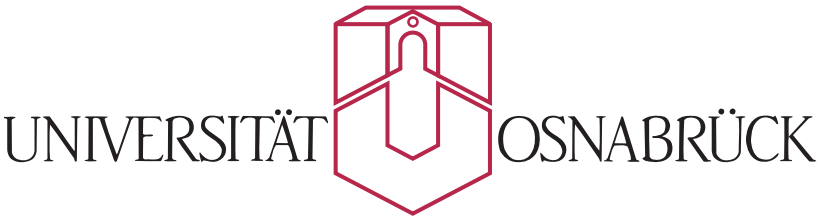


Context Effects in Early Visual Processing and Eye Movement Control

Dissertation
zur Erlangung des Grades
”Doktor der Naturwissenschaften“ (Dr. rer. nat.)
im Fachbereich Humanwissenschaften
der Universität Osnabrück

vorgelegt von

Nora Nortmann

Osnabrück, Februar 2014

Table of contents

Summary

There is a difference between the raw sensory input to the brain and our stable perception of entities in the environment. A first approach to investigate perception is to study relationships between properties of currently presented stimuli and biological correlates of perceptual processes. However, it is known that such processes are not only dependent on the current stimulus. Sampling of information and the concurrent neuronal processing of stimulus content rely on contextual relationships in the environment, and between the environment and the body. Perceptual processes dynamically adjust to relevant context, such as the current task of the organism and its immediate history. To understand perception, we have to study how processing of current stimulus content is influenced by such contextual factors. This thesis investigates the influence of such factors on visual processing. In particular, it investigates effects of temporal context in early visual processing and the effect of task context in eye movement control.

To investigate effects of contextual factors on early visual processing of current stimulus content, we study neuronal processing of visual information in the primary visual cortex. We use real-time optical imaging with voltage sensitive dyes to capture neuronal population activity in the millisecond range across several millimeters of cortical area. To characterize the cortical layout concerning the mapping of orientation, previous to further investigations, we use smoothly moving grating stimuli. Investigating responses to this stimulus type systematically, we find independent encoding of local contrast and orientation, and a direct mapping of current stimulus content onto cortical activity (Study 1). To investigate the influence of the previous stimulus as context on processing of current stimulus content, we use abrupt visual changes in sequences of modified natural images. In earlier studies, investigating relatively fast timescales, it was found that the primary visual cortex continuously represents current input (ongoing encoding), with little interference from past stimuli. We investigate whether this coding scheme generalizes to cases in which stimuli change more slowly, as frequently encountered in natural visual input. We use sequences of natural scene contours, comprised of vertically and horizontally filtered natural images, their superpositions, and a blank stimulus, presented with 10 or 33 Hz. We show that at the low temporal frequency, cortical activity patterns do not encode the present orientations but instead reflect their relative changes in time. For example, when a stimulus with horizontal orientation is followed by the superposition of both orientations, the pattern of cortical activity represents the newly added vertical orientations instead of the full sum of orientations. Correspondingly, contour removal from the superposition leads to the representation of orientations that have disappeared rather than those that remain. This is in sharp contrast to more rapid sequences for which we find an ongoing representation of present input, consistent with earlier studies. In summary, we find that for slow stimulus sequences, populations of neurons in the primary visual cortex are no longer tuned to orientations within individual stimuli but instead represent the difference between consecutive stimuli. Our results emphasize the influence of the temporal context on early visual processing and consequentially on information transmission to higher cortical areas (Study 2).

To study effects of contextual factors on the sampling of visual information, we focus on human eye movement control. The eyes are actively moved to sample visual information from the environment. Some traditional approaches predict eye movements solely on simple stimulus properties, such as local contrasts (stimulus-driven factors). Recent arguments, however, emphasize the influence of tasks (task context) and bodily factors (spatial bias). To investigate how contextual factors affect eye movement control, we quantify the relative influences of the task context, spatial biases and stimulus-driven factors. Participants view

and classify natural scenery and faces while their eye movements are recorded. The stimuli are composed of small image patches. For each of these patches we derive a measure that quantifies stimulus-driven factors, based on the image content of a patch, and spatial viewing biases, based on the location of the patch. Utilizing the participants' classification responses, we additionally derive a measure, which reflects the information content of a patch in the context of a given task. We show that the effect of spatial biases is highest, that task context is a close runner-up, and that stimulus-driven factors have, on average, a smaller influence. Remarkably, all three factors make independent and significant contributions to the selection of viewed locations. Hence, in addition to stimulus-driven factors and spatial biases, the task context contributes to visual sampling behavior and has to be considered in a model of human eye movements.

Visual processing of current stimulus content, in particular visual sampling behavior and early processing, is inherently dependent on context. We show that already in the first cortical stage, temporal context strongly affects the processing of new visual information and that visual sampling by eye movements is significantly influenced by the task context, independently of spatial factors and stimulus-driven factors. The empirical results presented provide foundations for an improved theoretical understanding of the role of context in perceptual processes.

1. General introduction

1.1 Views of perception

We can see leaves falling from a tree, hear the sound of a guitar, feel the softness of a cashmere jacket, smell roses, and taste the sweetness of a chocolate tart. We experience such rich percepts, but our senses, at the border between the environment and the skin and skull, are not receptive to objects in the environment. Instead, our sensors react to much simpler physical stimuli. Hair cells in the ear, for example, react to oscillations of air, and the receptors in the retina react to photons. This sensory input is influenced by objects in the environment in a complex way. For example, taking a closer look at the eye, we have to notice that the retina has different properties at its different locations. The spatial resolution is highest in the center and decreases towards the periphery (Sere et al., 2000). An objects' retinal activation pattern is dominated by the spatial resolution across the retina, changes in the wavelength sensitivity across the retina, and distortions in the retina itself. Making things more complicated, the position of the eyes in the world is not fixed. Body movements and eye movements lead to continuous image motion on the retina. These movements introduce changes in retinal activity even when the environment does not change. Thus, there is a large difference between the activity patterns at the receptor level and our stable perception of entities in the environment.

Given this divergence between the sensory input to the brain and the actual experience can lead to the question: How does the brain construct a faithful representation of the environment from the physical input obtained by our senses? Behind this question, there is often a view, in which the primary task of perception is essentially to compute a faithful representation of the environment from the information received (see Marr, 1982). Here, perception is a process that acts previous to cognitive capabilities, as it provides the input interface to the world. This process receives sensory information as input and provides a representation of the current outside environment as a result. Cognitive processes, here, work on the results of perception, which are internal representations. In the cognitivist view, there is a clear division of labor. Perceptual processes generate internal representations from sensory information. Cognitive processes, such as reasoning and planning, on the other hand, are processes that are completely internal. They, in turn, provide commands for motor control ("Sense-Think-Act Cycle" e.g. Malcolm et al., 1989; see Clark, 1998a, 1998b for subsumption). Consequently, based on the idea of such a division of labor, perception is viewed as rather independent of higher cognitive function and action. In the study of perception, accordingly, this view leads to an approach that tries to understand biological correlates of perception primarily based on the properties of currently presented stimuli.

Figure 1.1 Ambiguous image. Either a vase or two faces can be perceived. The image is an adaptation of the famous Rubin vase (Rubin, 1915).

However, perceptual processes are not only dependent on the current stimulus. This is especially apparent in cases, in which the same stimulus can lead to different percepts. This can be demonstrated with ambiguous images. Figure 1.1 shows an ambiguous image; either a vase or two faces can be perceived. Whether the observer perceives a vase or faces at a given time does not depend on the stimulus, as it is the same in both cases. Other factors determine how the brain processes the information. A possible factor can be information that has been processed prior to viewing an image. For example, an image that is similar to the ambiguous image, but biased to one of the two interpretations, influences observers to perceive later the respective object also in the ambiguous image (Leeper, 1935; Goolkasian, 1987; Goolkasian & Woodberry, 2010). Yet another factor is the task context of the observer. The vase or the faces can be easily seen, in the above example, given the task to look for either. Thus, perceptual processes are not solely determined by the stimulus, but influenced by contextual factors. Instead of constructing a representation of objective properties of external stimuli, perceptual processes dynamically favor an interpretation of the stimulus in dependence of the context at a given time.

Is there a general principle behind the influence of contextual factors on perceptual processing? On a systemic level, we can take a perspective in which brains "are not primarily designed to give faithful representations of the environment, but to ensure survival of their bearers and their genes" (König & Luksch, 1998). In such an action-oriented view (Engel et al, 2013), the most important goal of an animal is to produce beneficial behavior, which entails making fast decisions about action and solving specific imminent tasks. Perception, here, is thought to serve action. Perception, cognition and action are not seen as modular sequential processes that strictly divide labor, but instead they are seen as coupled to control a physical body that is situated in a real-world scenario (Clark, 1998b).[1] In the real world, sensory input is not arbitrary. Instead, there are dependencies in the world, including those between the world and the body. Such dependencies determine which entities provide information that is potentially useful in the processing of current stimulus content. These entities, accordingly, provide relevant context. An adaptation of the nervous system to such dependencies and a resulting use of the relevant context at every level, can allow the organism to facilitate processing of relevant information. Thus, direct sensory input alone is not sufficient for perceptual processes to provide an optimal base for the system to carry out its tasks. Additionally, these processes need to dynamically incorporate contextual factors in the processing of current stimulus content.

Concerning perception, action-oriented views of cognition do not focus primarily on the feedforward pathway from the sensors to higher areas. They oppose the classical emphasis on the questions of how information is propagated through the brain in a mainly feedforward way and how it then affects action in the end. Instead, action-oriented views emphasize the inter-connectivity of the brain (Clark, 1998b, Engel et al, 2013). Instead of sequentially processing of information, neuronal processes receive and generate feedback and can be influenced by internal dynamics, allowing them to adjust to relevant context. This also includes perceptual processes in early visual areas. There, processing of sensory information can depend on ongoing neuronal activity. Already in the primary visual cortex (V1) - the first cortical area receiving information from the eyes - feedback signals, interactions between neurons, and processes within individual neurons allow to constantly receive, process, and maintain information other than direct sensory input (see dedicated section 1.3 for details on this issue). Thus, in V1 internal dynamics and feedback signals provide capabilities for factors

[1] Maturana & Varela (1984) emphasize that there is a recurrent coupling between the world, the sensors, and the actuators for the purpose of producing useful to the survival of the organism.

other than those dependent on the current stimulus to take effect. Early visual processes can thus be dynamical, in the sense, that they do not process current stimuli in always the same way, but that they instead process new stimuli in dependence of varying contextual factors, including previous sensory input.

Sensory input changes over time due to self-motion and due to motion of objects in the environment. In such natural input, there are dependencies over time and space (see Simoncelli & Olshausen, 2001; Geisler, 2008 for reviews).[2] By natural selection and learning, the brain adapted to the properties of natural and ecologically relevant input (Barlow, 1961). This can allow the animal to more efficiently process information (Barlow, 1961), and enhance processing of potentially relevant information useful to satisfy its needs, and support fast choices of action.[3] Given dependencies across time, visual input continuously provides new relevant information, and thereby constantly provides context for the processing of future input. Because of an adaptation of the visual system to temporal dependencies in natural input, processing of new input might dynamically change dependent on the context of previous input.

Predictive coding theories (Rao & Ballard, 1999; Friston, 2005, 2010; Huang & Rao, 2011; Clark, 2013) even propose that the brain constantly makes predictions about future input. In these frameworks, internally generated predictions are constantly matched to new incoming information from the sensors. Only in case of a mismatch between the prediction and incoming information, deviations are propagated to higher areas as error signals, via the feedforward pathway.

Another factor that plays a role for the generation of beneficial behavior is related to the current needs of the animal. Which sensory data is especially informative depends on the context of the imminent tasks. The tasks can differ in the information that they require. For example, when looking for signs of human civilization in the wilderness, a manhole cover might be very informative. It is not so, when having to decide whether it is possible to keep walking straight ahead. The tree in front, blocking the view, however, is highly informative for choosing a walking path, as it indicates that it is better to turn, while it holds little information regarding signs of human civilization. Thus, how informative given stimulus content is, depends on the task. A bias towards processing of informative stimulus content necessitates a dynamical adjustment of perceptual processes to the task context.

Such a task-dependent adjustment can take effect via sampling of information from the environment.[4] Action-oriented views of cognition emphasize that vision is not a passive process, starting with the stimulus, which is passively fed to the nervous system, for it to be

[2] Not only are there, independent of the observer, statistical properties in the environment that can be used by the brain to more efficiently process information. Lawful changes in the input can also be introduced by action. When the animal moves its own sensors (e.g. by head and eye movement), the visual input changes in a lawful way. Such sensorimotor contingencies between visual input and action are learned by the brain (O'Regan & Noë, 2001) and can be used to predict future sensory input (see Engel et al., 29013 for a review). In addition to sensorimotor contingencies, which are direct relationships between sensor movement and changes in sensory input, lawful changes in the environment can also be introduced by action, when the animal changes objects and structures in it. In this context the notion of "extended mind" (Clark & Chalmers, 1989) is relevant. The proposition is that the human organism uses the environment as an extension of the mind, by structuring it in a way that helps to solve cognitive tasks (see Clark & Chalmers, 1989).

[3] The original idea is from Barlow (1953), when he discovered that response properties of frog retinal neurons match up to the frogs feeding behavior (Barlow, 1953). See (Martin, 1994) for review.

[4] The sampling of information can also influence what is later perceived. In the case of ambiguous images (example in Figure 1.1), Kietzmann et al. (2011) found that which of the two possible objects in the image is perceived by an observer does depend on which locations she has previously inspected (prior to object awareness).

left with the task to get the most usage out of it, but that instead, the animal itself actively selects sensory input by its own actions (Ballard, 1991; Ballard et al., 1997). As mentioned, the image on the retina is in constant motion, as the animal moves its body and eyes. The movements of the sensors in the environment determine which signals are received by them and can be processed further in the brain. The process of sampling sensory information from the environment is partly stimulus driven (e.g. Koch & Ullman, 1985; Itti & Koch, 2001), but also goal directed and depends on the current task of the observer (Yarbus, 1967; Land et al., 1999; see dedication section 1.4 about this issue).

