become translation invariant is sketched. Both subunits need to have receptive fields that are Gabor wavelets with a relative phaseshift of 90 deg. A bar of optimal orientation and spatial frequency is moved through the neurons receptive field. Whenever one subunit has a very positive or very negative activity the other one has an activity of about zero. The outputs of each subunit thus vary a lot while the sum of the outputs of the two subunits does not change much as the bar is moved.

**Fig. 2.**
Qualitative properties of the simulated neurons. (A) Pairs of receptive fields are shown of neurons with optimally stable activity. (B) The responses of two representative neurons to bars of changing orientations and widths (displayed at the optimal positions) are shown. (C) Responses of the same neurons to gratings of optimal phase and orientation but varying spatial frequency are shown. (D) The optimal orientation and optimal spatial frequency are plotted for all the simulated neurons.

**Fig. 3.**
Density distribution of properties of complex cells in primary visual cortex and of neurons with optimally stable activity. (A) The orientation tuning widths are shown for cortical complex cells in monkey cortex (Schiller et al. 1976a) and for the simulated neurons. (B) The selectivities to spatial frequency are shown for cortical complex cells in monkey cortex (Schiller et al. 1976b) and for the simulated neurons. (C) The relative modulation strengths are shown for a collection of 1061 cat complex cells replotted from (Skottun et al. 1991) along with the AC/DC ratio of the simulated neurons. (D) The distribution of aspect ratios is sketched for cat cortical neurons (Ohzawa and Freeman 1997). This is compared with the aspect ratios of the simulated neurons.

**Fig. 4.**
The influence of the parameters. (A) Neurons consisting of 8 halfsquaring \((f(x)=x^2 \text{ for } x>0, \ f(x)=0 \text{ otherwise})\) subunits are modeled. The histogram of
AC/DC ratios is shown for cat complex cells (Skottun et al. 1991) and for the optimized neurons with 8 halfsquaring subunits each. (B) Back in the simple two-subunit energy model, the objective functions $\Psi_{\text{stable}}$ (thick lines) and $\Psi_{\text{kurtosis}}$ (thin lines) are plotted as a function of the relative phase between subunits with Gabor shaped receptive fields. (C) Objective functions are plotted as a function of the aspect ratio. Error bars denote the s.e.m.

References


Figure 1

A

100 neurons

B

squared subunit activities S

neuron activity A
Figure 2

A

B

C

D

activity

orientation [deg]

spatial frequency [cyc/30 pixels]

activity

spatial frequency [cyc/30 pixels]

orientation [deg]
Figure 3

A

Physiology

Simulation

number of neurons
specificity for orientation [deg]

B

Physiology

Simulation

number of neurons

specificity to spatial frequency

C

Physiology

Complex

Simple

relative Modulation

Simulation

number of neurons

AC/DC ratio

D

Physiology

mean+2 std

Simulation

number of neurons

aspect ratio
Figure 4

A

B

C

\[ (x/a)^2 + (y/b)^2 = 1 \]
Extracting Slow Subspaces from Natural Videos Leads to Complex Cells

Christoph Kayser, Wolfgang Einhäuser, Olaf Dümmver, Peter König, and Konrad Körding

Institute of Neuroinformatics, ETH / University Zürich
Winternthurstr. 190, 8057 Zürich, Switzerland
{kayser,weinhaeu,olaf,peterk,koerding}@ini.phys.ethz.ch

Abstract. Natural videos obtained from a camera mounted on a cat’s head are used as stimuli for a network of subspace energy detectors. The network is trained by gradient ascent on an objective function defined by the squared temporal derivatives of the cells’ outputs. The resulting receptive fields are invariant to both contrast polarity and translation and thus resemble complex type receptive fields.

Keywords: Computational Neuroscience, Learning, Temporal Smoothness

1 Introduction

A large body of research addresses the problem of obtaining selective responses to a class of stimuli (e.g. Hebb 1949, Grossberg 1976, Oja 1982) but surprisingly few results exist on learning representations invariant to given transformations. But real world problems like recognition tasks do not only require the network to be specific to the relevant stimulus dimensions but also to be insensitive to the irrelevant dimensions (e.g. Fukushima 1988). In this paper we address the problem of learning translation invariance from natural video sequences, pursuing an objective function approach. We implement the temporal smoothness criterion as proposed by Hinton (1989) and used by Földiak (1991). A generative model containing slowly changing hidden variables is assumed. The effect of these hidden variables on linear subspaces can be described by a mixing matrix. This mixing matrix is inverted by the search for slowly varying subspace energy detectors. Instead of mathematically deriving the objective function for these subspaces from an explicit generative model we here explore the effect of a given function on learning of nonlinear detectors. We analyze the obtained slow components and compare them with properties of complex type receptive fields of cortical cells.