In conclusion, when assuming that the primary function of perception is to aid the generation of beneficial behavior, we do not expect perception to create a strictly objective representation of the environment in the animal's brain. Instead, perceptual processes, including neuronal processing and sampling of information, are viewed as dynamical, in the sense that they can incorporate relevant context and adjust the processing of current stimulus content to this context. Such context includes the current task of the animal and previous sensory input. The effect that one stimulus can lead to different percepts in the case of ambiguous images, is a byproduct, or consequence of a dynamical tuning of visual processes to relevant context. Context dependent perceptual processing of current stimulus content is not the exception, but an integral part of perception.

This thesis contributes to the investigation of the influence of contextual factors on visual processing of current stimulus content. To better understand how temporal context effects early visual processing of current stimulus content we study neuronal processing in the primary visual cortex. To characterize the cortical layout in V1 concerning the mapping of orientation, as a foundation for further investigation, we use moving grating stimuli. Investigating responses to this stimulus type systematically, we find simultaneous independent encoding of local contrast, in addition to orientation, and a direct mapping of current stimulus content onto cortical activity (Study 1, Section 2). In the main study, we investigate the influence of the context of previous stimulation on processing of current stimulus content using abrupt visual changes in sequences of modified natural images. We find a strong influence of the previous stimulus on the processing of new sensory input (Study 2, Section 3). To better understand how contextual factors interact with stimulus-driven factors in visual sampling behavior, we conduct a study to quantify the relative influences of the task context of the observer, spatial biases and stimulus-driven factors on eye movement control (Study 3, Section 4).

In the past, it has proven successful to understand a number of aspects of visual processing based on simple image properties. However, there are additional aspects that cannot be captured or predicted by just considering the current stimulus. Section 1.2 and 1.3 elaborate this issue, the earlier section focusing on neuronal processing in V1, and the latter focusing on the sampling of visual information by eye movements. The last section of this introduction (section 1.4) provides specific background and motivation of the presented empirical studies. These studies are covered subsequently in section 2 to 4. Section 5 gives an additional discussion. The appendix 6.1 provides supplementary information on the method of voltage sensitive dye imaging that is used in two of the presented studies. Appendix 6.2 documents an additional study, which investigates the mapping of visual input from the retina to V1 neurons (retinotopy) in rat.

1.2 Processing in the primary visual cortex (V1)

The visual system can be described as hierarchical network (Felleman & van Essen, 1991). Feedforward signals travel up the hierarchy, from the sensors, through early visual areas to higher areas. Feedback signals travel down the hierarchy. The retina, in the back of the eye, is the first stage of the visual processing hierarchy. Here, photoreceptors convert incoming light into chemical signals. Neurons in the retina, in particular interneurons and ganglion cells, translate these into electrical signals. Axons of ganglion cells form the optic nerve, which leaves the retina at the optic disc, and projects to the lateral geniculate nucleus (LGN) in the thalamus (see e.g. Kandel et al., 2000). The LGN projects to the primary visual cortex (V1).[5] V1 is thus the first cortical area that receives input from the eyes. V1 is a bottleneck for of visual information, as most visual information available to higher cortical areas passes through it.

1.2.1 Understanding V1 function in terms of simple stimulus properties

In the investigation of V1, we can build on a base of research that was done in this area over the last 50 years. V1 is now one of the best-investigated areas in the mammalian brain. The foundations go back to Hubel and Wiesel, who had great success in studying the neuronal properties in the cat and macaque visual system with relatively simple artificial stimuli such as oriented bars. An important idea in the study of properties of V1 neurons is the concept of a receptive field. A receptive field is a projection of a part of the retina onto the visual field, particularly, of that part of the retina that connects to a single cell, in a feedforward way. Hubel & Wiesel (1959) found that V1 neurons have local receptive fields that are selective for stimulus orientation. Later they found that neurons are grouped systematically in columns across the cortical surface in a topographical fashion, so that nearby cells prefer similar orientations and locations, thereby representing visual space and contours in overlaid retinotopic and orientation maps (Hubel & Wiesel, 1974; also see appendix for a retinotopic map in rat V1).[6] The findings provide evidence for the concept of a systematic mapping between visual space and cortical space.[7]

Over the years, a standard model of neurons in V1 was established. This model is widely agreed upon by researchers as a useful basis for further study and elaboration. Briefly, this model has three stages: In the first stage of processing, the input image is passed through one or more linear spatiotemporal filters (usually Gabor-like functions) - one for simple cells, two or more for complex cells (Hubel & Wiesel, 1962). Such filters compute one value, which, in intuitive terms, reflects how similar a particular part of the input image is to the filter. In the second step, the computed value is passed through a threshold function. In the case of two or more filters, as in complex cells, the values are combined (essentially summed) and then

[5] Cat V1 is widely used as a model system to study visual processing and also cerebral circuitry and processing in general (Payne & Peters, 2002). Throughout this thesis the definition of cat V1 as area 17 and 18 is used. Both areas receive their driving input from the LGN (Lee et al., 1998; Malpeli, 1983; Malpeli et al., 1986) and are structurally and functionally similar. For information on the historic development of the concept of cat V1 please see Payne & Peters (2002; pp. 1-129).

[6] With their work on the properties of neurons in the visual cortices and their functional layout (see Hubel & Wiesel, 1979 for overview) and development (e.g. Wiesel & Hubel, 1965; Hubel & Wiesel, 1963b) Hubel and Wiesel have laid the foundations for what is today a research field on its own: Visual Neuroscience. In 1981, they received a Nobel Prize for their discoveries concerning information processing in the visual system. The Nobel Prize in Physiology or Medicine was shared with Roger Wolcott Sperry, who did receive it for his work on the lateralization of brain function in split-brain patients (Sperry, 1967; Sperry, 1974).

[7] Hubel & Wiesel were, of course, aware that perceptual processes, including those reflected in their measured responses, could be influenced by factors not dependent on the current stimuli. This is apparent e.g., when Hubel & Wiesel (1959) discuss effects of anesthesia. Also Hubel (1959) carried out experiments in freely moving cats.

passed through the threshold function.[8] As a third step, the result is used as the rate parameter in a Poisson spiking process. The higher the rate is, the higher is the expectation of the spike count.[9] This model captures basic properties of cells in V1. In particular, it captures basic characteristics of simple and complex cells: e.g. their receptive field properties such as orientation, location and frequency tuning (Hubel & Wiesel, 1962; Foster et al., 1985).

The standard model of V1 neurons, essentially working like a Gabor filter bank, is a feedforward model with relatively simple functional units. Response properties of neurons are fixed and completely defined by the receptive field properties of the neuron. The model predicts that, whenever there is a change in the stimulus, this change is essentially mirrored in the cortical response. A Gabor filter bank faithfully maps present input, in dimensions of spatial frequency, orientation and location, onto cortical activity, leading to ongoing and feedforward encoding of present input, with minimal interactions between responses to successive stimuli (e.g. see Benucci et al. 2009).

The standard model, as described above or in a similar version, has been successful in predicting neuronal responses to a degree. It is widely accepted as a working model of V1 function (see e.g. Lennie, 2003 for overview) and is used to predict psychophysical performance (Graham & Nachmias 1971; Watson et al. 1983; Anderson et al. 1991; Bethge et al., 2007). It has also been shown to provide efficient representations of natural scenes (Olshausen & Field, 1996; Bell & Sejnowski, 1997; Hurri et al., 1997). In general, understanding V1 function in terms of a representation of the environment by predicting neuronal activity as the result from a filtering processes, in which neurons with Gabor-like stationary receptive fields filter given external stimuli, has been successful in the past.[10]

1.2.2 Understanding V1 function in terms of internal dynamics

Although the receptive field standard model is a useful conceptual tool to model neuronal response properties in V1, it fails to capture a number of established empirical results. There is evidence that early perceptual processes are influenced by internal neuronal dynamics and not by the current sensory stimulus alone in a purely feedforward way.[11] Arieli et al. (1996), for example, showed that the variability of activity in response to a specific stimulus could be predicted well from previous stimulus-independent activity. Fiser et al. (2004) showed that correlations found in spontaneous activity (neural behavior without stimulation) in V1 are only slightly modified by stimulation.

Internal neuronal dynamics can act on different levels. They concern processes within individual neurons, and communication between different neurons. Concerning

[8] The different filters that are combined in complex cells have similar locations and are tuned to a similar orientation, spatial frequency and temporal frequency. They are however different in phase, making them invariant to the exact position of the stripes of a grating or a bar of light (see Hubel & Wiesel (1962) for complex cell characterization).

[9] This description of the standard model is taken from Rust and Movshon (2005), who themselves propose extensions to it. But see also Carandini et al. (2005) and Olshausen & Field (2005) for similar descriptions of the standard model and reviews therein. Additionally, in many version of the standard model, there is a normalization step, in which responses are normalized by the responses of neighboring neurons (see Olshausen & Field, 2005).

[10] The assessment of the standard model in this paragraph is taken from Olshausen & Field (2005). See also König & Luksch (1998) for a similar assessment of its relevance.

[11] This section concentrates on temporal effects. However, there is extensive work on V1 neuronal response behavior focusing on the investigation of spatial context. Spatially, context presented outside of the receptive field influences a neurons' response behavior to stimuli presented inside its receptive field. Such effects are generally referred to as contextual modulation (of receptive field properties) and have led to the notion of the non-classical receptive field, spatially extending beyond classical receptive field borders, and including cortical interactions. See Lee (2009) for an overview. A comprehensive review of spatial contextual effect in visual processing in general is given by Albright and Stoner (2002).

communication between different neurons, we have to note that although V1 is the first cortical area that receives input from the eyes, most of its input comes from other neurons (Peters & Payne, 1993; Peters et al., 1994). Olshausen and Field (2005) estimate that between 60-80% of the response of a V1 neuron is a function of other neurons in the same area, or inputs other than those arising via the feedforward route from LGN. Thus, from an anatomical point of view, stimulus processing in V1 can be influenced by cortical mechanisms, via local intra-cortical circuits and feedback from higher areas.

Additionally, there is evidence for mechanisms within individual neurons that integrate input over time and lead to a changes in neuronal response properties depending on their history. For example, short-term synaptic plasticity describes synaptic changes found on time scales from a few milliseconds up to minutes. Although, these can also involve cortical circuitries (Zucker & Regehr, 2002), they rely on processes within individual neurons that go beyond what is covered in the standard model, mentioned above. For example, activity dependent synaptic depression due to depletion of vesicles of the presynaptic cell (see Zucker & Regehr, 2002 for review) can affect response properties of individual neurons over time. Thus, internal dynamics mediated by processes within individual neurons can influence how new sensory information is processed.

In respect to temporal context, internal neuronal dynamics can influence responses to presented stimuli, in dependence of previous input. Contrast adaptation, for example, which leads to stimulus specific changes in perceived contrast and orientation specific aftereffects (Blakemore & Campbell, 1969; Blakemore & Nachmias, 1971, see Kohn, 2007 for review), was shown to involve effects in V1 (in addition to the retina, e.g. Smirnakis et al., 1997). In V1, it reduces neuronal responsiveness and causes preferred orientations to shift away from the adapting orientation on short time scales (Movshon & Lennie, 1979; Müller et al., 1999; Felsen et al., 2002). Adaptation to a single orientation was also shown to enhance the representation of orthogonal orientations (Dragoi et al., 2002). Cortical mechanisms (Movshon & Lennie, 1979; Felsen et al., 2002) as well as cellular mechanisms (Sanchez-Vives et al., 2000), including synaptic depression (Müller et al., 1999), mentioned earlier, have been related to contrast adaptation. On the functional level, contrast adaptation has been suggested to increase efficiency in the visual system. According to this thought, its role is to adjust the gain of a neuron to avoid saturation and make better use of its limited dynamic range (Baccus & Meister, 2004).

Another neuronal property that depends on previous stimulation is response behavior immediately after a stimulus is turned off. Neuronal off responses are responses to a preferred stimulus, after it has disappeared (Coltheart, 1980; Duysens et al., 1985; McCormick et al., 2003). As these responses occur, after the stimulus is no longer visible, they do not correspond to current stimulus content. Such off-responses can depend on how long the previous stimulus was shown (Duysens et al., 1996) and are thus dependent on the history of stimulation. These effects, contrast adaptation and neuronal off responses, show that previous input can influence responses to current input already at early stages of visual processing. Thus, they are candidate mechanisms to enable processing of new information in context of the immediate past.

As described above, the standard simple receptive field model, that does not cover such time-dependent changes in neuronal response properties, is quite successful in predicting neuronal responses. Also on a population level, where responses of many neurons are captured at once, the average cortical response to an individual stimulus can be used to predict the response to that stimulus in a context or sequence of other stimuli well (Benucci et al., 2009). The

functional roles of feedback signals, neuronal interaction within the same area, and processes within individual neurons (that go beyond the mechanic of the standard model) and their influence in the processing of new sensory input are subject of ongoing research. It is possible that the influence of previous input on current processing is dependent on parameters not yet sufficiently explored. Investigating such influences in early visual processing is a necessary step towards an improved understanding of the role of context in perceptual processing in general.

To investigate cortical dynamics empirically, choices concerning the stimuli and the recording method have to be made. The remainder of this section will address these topics in turn.

The choice of stimuli

When planning to measure visual responses, appropriate stimuli have to be chosen. How does the choice of stimuli influence what information about neural function is revealed? Internal dynamics, which integrate information over time and across different neurons, might be expert to facilitate perceptual processing concerning the goals of the animal in its every day environment. This is because the organism is not a general processor. Instead, its body, its behavioral repertoire, its sensory apparatus, and its nervous system are all adapted by natural selection and learning to the kinds of stimuli, it has encountered during its evolutionary history and during its lifetime. Cortical processes are adapted to process natural input efficiently (Barlow, 1961). In case of visual processing, this results in an adaptation of the visual system to the statistical properties of natural images (Simoncelli & Olshausen, 2001; Felsen & Dan, 2005).[12] According to this line of thought, the visual system is specialized to process a certain subclass of stimuli: those whose statistics are similar to that of natural images. Stimuli that fulfill this criterion in some respect are said to be more ecologically plausible than stimuli that do not fulfill it.