2 Methods

The stimuli used to train our network consist of randomly chosen 10 by 10 patches sampled from a natural video recorded by a camera mounted on a cat’s head.
head (Betsch et al. submitted). Patches from the same spatial location within the image are taken from 2 subsequent images yielding a pair of intensity vectors $I_{t-1}$ and $I_t$ (images are sampled at 25 Hz). Each vector is normalized to mean zero. The complete stimulus set, consisting of 11000 such pairs, is reduced in dimensionality by PCA and whitened using the procedure described in Hyvärinen and Hoyer (2000). If not stated otherwise, the number of used principal components is 30 (in the following termed PCA dimension).

For the reported results the network consists of 5 neurons each of which sums the input of 4 sub-units (Fig. 1). Each sub-unit has a weight-vector associated and the activity of sub-unit $j$ of neuron $i$ is calculated as the product $A_{ij} = W_{ij} \cdot I$. The neurons are modelled as subspace energy detectors (Kohonen 1996) and their activity is calculated as $A_i = \sqrt{\sum_j A_{ij}^2}$. The analyzed objective function is

$$O_{time} := - \sum_{\text{cells } i} \left( \frac{\langle \frac{d}{dt} A_i \rangle^2}{\text{var}_t(A_i)} \right)_t$$

where the mean $\langle \cdot \rangle_t$ and the variance are taken over time. In order to implement this in discrete time, the derivative is approximated by the difference of the activities for two consecutive patches, $A_i(t) - A_i(t-1)$. The variance is furthermore replaced by the product of the standard deviation taken over all the activities for the patches $I_{t-1}$ times the standard deviation for the patches $I_t$. The network learns by changing the sub-unit weights $W_{ij}$ following the (analytically calculated) gradient of $O_{time}$ to a local maximum. The gradient ascent is controlled using adaptive stepsizes as described in Hyvärinen and Hoyer (2000) till a stationary state is reached. All sub-units are forced to be orthonormal in whitened space. The weights are randomly initialized with values between 0 and 1. The network layout together with two typical stimuli is shown Figure 1.

In order to quantify the properties of the learned cells their orientation and position specificity is calculated and displayed in $\theta$-$r$ diagrams: The cells are probed with Gaussian bars of defined orientation $\theta$ and position $r$ as stimuli and the resulting activities displayed. From these diagrams two parameters are extracted: The orientation specificity index ($\sigma_\theta$) is computed as the mean width of the orientation tuning over all positions. The position specificity index ($\sigma_r$) is
computed by first taking the standard deviation of the activity over all orientations at a fixed position and then averaging over all positions.

3 Results

In order to explore learning of invariant detectors a nonlinear network is implemented (see methods). We use neurons that compute the 2 norm of the corresponding sub-unit activities (Fig. 1). On the activities of these neurons an objective function characterizing their temporal smoothness, $O_{\text{time}}$, is defined and the network is trained till a stationary state is reached (Fig. 2A).

The resulting receptive fields of the sub-units largely resemble those of simple cells (Fig. 2B). After training every neuron receives input from a set of sub-units which all share the same orientation preference but differ in spatial localization. This is shown by the $\theta$-$r$ diagrams for the sub-units (Fig. 2C). Thus the resulting neurons are insensitive to the position of the stimuli and are therefore translation invariant (Fig. 2D). The system is also invariant with respect to the contrast polarity of the stimuli: The response for a bright bar on dark background is the same as for a dark bar on bright background. Note that this contrast polarity invariance is not learned by the network but instead is a built in feature of the transfer function of the neurons (since an even norm is used).

As an important control it is necessary to check that translation invariance is indeed a consequence of the temporal smoothness of the stimuli and not also an inherent network property. The stimulus vectors are randomly shuffled to destroy the temporal coherence of the pairs $\{I_{t-1}, I_t\}$. Figure 3A shows the resulting receptive fields of the sub-units, which no longer exhibit the systematic properties of those obtained with the stimuli in natural order. This shows that the correlations in the time domain of the video sequences are necessary for the learning of the complex like receptive fields.