Concerning processing in the visual cortex, evidence for such an adaptation of the visual system to statistics of natural stimuli has been found in previous studies. For example, Felsen et al. (2005) found an elevated sensitivity of V1 neurons to visual features when they were presented in natural scene context, as compared to random context. Mante at al. (2005) provide evidence that luminance and contrast are independent in natural scenes. They found that this independence corresponds to independent gain control mechanisms for luminance and contrast in early visual processing. Thus, they show how properties of mechanisms in visual processing can reflect properties of natural images.

Because of such adaptation, taking the statistical properties of natural stimuli into account can help, and might even be necessary, to understand neuronal function. Artificial stimuli, such as gratings and bars, are effective tools in the study of the visual system. They have the advantage that they are easy to control and that their properties are well understood (Rust & Movshon, 2005). However, they provide an unusual input to the visual system. Findings obtained with such stimuli might not always generalize to scenarios in which complex stimuli are used (see Felsen & Dan, 2005). It is possible, that internal dynamics involving mechanisms specialized to processes ecologically relevant input, cannot be probed with artificial stimuli. The more relevant the stimuli, the more influence might internal dynamics have on the processing of current stimulus content. Thus, to evaluate the generality of results

[12] See Hyvärinen et al. (2009) for a thorough discussion of natural image statistics.

obtained with artificial stimuli, it is important to validate such findings using stimuli that are more natural.

The choice of the recording method

When investigating neuronal processes, it is important to attend the question of what signals can be measured. Unfortunately, the data that can be obtained is limited. It is not possible to record all activity of all neurons at once with unlimited spatial and temporal resolution. Any recording of brain activity only provides a small window into the brain.

Non-invasive methods such as EEG, fMRI, or MEG can provide correlates of activity across different regions of the brain, measured simultaneously; however, spatially the measurements are coarse. These measurements, for example, do usually not resolve individual orientation columns in V1. A recent breakthrough was made with fMRI, where individual columns could be measured in humans with a seven Tesla scanner (Yacoub et al., 2008). Although these methods have made advancements, as for example, fMRI has been used to disentangle activity from different cortical layers (e.g. Goense & Logothetis, 2006; Ress et al., 2007; Olman et al., 2012), these non-invasive methods are limited in their use for the investigation of functional properties on the neuronal level.

Invasive techniques allow accessing the activity of individual neurons. Intra-cellular recordings, for example, report the currents within individual neurons. In extra-cellular recordings, the electrode does not penetrate the membrane of a cell, but instead is brought in close, to record voltage changes from the outside of either a single neuron or, when recording multi-unit activity, a few. The prevailing method to obtain neuronal responses is extra-cellular recording, which is the base for the majority of our knowledge about neuronal function in V1 (see Olshausen & Field, 2005). However, as both methods provide signals from individual cortical locations at a time, they are limited in their use for the investigation of neuronal interactions across different cortical locations.

Methods that overcome this limitation by simultaneously recording across cell populations are multi-electrode recordings and optical imaging techniques. In contrast to intrinsic optical imaging, optical imaging with voltage sensitive dye also provides a high temporal resolution in the millisecond range. Whereas the multi-electrode recording *samples* at individual locations, the voltage sensitive dye signal reflects spatial *averages* of neuronal activity at individual cortical locations.[13] Such simultaneous recordings at several cortical locations allow us to investigate dynamical aspects of cortical processing in large cell populations, which cover a variety of neurons tuned to different stimulus features.

In summary, the standard model, which predicts responses of V1 neurons based on fixed receptive field properties, has been successful. However, there is evidence that V1 neuronal response properties are dependent on temporal context. In the investigation of the dependency of early visual processing on temporal context and the related cortical dynamics, the use of ecologically relevant stimuli and the recording of many neurons simultaneously might be useful.

[13] Details of the method of voltage sensitive dye imaging and a comparison to other methods are given in a dedicated section in the Appendix (section 6.1).

1.3 Eye movement control

Early processing of visual input can be influenced by contextual factors. In addition, contextual factors can also affect visual processes by influencing the actions that are used to acquire the visual input. In general, visual perception is an active process. The active aspect of perception is obvious in the classical examples of active sensing (see Schroeder et al., 2010), such as whisking in rodents (Ritt et al., 2008), or echolocation in bats, dolphins, and fish (von der Emde & Schwartz, 2003; Thomas et al., 2004). The weakly electric fish Gnathonemus petersii does not passively receive input from the environment. It actively produces electrical discharges using an electric organ at its tail. Nearby objects distort the electrical field around the fish. These distortions are detected by the fish's electroreceptive skin surface. In that way, the fish actively samples the information needed for its imminent tasks, such as object detection and self-localization (von der Emde & Schwartz, 2003). In the case of human vision, this is not much different. Although the organism does not emit light, it performs selective actions: It actively changes its own position and the orientation of its sensors (e.g. the eyes) thus guiding its own information intake. This sampling process is an integral part of visual perception.

What light will hit the retina depends on the position of the eyes in the world. Not all active-sensing movements in vision are solely performed by the eye muscles. Movements of the eyes, the head and the entire body might synergistically be employed to select a new gaze location (see e.g. Einhäuser et al., 2009a). On the other hand, not every eye movement is performed with the intent to sample new information. Eye movements can be intended to compensate for a head movement and in effect stabilize the retinal input (Einhäuser et al., 2009b).[14] Here, we are not interested in these compensatory eye movements but in those that are part of the active selection of visual information. In the literature, the term 'overt visual attention' has become synonymous with these kinds of eye movements.

In humans, such eye movements consist of fixation periods and saccade movements (Buswell, 1935). During fixation, the eyes are rather stationary and center one location in the visual field. The centered part of the visual field is represented with the highest spatial acuity and it receives disproportionately more cortical processing resources than the surrounding areas (Tootell et al., 1982). Saccades are movements from one fixation location to the next. Human fixations in scene perception have an average duration of 330 ms (Rayner, 1998). This is the average time the human brain spends to process visual information with the eyes centered at one location, before selecting the next fixation location.

1.3.1 Stimulus-driven eye movement control

A first question is how the selection visual input is driven by the stimulus. Some locations in a scene are more salient than others and thereby attract attention. To predict stimulus-driven eye movement control in the viewing of images, the probability for fixation is based only on features that are extracted from the image. Luminance contrast, for example, is a local image property that can be used to predict how often an image location is fixated (Itti & Koch, 2001). Many different low-level image features can be used to compute fixation probabilities. These probabilities are then combined in a so-called saliency map. Such saliency maps are the basis for stimulus-driven models of human eye movements (e.g. Koch & Ullman, 1985). In principle, the selection of features for such models is unconstrained and present models

[14] In addition, there are also other types of eye movements, such as tremor, drifts and microsaccades, which are not in the focus of the present work (for review see Martinez-Conde et al., 2004).

include a considerable number of features (Wolfe & Horowitz, 2004; Betz et al., 2010). Using saliency maps, it is possible predict, to some extent, fixations in complex scenes based on simple stimulus properties alone (Privitera & Stark, 2000; Li, 2002; Parkhurst et al., 2002; Peters et al., 2005; Tatler et al., 2005).

1.3.2 Task context and spatial factors in eye movement control

There are additional factors that play a role in the selection of fixation locations that are not dependent on the specific scene or image that is viewed. Such are spatial constraints and properties of the oculomotor system itself, which lead to spatial biases on eye movements. For example, there is a tendency to fixate close to the center of an image (central bias, see Tatler et al., 2006 and Tatler, 2007). Other spatial biases concern the geometric properties of saccades (Brockmann & Geisel, 1999). One of these geometrical properties is a preference to select the successive fixation location rather close to the previous one, with the majority of saccades crossing a distance of 15 degrees of visual angle or less (Bahill et al., 1975). While it is clear that these spatial biases contribute to the selection of fixation points, their general influence yet needs to be quantified.

Simple stimulus features and spatial biases are not the only factors that influence the selection of fixation points. It has long been shown that eye movements depend on the task context of the observer (Buswell, 1935; Yarbus, 1967). The same picture is inspected differently, depending on whether observers, for example, are asked to estimate the age of people in it, or judge the material circumstances of a family in the picture (Yarbus, 1967). When perceiving a scene, the selection of sensory input differs with the given task, possibly related to how informative specific image regions are, in respect to that task. However, the objective quantification of the task-related information content, in specific image regions, remains a challenge. Information content is often determined intuitively or based on direct subjective ratings. Furthermore, it is unclear to what degree task-dependent local information content causally contributes to the selection of fixation points in relation to low-level image features.

Thus, low-level image features, spatial biases, and the task context all have an influence on eye movement control. An interesting topic is the interaction between these three factors. While each of these three factors has individually been shown to affect gaze, it is not clear to what degree and how these factors compete in eye movement control. In principle, it is possible that they are redundant to a large degree and predict similar fixation locations. A quantification of the relative influences of these factors is needed.

1.4 Studies, background and motivation

Perceptual processes are influenced by the current stimulus, but also depend on contextual factors. In early visual processing, simple receptive field models predict V1 responses to a degree. In the sampling of visual information, low-level image features can be used to predict eye movements. In both cases, however there is evidence that other factors play a role. In this thesis, we investigate specific influences of contextual factors and current stimulus content on visual processing. In V1, we first investigate the processing of current stimulus content in smooth visual motion and, second, the influence of previous input on the processing of current input in image sequences. Concerning the sampling of visual information by eye movements, we investigate the contributions of the task context of the observer in relation to stimulus-driven factors and spatial biases. This section elaborates detailed motivations for the presented studies.

1.4.1 Representation of current stimulus features in V1

Investigations of neuronal processing of current stimulus content are an important foundation for the investigation of contextual factors. In general, to characterize the cortical layout in V1 concerning the mapping of orientation, previous to further investigation, we use smoothly moving grating stimuli. In the first study, we investigate responses to this stimulus type systematically. In particular, we investigate to what extend information about current properties of this stimulus type is reflected in V1 population responses.

The receptive field of an individual V1 neuron, as mentioned above, has a location and an orientation preference. The finding that V1 neurons are topographically organized, thus mapping retinotopic location and orientation in overlaid retinotopic (coding stimulus local contrast) and orientation maps (coding stimulus orientation), is well accepted and has been reproduced across different species.[15] However, how these different information channels are simultaneously encoded across large neuronal populations remained unclear. Usually, visual stimuli provide both, orientation information as well as local contrast, i.e. location information. In a moving grating, both of these dimensions can easily be controlled experimentally. This stimulus type provides orientation information given by the orientation of the stripes and location information given by the exact positions of the moving stripes (the phase of the grating).

In particular, when a moving grating is presented, the orientation of the stripes leads to large differences in activity levels across the neuronal population, because of the orientation selectivity of individual cells. Neurons that are tuned to the orientation of the stripes respond strongly, while neurons that are tuned to the orthogonal orientation respond the least. In species in which orientation selective cells are topographically organized, the differences across the neuronal population can become visible as spatially structured patterns of activation on the cortex (e.g. Blasdel & Salama, 1986; Bonhoeffer & Grinvald, 1991; Bosking et al., 1997; Sharon & Grinvald, 2002). In the optical signal, the response to a preferred orientation is 20% stronger than the response the orthogonal orientation (Onat et al, 2011). Because of their high amplitude and their spatial structure, the response differences across V1 can be detected with various techniques. Even in fMRI experiments in human, these orientation patterns can be detected (Yacoub et al., 2008; Freeman et al., 2011).

This well-known orientation response, however, provides an ambiguous representation of the stimulus, as it does not contain information about the motion and location of the individual grating's stripes (the phase of the grating). It remained unclear how this second stimulus dimension, local contrast, is simultaneously represented to orientation across the two cortical maps. Given that the retinotopic representation of local contrast is independent of the cortical orientation map (Bosking et al., 2002; Buzas et al., 2003; Yu et al., 2005, but see Das & Gilbert, 1997), we hypothesize that smoothly travelling waves of cortical activity (Engel et al., 1994, 1997; Lee et al., 2005), which represent the current positions of the grating's stripes, should be detectable at the same time.

[15] The topographical organization of orientation selective cells is species-specific. Such a topographical arrangement of orientation preference is found in cat (Hubel & Wiesel, 1963a; Grinvald et al., 1986), primate (Hubel & Wiesel, 1974; Hubel et al., 1978), ferret (Redies et al, 1990), tree shrew (Humphrey et al., 1980a, 1980b), sheep (Clarke et al., 1976), and human (Yacoub et al., 2008). The situation is different in rodents, such as rats. Rats have orientation-tuned cells, but no orientation topography (Ohki et al., 2005).

Many recording methods are not suited to record these waves of activity across cortical orientation maps. Single electrode recordings do not provide a coherent population picture of activation across the cortex due to restrictions in spatial sampling. Intrinsic optical imaging and fMRI are limited in temporal resolution because of their dependence on slow hemodynamic processes. In Study 1, we overcame these limitations by using voltage-sensitive dye imaging, which allowed us to record population activity in the millisecond range across several millimeters of cat V1.

To investigate the retinotopic representation of visual motion, simultaneously to the orientation representation, we show smoothly moving gratings of different orientations, temporal, and spatial frequencies. Using singular value decomposition, we are able to separate the responses to the grating's global orientation and its individual moving stripes. The orientation response matches the orientation of the grating, and the retinotopic response, which is two orders of magnitude smaller, precisely matches the actual movement of the gratings' stripes in space and time. Capturing and resolving both responses simultaneously, we are able to show an independent processing of the two stimulus features across the neuronal population.

1.4.2 Influences of the past stimulus context on responses in V1

In the second study, we focus on the influence of the context of past stimulation on the processing of current stimuli. As the response to the orientation of a stimulus is a dominant response in V1, as elaborated above (see Study 1; the orientation response is much higher than e.g. the retinotopic response), we concentrated on the processing of changes in orientation in the second study. Previous studies, that investigated responses to rapid changes of orientation, found a representation of current image content. For example, Ringach et al. (1997) and Benucci et al. (2009) investigated V1 responses to stimulus sequences composed of gratings that differed in orientation content. [16] Ringach et al. (1997) made electrophysiological recordings and Benucci et al. (2009) additionally used voltage-sensitive dye imaging. Data from these studies provided evidence that V1 continuously represents present input, with little interference from previous stimulation. This is referred to as 'ongoing encoding scheme' in the following.

In the investigation of the influence of previous stimuli onto processing of current stimuli, the considered timescales might play a role. In the referred studies the investigated time scale is relatively short: Ringach et al. (1997) used a presentation duration of 17 ms for each single grating; Benucci et al. (2009) used a presentation duration of 32 ms. On what timescales do changes occur in the natural input to the cat visual system? Kayser et al. (2004) investigated temporal correlations of orientations in the natural movies (Betsch et al., 2004). They found long lasting temporal correlations between collinear orientations, which are significant over several hundreds of milliseconds and strongest for the cardinal orientations.[17] These findings indicate that under natural conditions, collinear orientations can be present for hundreds of milliseconds, before a change of orientation occurs. Mechanisms such as adaptation (Movshon & Lennie, 1979; Müller et al., 1999; Felsen et al., 2002; Dragoi et al., 2002) and off-responses (Duysens et al., 1996, Bair et al., 2002) are known to be time-dependent. They might have a stronger effect with time, interfere with an independent processing of individual stimuli in a sequence, and lead to an increased influence of the preceding stimulus on the

[16] This is not a complete list. Other related work is referenced in the study itself (section 3).
[17] Kayser et al. (2004) report decay time constants for 0°: >1 s, 45°: 490 ms, 90°: 900 ms, 135°: 360 ms.

processing of the next stimulus. Thus, responses to changes in orientation content might strongly depend on the timescale under consideration.

Considering the potential time dependency of the effects of previous stimuli onto current processing, an empirical study to investigate whether the ongoing encoding scheme transfers to longer timescales is needed. Therefore, in Study 2, we investigate whether the ongoing encoding scheme, which is found for sequences of oriented stimuli shown with short presentation durations (<32 ms) in V1, does generalize to sequences presented with longer durations. A significant deviation from this encoding scheme would indicate an influence of the previous stimulus on the processing of the next.

Concerning the choice of stimuli we take into account that, as mentioned earlier, response characteristics obtained with simple artificial stimuli, like the gratings, do not necessarily generalize towards input of ecological relevance (Smyth et al., 2003; David et al., 2004; David & Gallant, 2005; Felsen et al. 2005; Haider et al., 2010; Fournier et al., 2011; Onat et al. 2011). That is, response behavior to natural stimuli can deviate significantly from predictions based on simple parameterized stimuli (see Carandini et al. (2005) and Olshausen & Field (2005) for reviews). Thus, it is important to validate findings obtained with artificial stimuli using more natural stimulus conditions (Felsen & Dan, 2005).

For this reason, we use types of stimuli with different degrees of complexity. We use simple artificial grating stimuli for comparison with the literature and additionally we use stimuli that are more natural. These were derived by extracting oriented contours from natural images by filtering them along the orientation dimension (horizontal or vertical) in Fourier space. We either apply a filter that is narrow in orientation space, resulting in vertical and horizontal gratings (of multiple superposed spatial frequencies), for which the superposition is similar to a plaid pattern, or we use a wide filter, that is more tolerant to content from neighboring orientations. Here, the superposition of the filtered horizontal and vertical image adds up to the unfiltered natural image. In contrast to the gratings, these images retain important properties of natural stimuli, such as their phase relationships and the typical 1/f fall-off of amplitudes along the spatial dimension (Simoncelli & Olshausen, 2001; Geisler, 2008; Hyvärinen et al., 2009).

For a comparison with the literature, we present the oriented stimuli and their compound superpositions within 33-Hz sequences (30 ms presentation duration for each image). For the investigation of longer stimulus durations, we present them within 10-Hz sequences (100 ms presentation duration for each individual image). To capture population activity evoked by these sequences of oriented stimuli across V1, we use voltage-sensitive dye imaging. Based on responses to pairs of stimuli within the sequence (pair-triggered averages), we investigate the influence of the preceding stimulus on the response to the next. Indeed, we find that the established view of cortical responses to sudden changes in orientation, which is that they represent the stimulus currently shown ('ongoing encoding' scheme), does not generalize to longer stimulus durations. Instead, the observed responses provide evidence that information from the previous and the current stimulus is combined and another coding scheme has to be considered to explain the observed data.

1.4.3 Influences of the task context on visual sensory sampling

In the third study, we focus on the in influences of context on the sampling of visual information by eye movements. To investigate such influences, we concentrate on a quantification of the influence of the task context on human eye movements in relation to

influences of low-level image features and spatial biases. As described above, it is widely acknowledged that low-level image features, the task of the observer, and spatial biases contribute to the selection of fixation points. However, no study has comprehensively investigated the relative influences of these three factors on eye movement control. An open topic is the potential explanatory overlap of these three factors. A point of concern is that even if low-level image features, such as luminance contrast, are good correlates of fixation probability, it is not clear to what extend they drive attention causally (Privitera & Stark, 2000; Carmi & Itti, 2006; Tatler, 2007). For example, the photographer places the object of interest in the center of the picture. In this case, this object might be task relevant, have high low-level feature values (e.g. high contrast), and its position coincides with the central fixation bias. In such a scenario, any of the three factors alone might be as good a predictor for eye movements as all of them combined. Similar problems are encountered in studies that investigate correlations between individual factors and eye movements, without excluding the contributions from other factors. Thus, a disentangled quantification of the individual influences of each of the three, above-mentioned factors is needed.

To quantify the relative contribution of low-level image features, task context, and spatial biases to the selection of fixation points, we sample non-overlapping image patches (also called 'bubbles'; Gosselin & Schyns, 2001, 2002; Schyns et al., 2002; Vinette et al., 2004) from forest scenes and face images. These isolated patches are shown to human participants in different configurations. The participant's eye movements are recorded while they have to solve a classification task. We developed four measures. The stimulus-dependent measure captures low-level feature contrast. It is based on the luminance and texture distribution within each image patch. The second measure, describing the spatial characteristics of eye movements, builds on a baseline study and takes into consideration the global fixation bias and geometrical properties of saccades. The third, task-related measure ignores image features and spatial biases, but quantifies how much information an image patch contains in the context of a specific classification task, based on participant's decision responses. By evaluating the eye-tracking data of the main study, we obtain the fourth measure that captures the empirical salience of each image patch. Having applied these four measures to each image patch, we are able to quantify the overall contributions as well as the individual, non-redundant, contributions of task context, low-level image features, and spatial biases to fixation behavior.

1.5 References

Albright TD, Stoner GR (2002) Contextual influences on visual processing. Annu Rev Neurosci 25: 339–79.

Anderson ML (2003) Embodied Cognition: A field guide. Artif Intell 149: 91–130.

Arieli A, Sterkin A, Grinvald A, Aertsen A (1996) Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. Science 273: 1868–1871.

Baccus SA, Meister M (2004) Retina versus cortex: contrast adaptation in parallel visual pathways. Neuron 42: 5–7.

Bahill AT, Adler D, Stark L (1975) Most naturally occurring human saccades have magnitudes of 15 degrees or less. Invest Ophthalmol 14: 468–469.

Bair W, Cavanaugh JR., Smith MA, Movshon JA (2002) The timing of response onset and offset in macaque visual neurons. J Neurosci 22: 3189–3205.

Ballard DH (1991) Animate vision. Arti Intell 48: 57–86.

Ballard DH, Hayhoe MM, Pook PK, Rao RP (1997) Deictic codes for the embodiment of cognition. Behav Brain Sci 20: 723–767.

Barlow HB (1953) Summation and inhibition in the frog's retina. J Physiol-London 119: 69–

78.
Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: Rosenblith W (editor) Sensory communication. Cambridge (MA): MIT Press: 217–234.
Bell AJ, Sejnowski TJ (1997) The independent components of natural images are edge filters. Vision Res 37: 3327–3338.
Benucci A, Ringach DL, Carandini M (2009) Coding of stimulus sequences by population responses in visual cortex. Nat Neurosci 12: 1317–1324.
Bethge M, Wiecki TV, Wichmann FA (2007) The Independent Components of Natural Images are Perceptually Dependent. Proceedings of SPIE Human Vision and Electronic Imaging XII (EI105).
Betsch B, Einhäuser W, Körding KP, König P (2004) The world from a cat's perspective – statistics of natural videos. Biol Cybern 90: 41–50.
Betz T, Kietzmann TC, Wilming N and König P (2010). Investigating task-dependent top-down effects on overt visual attention. J Vis 10: 1-14.
Blakemore C, Campbell FW (1969) On the existence of neurons in the human visual system selectively sensitive to the orientation and size of the retinal images. J Physiol 202: 237–260.
Blakemore C, Nachmias J (1971) The orientation specificity of several visual aftereffects. J Physiol 213: 157–174.
Blasdel GG, Salama G (1986) Voltage-sensitive dyes reveal a modular organization in the monkey striate cortex. Nature 321: 579–585.
Bonhoeffer T, Grinvald A (1991) Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. Nature 353: 429–431.
Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation selectivity and arrangement of horizontal connections in tree shrew striate cortex. J Neurosci 17: 2112–2127.
Bosking WH, Crowley JC, Fitzpatrick D (2002) Spatial coding of position and orientation in primary visual cortex. Nat Neurosci 5: 874–882.
Brockmann D, Geisel T (1999) Are human scanpaths Levy flights? Artificial Neural Networks 1: 263–268.
Buswell GT (1935) How people look at pictures. A study of the psychology of perception in art. Chicago: University of Chicago Press.
Buzas P, Volgushev M, Eysel UT, Kisvarday ZF (2003) Independence of visuotopic representation and orientation map in the visual cortex of the cat. Eur J Neurosci 18: 957–968.
Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC (2005) Do we know what the early visual system does? J. Neurosci 25: 10577–10597.
Carmi R, Itti L (2006) Visual causes versus correlates of attentional selection in dynamic scenes. Vision Res 46: 4333–4345.
Clark A, Chalmers DJ (1998) The extended mind. Analysis 58: 7–19.
Clark A (1998a) Embodiment and the Philosophy of Mind. In: O'Hear A (editor) Current Issues in Philosophy of Mind: Royal Institute of Philosophy, Supplement 43, Cambridge University Press: 35–52.
Clark A (1998b) Where brain, body, and world collide. Daedalus 127: 257–280.
Clark A (2013) Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. Behav and Brain Sci 36: 181–204.
Clarke PGH, Donaldson IML, Whitteridge D (1976) Binocular visual mechanisms in cortical areas I and II of the sheep. J Physiol 256: 509–526.
Coltheart M (1980) Iconic memory and visible persistence. Percept. Psychophys 27: 183–228.
Das A, Gilbert CD (1997) Distortions of visuotopic map match orientation singularities in primary visual cortex. Nature 387: 594–598.

David SV, Vinje WE, Gallant JL (2004) Natural stimulus statistics alter the receptive field structure of v1 neurons. J Neurosci 24: 6991–7006.
David SV, Gallant JL (2005) Predicting neuronal responses during natural vision. Network 16: 239–260.
Dragoi V, Sharma J, Miller EK, Sur M (2002) Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. Nat Neurosci 5: 883–891.
Duysens J, Orban GA, Cremieux J, Maes H (1985) Visual cortical correlates of visible persistence. Vision Res 25: 171–178.
Duysens J, Schaafsma SJ, Orban GA (1996) Cortical off response tuning for stimulus duration. Vision Res 36: 3243–3251.
Einhäuser W, Moeller GU, Schumann F, Conradt J, Vockeroth J, Bartl K, Schneider E, König P (2009a) Eye–Head Coordination during Free Exploration in Human and Cat. Ann NY Acad Sci 1164: 353–366.
Einhäuser W, Schumann F, Vockeroth J, Bartl K, Cerf M, Harel J, Schneider E and König P (2009b) Distinct roles for eye and head movements in selecting salient image parts during natural exploration. Ann NY Acad Sci 1164: 188-193.
Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN, (1994) fMRI of human visual cortex. Nature 369: 525.
Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex 7: 181–192.
Engel AK, Maye A, Kurthen M, König P (2013) Where's the action? The pragmatic turn in cognitive science. Trends Cogn Sci 17: 202–209.
Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in primate cerebral cortex. Cereb Cortex 1: 1–47.
Felsen G, Shen YS, Yao H, Spor G, Li C, Dan Y (2002) Dynamic modification of cortical orientation tuning mediated by recurrent connections. Neuron 36: 945–954.
Felsen G and Dan Y (2005) A natural approach to studying vision. Nat Neurosci 8: 1643–1646.
Felsen G, Touryan J, Han F, Dan Y (2005) Cortical sensitivity to visual features in natural scenes. PLoS Biol 3: e342.
Fiser J, Chiu C, Weliky M (2004) Small modulation of ongoing cortical dynamics by sensory input during natural vision. Nature 431: 573–578.
Foster KH, Gaska JP, Nagler M, Pollen DA (1985). Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. J Physiol-London 365: 331–363.
Fournier J, Monier C, Pananceau M, Frégnac Y (2011) Adaptation of the simple or complex nature of V1 receptive fields to visual statistics. Nature Neurosci 14: 1053–1060.
Freeman J, Brouwer GJ, Heeger DJ, Merriam EP (2011) Orientation decoding depends on maps, not columns. J Neurosci 31: 4792–4804.
Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360: 815–836.
Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11: 127–138.
Geisler WS (2008) Visual Perception and the Statistical Properties of Natural Scenes. Annu Rev Psychol 59: 167–92.
Goense JBM, Logothetis NK (2006) Laminar specificity in monkey V1 using high-resolution SE-fMRI. Magn Reson Imaging 24: 4381–392.
Goolkasian P (1987) Ambiguous figures: Role of context and critical features. J Gen Psychol 114: 217–228.
Goolkasian P, Woodberry C (2010) Priming effects with ambiguous figures. Attention, Perception, & Psychophysics 72: 168–178.