Since the temporal correlation between patches in natural videos decays gradually over time (Betsch et al. Submitted) we pair frames of larger temporal distances ($\{I_{t-\Delta_n}, I_t\}$ instead of $\{I_{t-1}, I_t\}$). As expected, with growing time shift $\Delta_n$ orientation specificity decreases and the cells become more specific to position (Fig. 3B). In the limit of no correlation (large temporal distances or randomly paired frames) position and orientation specificity index become identical within error range.

In the current implementation the stimuli are whitened and all principal components up to the given PCA dimension are amplified to amplitude one whereas the other amplitudes are set to zero. One reason for this preprocessing is the large decrease in computation time when using fewer dimensions. To assess the effect of the choice of the PCA dimension the position and orientation specificity is computed for different dimensions (Fig. 3C). None of these quantities changes significantly. Inspection of the resulting sub-unit receptive fields and $\theta$-$r$ diagrams reveals that still complex like receptive fields are obtained (data not shown). But since the dimension of the stimulus space is now much larger than the number of feature detectors, the coverage of the stimulus space is coarse and most complex cells have similar preferences.
Fig. 2. Results. A) The objective function is optimized till a stationary state is reached. B) Receptive fields of the sub-units after 175 iterations. C) $\theta - r$ diagrams for these sub-units. The diagram shows the response strength of the unit for bars of different position (x-axis) and different orientation (y-axis). D) $\theta - r$ diagrams for the complex cells.

4 Discussion

The presented results show that complex like receptive fields can be learned by extracting the slowly varying subspaces of natural stimuli. The obtained receptive fields are comparable to those of Hyvärinen and Hoyer (2000) who use a different approach, independent subspace analysis (ISA). ISA uses the same network layout but implements a different objective, independence of the cells’ responses, which is comparable to sparse coding. Whereas they use natural photographs taken from PhotoCDs we exploit the temporal domain of natural image sequences.

Another network for learning transformation invariant filters is the adaptive-subspace self-organizing map (ASSOM) proposed by Kohonen (1996). There also the neurons are modelled as sub-space energy detectors but the network learns a two dimensional map such that the activity maximum moves slowly over the network. The cells are implicitly forced to extract slowly varying features resulting in an approach comparable to the work of Foldiak and to the one presented here. Opposed to the ASSOM, the objective function approach incorporates the temporal smoothness in an explicit way and the results shown here were obtained from more natural stimuli.

The fact that quite different objectives lead to similar receptive fields poses the question to which degree the objectives of temporal smoothness and independence are equivalent.
Fig. 3. Controls. A) (Left) Receptive fields of the sub-units for a network trained with randomly paired stimuli (no temporal coherence). (Rightmost column) $\theta - r$ diagrams for the (no longer complex) cells. B) Increasing the time lag $\Delta N$ between two subsequent stimuli decreases the orientation specificity $\sigma_\theta$ (circles) and increases the position specificity $\sigma_r$ (diamonds). Errorbars denote the standard deviation over all cells in the network. C) $\sigma_\theta$ (circles) and $\sigma_r$ (diamonds) are shown as a function of the PCA Dimension.

It is interesting to note that the temporal smoothness function is very well compatible with a number of physiological mechanisms found in the mammalian cortex (Körding and König 2000). In this respect it is of importance that optimizing the objective function only needs information locally available to the cell.

A number of issues remain for further research: Different PCA dimensions require different subspace sizes and different numbers of neurons for optimal stimulus space coverage. Incorporating a dynamic subspace size in the objective function approach might recruit the optimal number of sub-units needed.

The presented results are obtained by using the 2-Norm of the subspace as transfer function for the cells. In this way the network becomes very similar to the classical energy models for complex cells which are supported by electrophysiological evidence. Some research on the other hand advocates stronger nonlinearities. Riesenhuber and Poggio (2000) for example propose the max function, which corresponds to the infinity norm. It seems likely that this network property can also be learned using the same objective function. Learning the norm of the sub-spaces might be worthwhile since it incorporates learning the nonlinearity of the network. Furthermore this could also lead to an explicitly learned contrast polarity invariance which so far is built in.
Concluding, temporal coherence is a method for learning complex type receptive fields from natural videos, and seems very well suited for learning different network properties of biological systems in which temporal information is ubiquitous.

Acknowledgments

This work was supported by the SNF (CK, PK) and the Boehringer Ingelheim Fonds (KPK). Furthermore we thank A. Hyvärinen and P. Hoyer for making their code available to the public. We are grateful to B. Betsch and C. Arielle for help with the acquisition of the stimulation videos.

References