Gosselin F, Schyns PG (2001) Bubbles: a technique to reveal the use of information in recognition tasks. Vision Res 41: 2261–2271.
Gosselin F, Schyns PG (2002) RAP: a new framework for visual categorization. Trends Cogn Sci 6: 70–77.
Graham N, Nachmias J (1971) Detection of grating patterns containing two spatial frequencies: A test of single-channel and multiple-channels models. Vision Res 11: 251–259.
Grinvald A, Lieke E, Frostig RD, Gilbert CD, Wiesel TN (1986) Functional architecture of cortex revealed by optical imaging of intrinsic signals. Nature 324: 361–364.
Haider B, Krause MR, Duque A, Yu Y, Touryan J, Mazer JA, McCormick DA (2010) Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. Neuron 65: 107–121.
Huang Y, Rao RPN (2011) Predictive coding. WIREs Cogn Sci 2: 580–593.
Hubel DH (1959) Single unit activity in striate cortex of unrestrained cats. J Physiol-London 147: 226–238.
Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. J Physiol-London 148: 574–91.
Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol-London 160: 106–154.
Hubel DH, Wiesel TN (1963a) Shape and Arrangement of columns in cat's striate cortex. J Physiol-London 165: 559–568.
Hubel DH, Wiesel TN (1963b). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. J Neurophysiol 26: 994–1002.
Hubel DH, Wiesel TN (1974) Sequence regularity and geometry of orientation columns in the monkey striate cortex. J Comp Neurol 158: 267–293.
Hubel DH, Wiesel TN, Stryker MP (1978) Anatomical demonstration of orientation columns in macaque monkey. J Comp Neur 177: 361–380.
Hubel DH, Wiesel TN (1979) Brain Mechanisms of Vision. Sci Am 241: 150–162.
Humphrey AL, Norton TT (1980a) Topographic organization of the orientation column system in the striate cortex of the tree shrew (Tupaia glis). I. Microelectrode recording. J Comp Neur 192: 531–547.
Humphrey AL, Skeen LC, Norton TT (1980b) Topographic organization of the orientation column system in the striate cortex of the tree shrew (Tupaia glis). II. Deoxyglucose mapping. J Comp Neur 192: 549–566.
Hurri J, Hyvärinen A, Oja E (1997) Wavelets and Natural Image Statistics. In Proc Scandinavian Conf on Image Analysis '97. 13–18.
Hyvärinen A, Hurri J, Hoyer PO (2009) Natural Image Statistics. A probabilistic approach to early computational vision. Springer-Verlag.
Itti L, Koch C (2001) Computational modelling of visual attention. Nat Rev Neurosci 2: 194–203.
Kandel E, Schwartz JH, Jessell TM (2000) Principles of neural science, fourth edition. Elsevier: NewYork: Chapter 26.
Kayser C, Einhäuser W, König P (2004) Temporal correlations of orientations in natural scenes. Neurocomputing 52-54: 117–123.
Kietzmann TC, Geuter S, König P (2011) Overt Visual Attention as a Causal Factor of Perceptual Awareness. Plos One 6: 1–9.
Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol 4: 219–227.
Kohn A (2007) Visual adaptation: Physiology, mechanisms, and functional benefits. J Neurophysiol 97: 3155–3164.
König P, Luksch H (1998) Active sensing - closing multiple loops. Z Naturforsch C 53: 542-

549.
Land M, Mennie N, Rusted J (1999) The roles of vision and eye movements in the control of activities of daily living. Perception 28: 1311–1328.
Lee C, Weyand TG, Malpeli JG (1998). Thalamic control of cat area-18 supragranular layers: Simple cells, complex cells, and cells projecting to the suprasylvian visual area. Vis Neurosci 15: 27–35.
Lee SH, Blake R, Heeger DJ (2005) Traveling waves of activity in primary visual cortex during binocular rivalry. Nat Neurosci 8: 22–23.
Lee TS (2009) Contextual Influences in Visual Processing. In: Binder MD, Hirokawa N, Windhorst U (eds) Encyclopedia of Neuroscience. Springer-Verlag: 867–871.
Leeper R (1935) A study of a neglected portion of the field of learning: The development of sensory organization. J Genet Psychol 46: 41–75.
Lennie P (2003) Receptive fields. Curr Biol 13: R216-R219.
Li Z (2002) A saliency map in primary visual cortex. Trends Cogn Sci 6: 9–16.
Malcolm C, Smithers T, Hallam J (1989) An Emerging Paradigm in Robot Architecture. Department of Artificial Intelligence, University of Edinburgh. Research Paper 447.
Malpeli JG (1983) Activity of cells in area 17 of the cat in absence of input from layer A of lateral geniculate nucleus. J Neurophys 49: 595–610.
Malpeli JG, Choongkil L, Schwark HD, Weyand TG (1986). Cat area 17. I. Pattern of the thalamic control of cortical layers. J Neurophys 56: 1062–1073.
Mante V, Frazor RA, Bonin V, Geisler WS, Carandini M (2005) Independence of luminance and contrast in natural scenes and in the early visual system. Nat Neurosci 8: 1690–1697.
Marr D (1982) Vision. San Francisco, CA: W. H. Freeman.
Martin KAC (1994) A brief history of the "feature detector". Cereb Cortex 4: 1–7.
Martinez-Conde S, Macknik SL & Hubel DH (2004) The role of fixational eye movements in visual perception. Nat Rev Neurosci 5: 229–240.
Maturana HR, Varela FJ (1984) El árbol del concocimiento. German version: Der Baum der Erkenntnis. Die biologischen Wurzeln menschlichen Erkennens. Goldmann Verlag. 1987, 11. Edition.
McCormick DA, Shu Y, Hasenstaub A, Sanchez-Vives M, Badoual M, Bal T (2003) Persistent cortical activity: mechanisms of generation and effects on neuronal excitability. Cereb Cortex 13: 1219–1231.
Movshon JA, Lennie P (1979) Pattern-selective adaptation in visual cortical neurones. Nature 278: 850–852.
Müller JR, Metha AB, Krauskopf J, Lennie P (1999) Rapid adaptation in visual cortex to the structure of images. Science 285: 1405–1408.
Ohki K, Chung S, Ch'ng YH, Kara P, Reid RC (2005) Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. Nature 433: 597–608.
Olman CA, Harel N, Feinberg DA, He S, Zhang P, Ugurbil K, Yacoub E (2012) Layer-Specific fMRI Reflects Different Neuronal Computations at Different Depths in Human V1. PLoS ONE 7: e32536.
Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381: 607–609.
Olshausen BA, Field DJ (2005) How close are we to understanding V1? Neural Comput 17: 1665–1699.
Onat S, König P, Jancke D (2011). Natural scene evoked population dynamics across cat primary visual cortex captured with voltage-sensitive dye imaging. Cereb Cortex 21: 2542-2554.
O'Regan JK, Noë A (2001) The sensorimotor account of vision and visual consciousness. Behav Brain Sci 24: 939–1031.
Parkhurst D, Law K, Niebur E (2002) Modeling the role of salience in the allocation of overt

visual attention. Vision Res 42: 107–123.
Payne BR, Peters A (2002) The cat primary visual cortex. San Diego (CA): Academic Press.
Peters A, Payne BR (1993) Numerical relationships between geniculocortical afferents and pyramidal cell modules in cat primary visual cortex. Cereb Cortex 3: 69–78.
Peters A, Payne BR, Budd J (1994) A numerical analysis of the geniculocortical input to striate cortex in the monkey. Cereb Cortex 4: 215–29.
Peters RJ, Iyer A, Itti L, Koch C (2005) Components of bottom-up gaze allocation in natural images. Vision Res 45: 2397–2416.
Privitera CM, Stark LW (2000) Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations. IEEE Trans Pattern Anal Mach Intell 22: 970–982.
Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2: 79–87.
Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. Psychol Bull 124: 372–422.
Redies C, Diksic M, Riml H (1990) Functional organization in the ferret visual cortex: A double-label 2-deoxyglucose study. J Neurosci 10: 2791–2803.
Ress D, Glover GH, Liu J, Wandell B (2007) Laminar profiles of functional activity in the human brain. Neuroimage 34: 174–84.
Ringach DL, Hawken MJ, Shapley R (1997) Dynamics of orientation tuning in macaque primary visual cortex. Nature 387: 281–284.
Ritt JT, Andermann ML, Moore CI (2008) Emdodied information processing: vibrissa mechanics and texture features shape micromotions in actively sensing rats. Neuron 57: 599–613.
Rubin E (1915) Synsoplevede Figurer. Copenhagen: Gyldendals.
Rust NC, Movshon JA (2005) In praise of artifice. Nat Neurosci 8: 1647–1650.
Sanchez-Vives MV, Nowak LG, McCormick DA (2000) Cellular mechanisms of long-lasting adaptation in visual cortical neurons. J Neurosci 20: 4286–4299.
Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P (2010) Dynamics of Active Sensing and Perceptual Selection. Curr Opin Neurobiol 20: 172–176.
Schyns PG, Bonnar L, Gosselin F (2002) Show me the features! Understanding recognition from the use of visual information. Psychol Sci 13: 402–409.
Sere B, Marendaz C, Herault J (2000) Nonhomogeneous resolution of images of natural scenes. Perception 29: 1403–1412.
Sharon D, Grinvald A (2002) Dynamics and constancy in cortical spatiotemporal patterns of orientation processing. Science 295: 512–515.
Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24: 1193–1216.
Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M (1997) Adaptation of retinal processing to image contrast and spatial scale. Nature 386: 69–73.
Smyth D, Willmore B, Baker GE, Thompson ID, Tolhurst DJ (2003) The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. J Neurosci 23: 4746–4759.
Sperry RW (1967) Split-Brain. Approach to Learning Problems. In: Quarton GC, Melnechuk T, Schmitt FO (eds) The Neurosciences: A Study Program. New York: Rockefeller University Press: 714–722.
Sperry RW (1974) Lateral Specialization in the Surgically Separated Hemispheres. In: Schmitt FO, Worden FG (eds) The Neurosciences Third Study Program. Cambridge (MA): MIT Press: 5–19.
Tatler BW, Baddeley RJ, Gilchrist ID (2005) Visual correlates of fixation selection: effects of scale and time. Vision Res 45: 643–659.
Tatler BW, Baddeley RJ, Vincent BT (2006) The long and the short of it: spatial statistics at

fixation vary with saccade amplitude and task. Vision Res 46: 1857–1862.
Tatler BW (2007) The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. J Vis 7: 1–17.
Thomas J, Moss C, Vater M (2004) Echolocation in bats and dolphins. Chicago: University of Chicago Press.
Tootell RB, Silverman MS, Switkes E, De Valois RL (1982) Deoxyglucose analysis of retinotopic organization in primate striate cortex. Science 218: 902–904.
Vinette G, Gosselin F, Schyns PG (2004) Spatio-temporal dynamics of face recognition in a flash: it's in the eyes. Cogn Sci 28: 289–301.
von der Emde G, Schwartz S (2003) Imaging of objects through electrolocation in Gnathonemus petersii. Journal of Physiology-Paris. 96: 431–444.
Watson AB, Barlow HB, Robson JG (1983) What does the eye see best? Nature 302: 419–422.
Wiesel TN, Hubel DH (1965) Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. J Neurophysiol 28: 1029–1040.
Wolfe JM, Horowitz TS (2004) What attributes guide the deployment of visual attention and how do they do it? Nat Rev Neurosci 5: 495–501.
Yacoub E, Harel N, Ugurbil K (2008) High-field fMRI unveils orientation columns in humans. Proc Natl Acad Sci USA 105: 10607–10612.
Yarbus AL (1967) Eye movements and vision. New York: Plenum.
Yu H, Farley BJ, Jin DZ, Sur M (2005) The coordinated mapping of visual space and response features in visual cortex. Neuron 47: 267–280.
Zucker RS, Regehr WG (2002) Short-term synaptic plasticity. Annu Rev Physiol 64: 355–405.

2. Independent encoding of grating motion across stationary feature maps in primary visual cortex visualized with voltage-sensitive dye imaging (Study 1)

This section is based on the following publication in NeuroImage: S Onat[1], N Nortmann[123], S Rekauzke[23], P König[1], D Jancke[23] (2011) Independent encoding of grating motion across stationary feature maps in primary visual cortex visualized with voltage-sensitive dye imaging. NeuroImage 55: 1763-1770. *[1]Institute of Cognitive Science, Department of Neurobiopsychology, University Osnabrück, 49069 Osnabrück, Germany. [2]Cognitive Neurobiology, Ruhr-University Bochum, 44780 Bochum, Germany. [3]Bernstein Group for Computational Neuroscience, Institut für Neuroinformatik, Ruhr-University Bochum, 44780 Bochum, Germany.*

2.1 Abstract

In early visual cortex different stimulus parameters are represented in overlaid feature maps. Such functioning was extensively explored by the use of drifting gratings characterized by orientation, spatial–temporal frequency, and direction of motion. However surprisingly, the direct cortical visuotopic drift of the gratings' stripy pattern has never been detected simultaneously to these stationary feature maps. It therefore remains to be demonstrated how physical signals of grating motion across the cortex are represented independently of other parametric maps and thus, how multi-dimensional input is processed independently to enable effective read-out further downstream.

Taking advantage of the high spatial and temporal resolution of voltage-sensitive dye imaging, we here show the real-time encoding of position and orientation. By decomposing the cortical responses to drifting gratings we visualize the typical emergence of stationary orientation maps in which specific domains exhibited highest amplitudes. Simultaneously to these patchy maps, we demonstrate coherently propagating waves of activity that precisely matched the actual movement of the gratings in space and time, most dominantly for spatial frequencies lower than the preferred range. Thus, the primary visual cortex multiplexes information about retinotopic motion by additional temporal modulation of stationary orientation signals. These signals may be used to variably extract coarse-grained object motion and form information at higher visual processing stages.

2.2 Introduction

Fifty years ago, neurons in primary visual cortex were discovered to have local receptive fields that are selective for stimulus orientation (Hubel & Wiesel, 1959). Moreover, it has been found that neurons are grouped systematically across the cortical plane representing visual space and contours in overlaid retinotopic and orientation maps (Hubel & Wiesel, 1974). Intrinsic optical imaging revealed further details of the distinct cortical functional organization establishing the columnar arrangement of ocular dominance, motion direction, and spatio-temporal frequency maps (Blasdel & Salama, 1986; Bonhoeffer & Grinvald, 1991; Bosking et al., 1997; Hübener et al., 1997; Shoham et al., 1997; Xu et al., 2006). But how are these different information channels efficiently encoded across large neuronal populations without undesirable interferences?

In primary visual cortex the precise temporal structure of the neuronal spike patterns in the gamma range [40–80 Hz] was proposed as a fundamental mechanism to enable binding as well as segmentation of visual features (Abeles, 1991; Gray et al., 1989), in addition to mechanisms in which amplitudes of instantaneous firing rates of coactive neurons (Roelfsema

et al., 2004), or entire spike trains (Jancke, 2000; Richmond et al., 1990), all play an important role in processing (Kayser et al., 2009). Intrinsic optical imaging research to date shows that moving gratings – as a means to provide simultaneous multi-dimensional visual input – are cortically represented by unique patchy activity patterns where the locations of the activated domains are interpreted to either reflect the intersection of orientation, spatial frequency, and direction maps (Blasdel & Salama, 1986; Bonhoeffer & Grinvald, 1991; Bosking et al., 1997; Hubel & Wiesel, 1974; Hübener et al., 1997; Shoham et al., 1997; Xu et al., 2006), or to encode the overall spatio-temporal energy of these features (Basole et al., 2003; Mante & Carandini, 2003; see Issa et al., 2008 for recent review). In either case, since multiple feature combinations can produce an identical stationary map (Basole et al., 2003) its information content remains ambiguous (Movshon et al., 1985; Rust et al., 2006; Zhang & Britten, 2006). Particularly in the event of moving gratings such feature map fails to additionally capture the physical motion of the grating across the retinotopic representation and therefore, neglects the permanent space–time changes in luminance caused by the gratings' stripes. Instead, given that retinotopy is independent of orientation (Bosking et al., 2002; Buzas et al., 2003; Yu et al., 2005, but see Das & Gilbert, 1997), smoothly traveling waves of activity (Engel et al., 1994, 1997; Lee et al., 2005) should be detectable at the same time.

However, the applied recording methods so far were limited in sampling: single electrode recordings hamper a coherent population picture of activation across the cortex due to restrictions in spatial sampling, whereas intrinsic optical imaging is limited in temporal resolution because of its dependence on slow hemodynamic processes. Also with fMRI, as another method that uses hemodynamic signals, retinotopic waves and orientation maps were never captured simultaneously (Yacoub et al., 2008). To overcome these limitations, we made use of an imaging technique that employs voltage-sensitive dye to record population activity in the millisecond range across several millimeters of cat visual cortex with high resolution (Grinvald et al., 1994; Jancke et al., 2004; see Grinvald & Hildesheim, 2004 for review).

2.3 Materials and methods

We recorded from cat primary visual cortical area spanning A17 and A18 (11 animals, 12 hemispheres). Detailed analysis was performed in 6 experiments, in which the compound spatio-temporal activity was found. Standard surgical and experimental procedures were used, approved by the German Animal Care and Use Committee (AZ 9.93.2.10.32.07.032) in accordance with the Deutsche Tierschutzgesetz and NIH guidelines.

2.3.1 Animal preparation

Animals were initially anesthetized with ketamine (15 mg kg^{-1} i. m.) and xylazine (1 mg kg^{-1} i.m.), supplemented with atropine (0.05 mg kg^{-1} i.m.). After tracheotomy, animals were artificially respirated, continuously anesthetized with 0.8–1.5% isoflurane in a 1:1 mixture of O2/N2O, and fed intravenously. Heart rate, intratracheal pressure, expired CO2, body temperature, and EEG were monitored during the entire experiment. The skull was opened above area 17/18 and the dura was resected. Paralysis was induced and maintained by Alloferin®. Eyes were covered with zero-power contact lenses as protectives. External lenses were used to focus the eyes on the screen. To control for eye drift, the position of the area centralis and receptive field positions were repeatedly measured. A stainless steel chamber was mounted and the cortex was stained for 2–3 h with voltage-sensitive dye (RH-1691), and unbound dye was subsequently washed out.

2.3.2 Stimulus presentation

Gratings (0.1/0.2/0.4 c/°, 6.25 Hz were presented on a Monitor (100 Hz, mean luminance 11 cd m− 2, Sony Triniton GDM-FW900, Japan) covering a visual field of 30°×40°. To detect the area 17/18 border, gratings of 0.6 c/° and 2 Hz were additionally used. Eyes were converged using a prism. Stimuli were randomly displayed for 2 s (including 200 ms of prestimulus duration), followed by a uniform gray screen (blank condition) presented for 15 s during the interstimulus intervals.

2.3.3 Data acquisition and pre-processing

Optical imaging was accomplished using an Imager 3001 (Optical Imaging Inc, Mountainside, NY) The camera was focused ~ 400 nm below the cortical surface. Data acquisition onset was synchronized with heart-beat signal. For detection of changes in fluorescence the cortex was illuminated with light of 630 ± 10 nm and emitted light was high-pass filtered with a cut-off of 665 nm (camera frame rate 220 Hz). Divisive normalization was performed by dividing each pixel value by its DC level during 200 ms pre-stimulus time (in which a uniform gray screen was presented); heart-beat and respiration-related artifacts were removed by subtracting the average blank signal.

2.3.4 Singular value decomposition analysis

Evoked signals were computed by averaging 25 to 50 stimulus repetitions. For singular value decomposition (SVD) analysis we chose the time interval in which activity roughly stabilized (>400 ms). Technically, the entire time course or a time window starting closer to response onset could be used. However, inclusion of the transient response part, characterized by huge changes in overall activity, reduces sensitivity of the method to the later oscillatory part of the responses. We applied SVD to the evoked activity by using *svd* command of Matlab software (Mathworks, Natick, MA, USA) after transforming each frame into a vector. SVD transforms the data matrix representing the evoked signal, $S(t,x)$, into a weighted sum of N space–time separable matrices, $g_i(t,x)$, such that $S(t,x) = \sum_i^N \gamma_i g_i(\mathrm{t},\mathrm{x})$; where γ_i represents the weight of the i^{th} modes and N, the number of frames. The matrix $g_i(t,x)$ can be described as the outer product of two vectors $u_i(x)$ and $v_i(t)$ representing the i^{th} spatial and temporal modes. Interpreting the rows of $S(t,x)$ as samples and the columns as features, $u_i(x)$ and $v_i(t)$ are the eigenvectors of the sample and feature covariance matrices i.e. principal components, respectively. Singular values, γ_i increased linearly with logarithmic scale and significant components were detected by identifying the first component with significant increase in its weight with respect to the previous components. To selectively reconstruct the oscillatory activity, we selected those components with a prominent peak at the required frequency. In some cases fitting harmonic functions to the temporal modes at the required frequency beforehand improved the results.

2.4 Results

Cortical responses to drifting gratings were measured that moved either rightwards or downwards across the contra-lateral visual field (Fig. 2.1a). These stimuli give rise to complementary activity pattern as demonstrated by the difference map (Fig. 2.1a, second row, leftmost). The lower graph in Fig. 2.1a shows the evoked time courses of activity of one experiment averaged across the imaged region for each grating direction. Activity rises steeply after stimulus onset, suspended by an intermediate deceleration–acceleration notch after ~85 ms (Sharon & Grinvald, 2002), reaching maximum levels after 300 ms, followed by a slow monotonic decay.

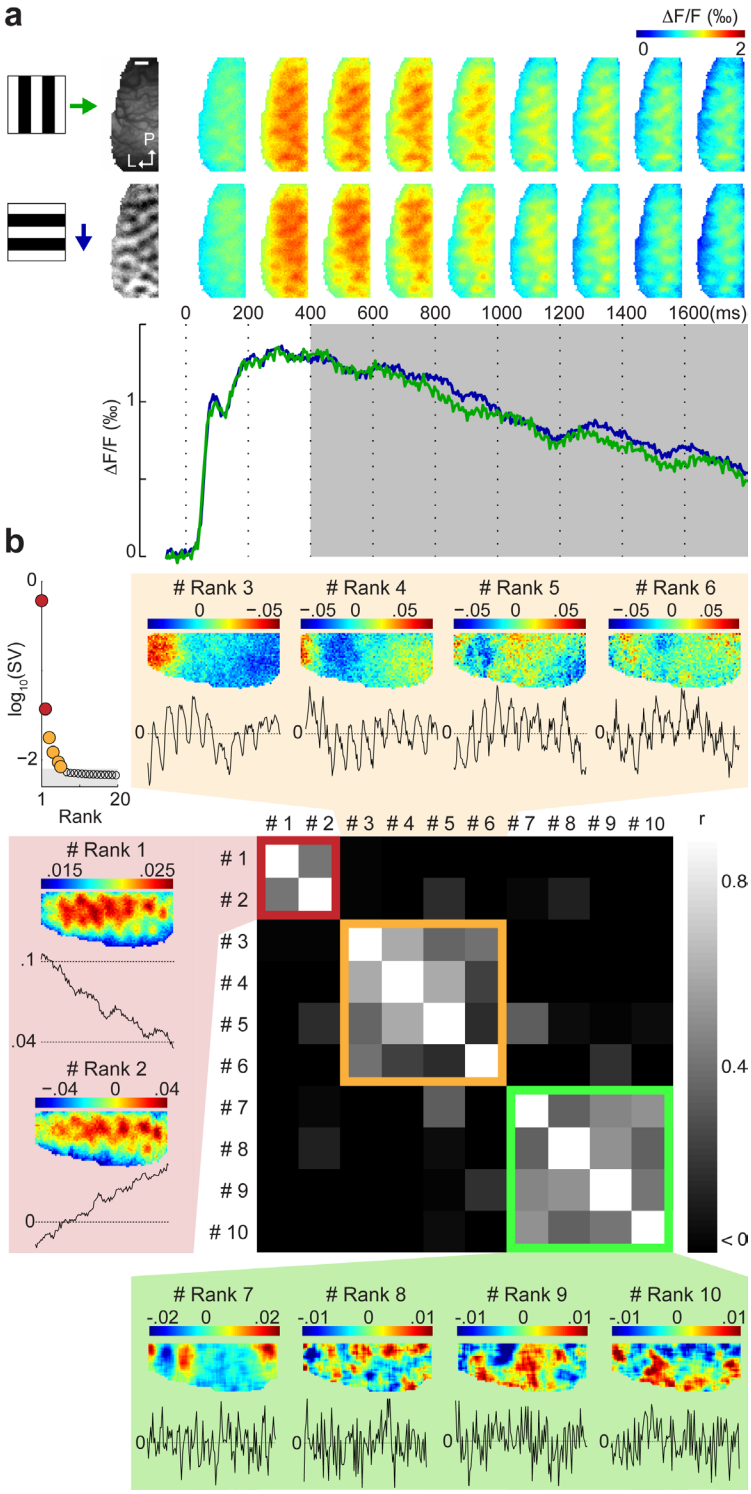


Figure 2.1 Decomposition of evoked cortical responses to gratings of 0.2 c/° drifting for 2 s at a temporal frequency of 6.25 Hz. (a) Evoked spatio-temporal activity patterns (top rows) and time courses obtained by spatial averages across the images (bottom traces) expressed as fractional change in fluorescence relative to blank condition (ΔF/F). Top left frame shows the vascular image of the recorded right hemisphere, P = posterior, L = lateral; here and in all figures scale bar 1 mm. Leftmost frame in 2nd row depicts the time-averaged orientation map derived by subtracting evoked responses to the vertical grating from horizontal. Green trace = responses to vertical grating, drifting rightwards in visual space; blue trace = horizontal grating, drifting downwards. (b) Top left corner, singular values, $\boldsymbol{g_i}$, ranked in order of their contributions. Components of significant contribution to variance are colored (grey area depicts significance level). The contribution of each single SVD component to single recorded trials (n=35) was computed, their correlations across trials are represented as a matrix. Spatial ($\boldsymbol{u_i(x)}$) and temporal ($\boldsymbol{v_i(t)}$) modes of the SVD components were clustered according to their correlation (red, yellow, and green boxes; curves represent weight of each spatial mode [y-axes] as a function of time [400–1800 ms]).

Singular-value decomposition (SVD) was applied to the later part of the evoked responses (Fig. 2.1a, grey area in bottom traces, 400–1800 ms) excluding the fast transient onset (see Material and methods, SVD analysis). SVD decomposes the evoked signal into distinct orthonormal components that are separable into their spatial and temporal modes. The relative importance of each component, i.e. fraction of variance explained, is captured by the size of the singular value (Fig. 2.1b, top-left). We observed 6 significant singular values spanning a range of two orders of magnitude. This shows that the responses to gratings are characterized by compound dynamics that result in spatio-temporal inseparability of activity.

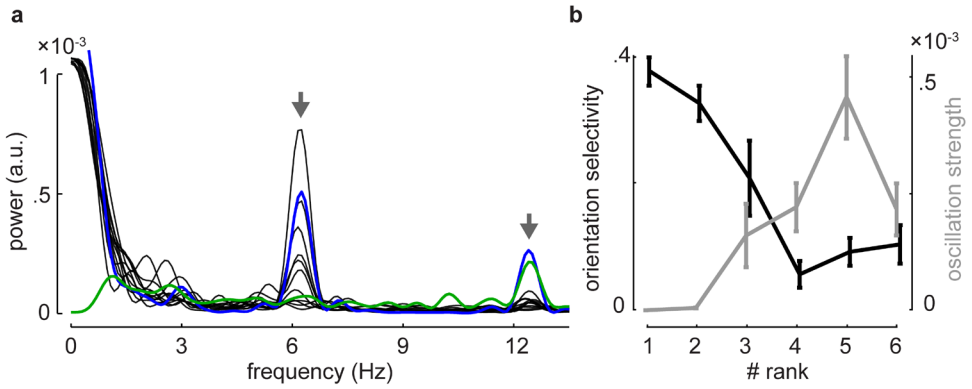


Figure 2.2 (a) Power spectra of temporal modes averaged across significant components for all experiments (black lines). In 6 of 12 hemispheres a prominent peak at 6.25 Hz was observed (see arrowhead). For the experiment presented in Fig. 2.1, components 7 to 10 were characterized by another prominent peak at 12.5 Hz (green line, see 2nd arrowhead). Blue line depicts spectrum characteristics of responses to a grating of lower spatial frequency (0.1 c/°, cf. Fig. 2.10, bottom rows). (b) Orientation selectivity of spatial modes (black line) and power at 6.25 ± 1 Hz (gray line) are shown for the first 6 SVD components across all experiments. Orientation selectivity was evaluated by quantifying the correlation between spatial modes and differential orientation map (shown in Fig. 2.1a). Power at 6.25 Hz (± 1.00 Hz) was computed for each component. Differences in the singular values were accounted by weighting each component's power by its singular value. With increasing rank the power at 6.25 Hz increased while the correlation of the spatial modes to the orientation maps decreased. Error bars indicate SEM. As a control, we additionally compared the power at 6.25 (± 1.00 Hz) to a baseline (i.e. power at flanking frequencies, 3.25 Hz–5.25 Hz and 7.25 Hz–9.25 Hz). For the first cluster (components 1 to 2) the power at 6.25 Hz was on average 1.21 ± 0.18 times higher, but was increased by a factor of 8.63 ± 1.56 for the second cluster (components 3 to 6), indicating that oscillatory components were almost independent from cortical representation of orientation.

To infer the biological significance of the SVD components we reasoned that the contributions of individual components must exhibit co-variation across trials. Indeed, the respective correlation matrix revealed different clusters of SVD components (Fig. 2.1b). The first cluster encompasses two SVD components with highest singular values, each displaying non-oscillatory tonic activity modes (Fig. 2.1b, red/left). This was observed in all measured hemispheres. The second cluster included 4 SVD components with smaller singular values and strong oscillatory activity (Fig. 2.1b, orange/top) matching the fundamental frequency of the moving grating (6.25 Hz). For the present example, additional SVD components followed (Fig. 2.1b, green/ bottom) displaying distinct oscillations at the 2nd harmonic of the grating's temporal frequency. We found that these oscillatory dynamics were almost exclusively accounted for by those higher order SVD components that did not have orientation selective spatial modes (see Fig. 2.2a for average power spectra across the significant temporal modes and Fig. 2.2b for relationship between their oscillatory strength and orientation selectivity across all experiments).

Reconstruction of activity using the first cluster of SVD components displayed the typical patchy structure of stationary orientation maps characterized by repeating local domains of peak amplitudes (Fig. 2.3, see contours in red panels, first and third rows). The regions maximally activated by the different orientations are largely non-overlapping resulting, as expected, in orthogonal maps of activation (Fig. 2.5). Bear in mind that other stimulus dimensions are mapped within primary visual cortex as well. For example, regions differ in their responsiveness to stimuli of different spatial frequency (Shoham et al., 1997). Yet, orientation maps determined by gratings of different spatial–temporal frequencies are highly overlapping (Fig. 2.6). Thus, we conclude that in the present experiment the response captured by the first cluster of SVD components is dominated by the orientation selectivity of neurons in primary visual cortex.

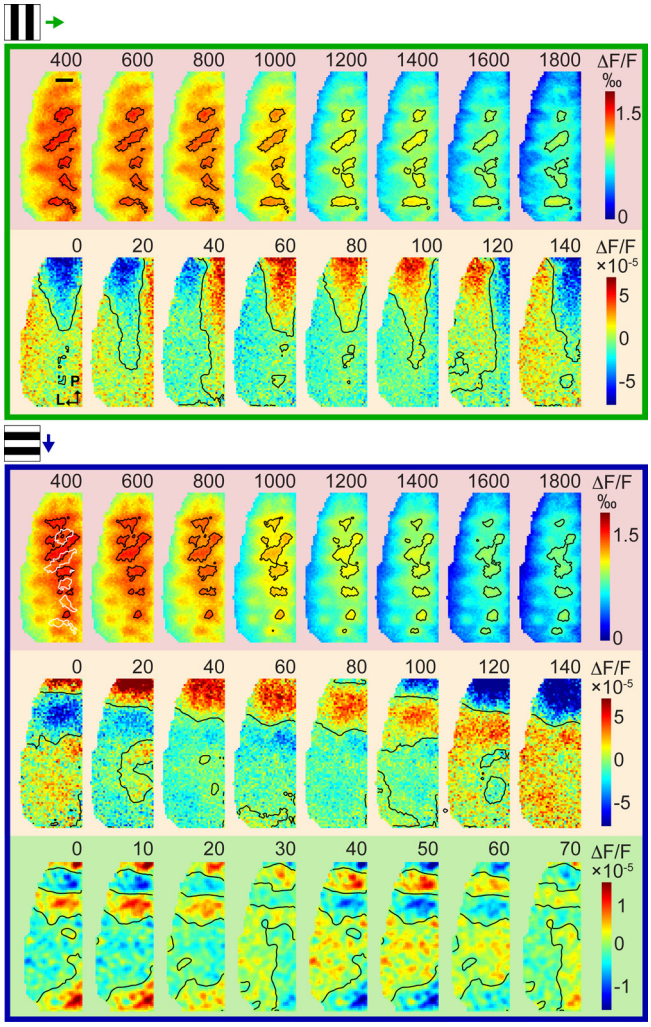


Figure 2.3 Propagation of retinotopic cortical activity across stationary orientation maps. Icons on top sketch stimulus conditions. Within each panel, spatio-temporal activity dynamics represented by different SVD components are shown (color code as in Fig. 2.1b): tonic components (red panels, top rows), oscillatory components with principle oscillation frequencies at 6.25 (orange panels, second rows) and 12.5 Hz (green panel, bottom row). Time after stimulus onset (red panels) or relative to the propagation period (orange and green panels) is indicated above each single frame. Note the difference in the time-scales. To capture a full cycle of propagating activity, only a single period (duration 160 or 80 ms) is depicted (multiple cycles were averaged). Contour lines are drawn around 90th percentiles of activity for the tonic components (contours in first frame of upper box were copied as white outlines to first frame in lower box for comparison) and at zero crossings for the oscillatory components. Color bars depict activity levels, ΔF/F; note the two-order difference in amplitudes between the two scales. See Fig. 2.4 for rough retinotopic sampling of the imaged cortical area using electrophysiology.

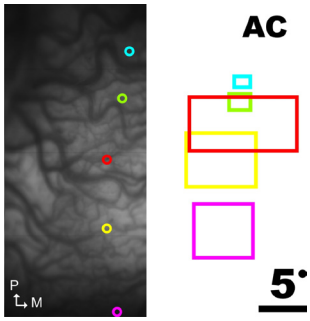


Figure 2.4 Relationship between direction of motion across the cortex and its retinotopic layout. Several penetration sites along the posterior-anterior cortical axis were done prior to the recording and the corresponding receptive field locations (hand-mapped) in visual space are shown. Colors match recording sites (AC=projection of area centralis). Note that for the grating moving downwards (Fig. 2.3, bottom rows) activity propagated from posterior towards anterior regions in the expected cortical direction.

Overlaid on these maps, the second cluster of oscillatory SVD components (orange panels/second and fourth rows) revealed cyclic waves of activity that propagated either medial to lateral (right to left in image frames) or in posterior–anterior direction (from top downwards) across the cortex depending on the gratings' drifting direction. Evidently, the retinotopic propagation was not restricted to particular regions as it passed coherently through both preferred orientation and orthogonally tuned domains. Moreover, the fact that these waves propagated with the same temporal frequency as the grating suggests an underlying asymmetry in the responses to its dark and bright stripes (see Discussion). In addition, the higher order SVD components (7 to 10) exhibited waves of half spatial wavelength, thus reporting the individual contrast changes at each of the grating edges (Fig. 2.3, last row).

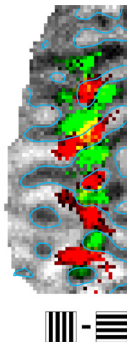


Figure 2.5 Orientation related activity components derived from SVD analysis overlaid on conventional subtractive map (vertical minus horizontal, cf. Fig. 2.1; grey values were clipped to ±2 SD). Most active pixels of tonic SVD components (90th percentile of activity) were plotted for each of the single conditions, vertical grating (red), horizontal (green; yellow pixels mark minor overlap of red and green). Brightness is proportional to the number of temporal frames a given pixel was above threshold (with brightest pixels denoting the entire length of the analyzed temporal interval [400-1800 ms]).

We next extracted the phase of the best fitting harmonic functions (Fig. 2.7a) for each pixel and plotted these values topographically (Fig. 2.7b). The change in phase as a function of cortical distance (Fig. 2.7c) resulted in values of ~ 30 and ~ 39 mm/s (for medial–lateral and anterior–posterior direction, respectively). Given the speed of the drifting gratings (~31 °/s) this cortical propagation speed matches to the known magnification factor of the central visual field representation of area 17 (Tusa et al., 1978). Hence, overall the oscillatory SVD components reflected the shifts of the gratings' stripes with high spatial and temporal accuracy. Notably in this experiment the signal-to-noise ratio of the propagating waves was highest medial to the 17/18 areal border (Fig. 2.7b, see dashed white lines in the respective phase plots, and Fig. 2.8). However, when a grating of lower spatial frequency (0.1 c/°) was used in another experiment, propagation was evoked concurrently across both primary visual areas (Fig. 2.7e; see Fig 2.9 for space–time diagram). We calculated the power of the oscillatory components dependent on the grating spatial frequency for each area 17 and 18 separately. At all spatial frequencies investigated, oscillatory power in area 17 was higher than in area 18 (Fig. 2.7f), on average by a factor of 1.7. This may suggest that the emergence of retinotopic grating propagation is area specific and might lead to the situation where the retinotopic component is independently processed in area 17 while in area 18 such multiplexing is absent or weak.

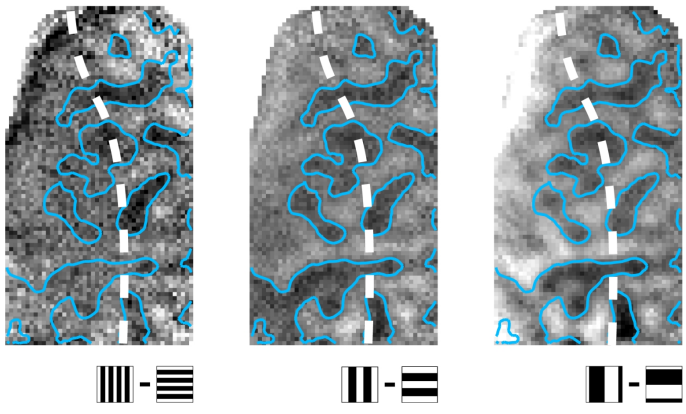


Figure 2.6 Subtractive orientation maps (vertical-horizontal) derived from different spatio-temporal frequencies of the gratings. From left to right: 0.6 c/° (2 Hz), 0.2, 0.1 c/° at 6.25 Hz. Each map depicts time average, 400-1800 ms, and was clipped to ±2 SD. Blue contours separate pixels preferring vertical (bright) from horizontal (dark). To facilitate comparison, contours from the middle plot were copied to the left and right plots. High spatial frequencies produced a strong pattern in area 17 (right to dashed line). As expected, with decreasing spatial- and increasing temporal frequencies also orientation domains in area 18 emerged (left to dashed line). Importantly, as shown in Shoham et al. (1997), in area 17 small systematic differences between orientation maps of different spatial frequencies reflect an additional underlying mapping of distinct low frequency domains (preferring high speeds) and high frequency domains (preferring low speeds).

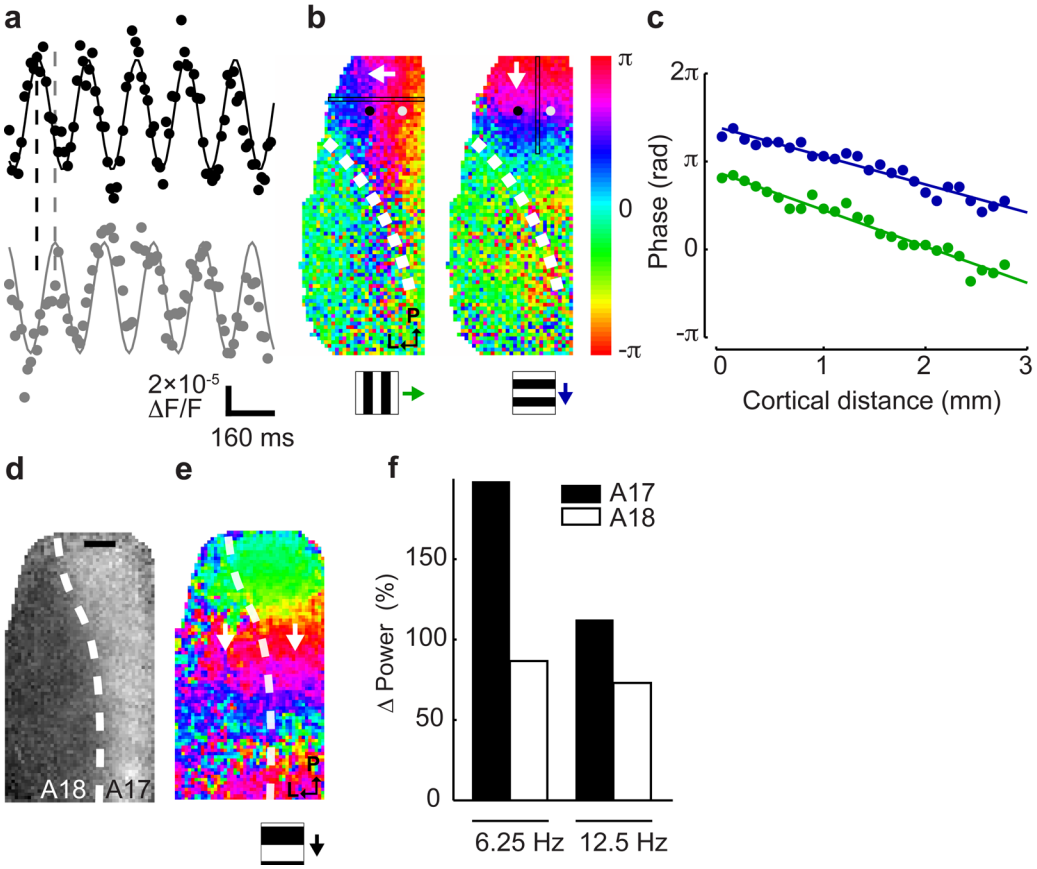


Figure 2.7 Phase vs. space: propagation speed of cortical activity. (a) Oscillatory activity around two pixels separated by 1 mm along the medial–lateral axis (grey and black dots in panel (b). Oscillatory activity was best described by a harmonic function with a period of 160 ms (black and grey fitted curves). (b) The phase of each pixel is shown topographically for both grating conditions. Estimate of A17/18 border is shown as stippled line. Arrows indicates direction of motion. (c) Change of phase as a function of space within circumscribed regions (see vertical and horizontal rectangles in (b). The slope of the fitting lines was on average across both propagation directions ~1.2 radians per mm. (d) Another experiment (right hemisphere) where the A17/18 border was mapped with gratings of 0.6 c/°. Stippled line delineates at 75th percentile. (e) Phase map computed from oscillatory components in response to a downward drifting grating of 0.1 c/° spatial frequency. (f) Increase of oscillatory power in A17 and A18 in percentage of baseline power at 6.25 Hz and 12.5 Hz across all conditions.

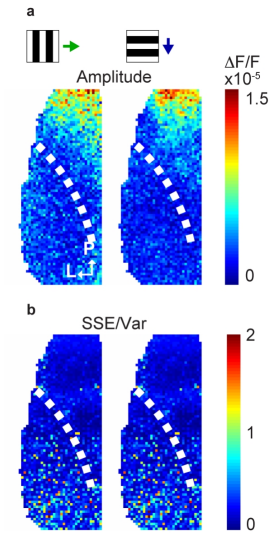


Figure 2.8 Amplitudes and signal-to-noise of oscillatory activity in response to a 0.2 c/deg spatial frequency grating. Using significant oscillatory SVD components, propagating waves of activity were reconstructed representing vertical (left) and horizontal (right) grating motion. Different cycles of propagation were averaged. To the temporal course of activity at each pixel a harmonic function was fitted (cf. Fig. 2.7a). (a) The best fitting amplitude parameters are represented topologically. (b) The goodness of the fit was evaluated by measuring the squared sums of errors normalized by the variance for each pixel separately. Note that both amplitudes and signal-to-noise ratios decreased gradually with increasing proximity to the area 18 border (roughly estimated by stereotactic coordinates, see stippled line).

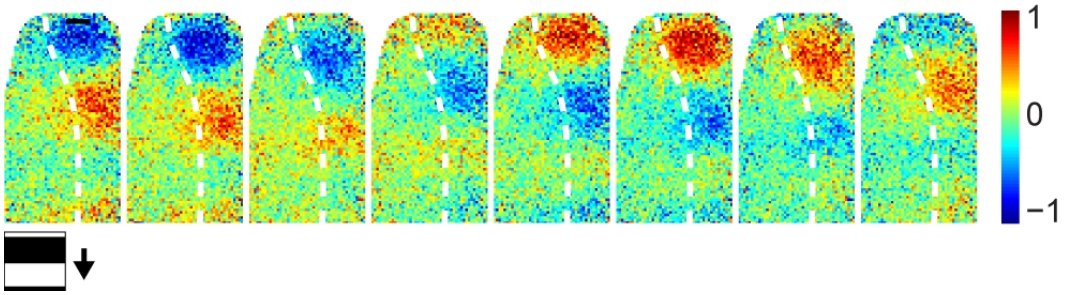


Figure 2.9 Propagation of activity across primary visual cortex (right hemisphere) in response to a grating with 0.1 c/° spatial frequency moving downwards. Propagation is most prominently seen across area 17 (right from stippled white line). Activity was reconstructed using SVD components with dominant power at 6.25 Hz. Same conventions as in Fig. 2.3 and 2.7; the data are depicted as phase plot in Fig. 2.7e; see space-time plot of 12.5 Hz related components across area 17 in Fig. 2.10 bottom row).

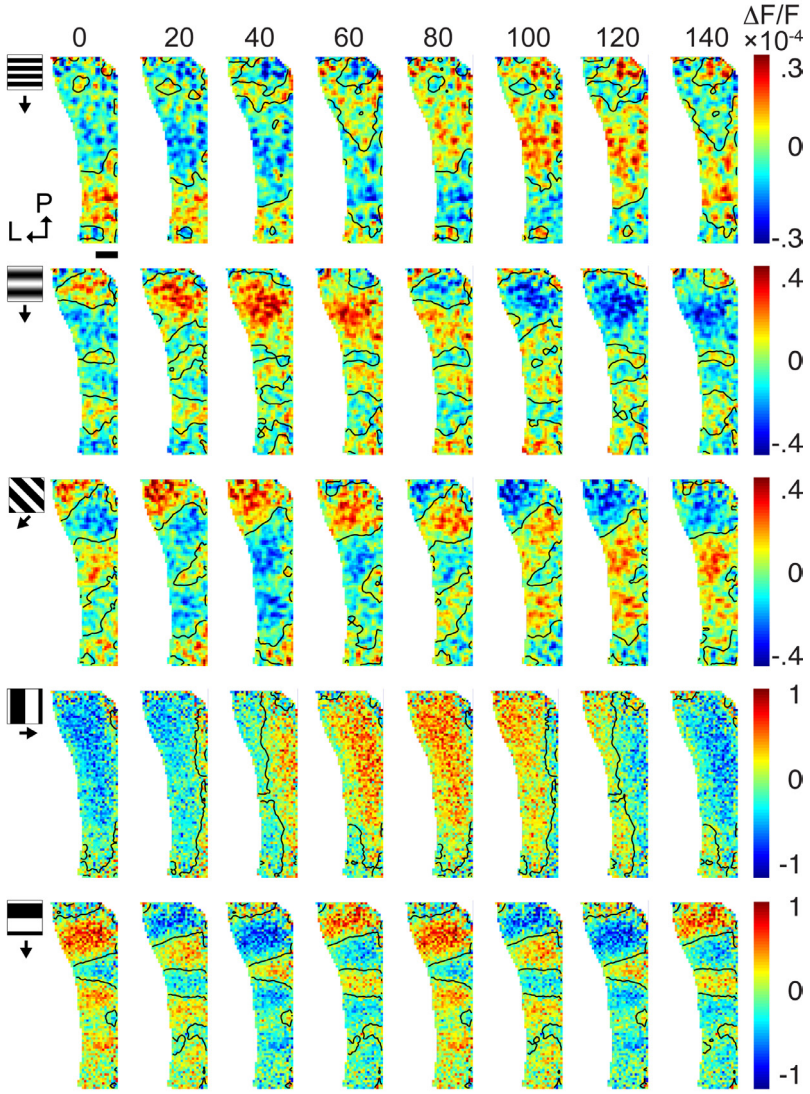


Figure 2.10 Retinotopic propagation of cortical activity in response to gratings varying spatial parameters. SVD components with principle oscillation at either 6.25 Hz (rows 1–4) or 12.5 Hz (last row) were used to reconstruct the oscillatory dynamics in A17 (as outlined in Fig. 2.7d). Icons at left sketch the different grating conditions from top to bottom: 0.4 c/°, 0.2 c/° (sinusoidal), 0.2 c/° (oblique), 0.1 c/° (vertical and horizontal). Modulation amplitudes were highest for low frequency gratings. Note the systematic change in propagation direction and spatial wavelength of activity. Contour lines are drawn at zero crossings. Color bars depict activity levels, ΔF/F.

We additionally explored further systematic variations of the gratings' spatial frequency and drifting direction in this experiment (Fig. 2.10). The highest spatial frequency measured (0.4 c/°) evoked activity with principle oscillation matching the grating temporal frequency. However, the retinotopic modulations of these waves were less pronounced, which could be due to limitations in the spatial resolution of our method or caused by likely differences in physiological mechanisms (Movshon et al., 1978a,b). Wave amplitudes and retino-topic spatial resolution increased then systematically with lower spatial frequencies (Fig. 2.10, top to bottom). Importantly, a sinusoidal grating (0.2 c/°) showed comparable propagation characteristics as a square-wave grating of equal spatial frequency (Fig. 2.10, second and third row, respectively. Note that the third row represents an oblique grating). With lower spatial frequencies of the gratings (0.1 c/°, Fig. 2.10 bottom rows), additional SVD components with power at 12.5 Hz, representing each grating edge, became increasingly significant (Fig. 2.10, last row).

2.5 Discussion

We have demonstrated how the primary visual cortex – as the cortical bottleneck along the visual pathway – disambiguates two parameters, visual space and orientation: Besides representation of stationary orientation maps we show additional response components that simultaneously signal retinotopic grating motion through small oscillatory modulations of synaptic activity. Thus, our study demonstrates for the first time directly, how retinotopic millisecond propagation of grating motion is represented across stationary feature maps. We found that only a small portion of the entire signal amplitude represents such physical motion. Importantly, these retinotopic waves were homogenously traversing the orientation maps indicating independent and parallel multiplexing of the retinotopic signals.

2.5.1 Sources of the optical signal

The dye signals in cortical responses to visual stimulation are commonly of amplitudes in the range of maximal 2‰ relative to baseline fluorescence levels (Grinvald & Hildesheim, 2004; Jancke et al., 2004; Sharon & Grinvald, 2002; Sharon et al., 2007). The oscillatory signal reported here was even two orders of magnitude lower and only detectable in our highest quality recordings as the functional signal is most sensitive to various sources of noise (Grinvald & Hildesheim, 2004; Reynaud et al., 2011). Besides noise due to heart-beat pulsations and breathing artifacts (Grinvald & Hildesheim, 2004), which were largely eliminated by our data recording and processing procedures, shot-noise induced by statistical fluctuations of the light emitting process is a major technical obstacle to detect small activity changes. Therefore, the intensity of staining, i.e. baseline fluorescence levels, should ideally be as high as possible (Grinvald et al., 1988; Ross et al., 1977). However, such requested high levels of staining are not achievable in all preparations and consequently, recordings of the oscillatory components were not revealed in every experiment. Thus, the limiting factor here is most likely of technical origin rather than reflecting a genuine physiological variability across animals.

Extra-cellular recordings performed in parallel to the optical measurements confirmed that high amplitudes of the dye signal corresponded to spiking cortical activity. Yet, we cannot estimate the relative contribution of various cell types as the optical signal captures activity of a large pool of neurons. Most likely the oscillatory signal is dominated by complex cells, since simple cell activity is phase dependent and may therefore be cancelled out at the population level where phases are arbitrarily mixed (Benucci et al., 2007). Moreover, VSD imaging emphasizes activity in upper cortical layers 2/3 (Petersen et al., 2003) and therefore

highlights intra-layer propagation of activity via long-range horizontal connections, including subthreshold postsynaptic cortical activation (Jancke et al., 2004).

The amplitudes of the retinotopic waves detected here were most prominent when the grating's spatial frequency was at the lower end of the known bandpass characteristics of neurons in the respective visual area (Movshon et al., 1978b). This suggests that the relative width of the grating period plays a critical role for the emergence of the retinotopic waves (Maffei & Fiorentini, 1973; Movshon et al., 1978a). In view of the high power at the fundamental, compared to double temporal frequency, these retinotopic waves may result from an asymmetry in the responses to dark and bright (Jin et al., 2008). Indeed, a recent study in macaque primary visual cortex (V1) showed that neuronal responses in layers 2/3 (the layers our method is most sensitive to) was higher to “black” as compared to “white” stimuli (Yeh et al., 2009). Consequently, a traveling wave may emerge from units that are periodically driven by the gratings' dark edges. However, oscillatory signals were also evident for sinusoidal gratings devoid of sharp edges, for higher spatial frequencies (0.4 cycles/°) close to the preferred range of area 17 (Pollen et al., 1978), and uniformly encompass regions of preferred and opposite orientation selectivity. Altogether these observations indicate that the periodic drive constitutes the main but not necessarily the full explanation of the current findings. Also top-down influences may be involved (Ahmed et al., 2008) thus, further studies are needed to identify the various possible mechanisms contributing to retinotopic grating propagation.

2.5.2 Space–time inseparable dynamics revealed by SVD analysis

In general, SVD and other directly related procedures such as principal component analysis (PCA or Karhunen–Loève transform) are analytical matrix decomposition methods. SVD and related statistical procedures are thus helpful mathematical tools to separate spatial patterns given by eigenvalues of the autocorrelation matrix from temporal modes in complex physical systems (Aubry et al., 1991; Sirovich, 1987). Such procedures were also successfully applied in waveform analysis of evoked neuronal potentials (Glaser & Ruchkin, 1976) and spike response patterns (Abeles & Goldstein, 1977). In a first attempt to model optical imaging data by Karhunen–Loéve decomposition, different modes could be identified that resembled the layout of intrinsically recorded maps of ocular dominance, orientation, and direction in primary visual cortex (Sirovich et al., 1996). PCA is based on the correlations that exist in the multidimensional data and describes it in a new coordinate system where the basis vectors are linearly decorrelated and ranked in descending order according to the explained variance. SVD (or temporal PCA) extends the PCA and explains the data as weighted sums of space–time separable matrices, where the temporal and spatial modes correspond to principal components. The SVD algorithm is therefore mathematically equivalent to applying PCA in a specific form to spatial and temporal dimensions (Sornborger et al., 2003; Strang, 2007; see Material and methods, Singular value decomposition analysis section). In the simplest case, where the responses are space–time separable, the dynamics can be described as the outer product of temporal and spatial modes. In the presence of strong space–time correlations the outer product fails to account for the full variance. However, the residuals can be subject to the same analysis, successively explaining the remaining dynamics. SVD implicitly iterates this process and describes the data with a finite sum of products of temporal and spatial modes. Specifically, the algorithm refers only to the covariance matrix and not to higher order moments.

Another decomposition method, independent component analysis (ICA) takes a completely different approach. It describes the dataset in a coordinate system where statistical dependencies between different dimensions are minimal. For example, it has been proven

useful in cleaning EEG recordings of biological artifacts like heart-beat and eye movements (Jung et al., 2000). An important aspect of ICA is that the number of sources that are responsible for generating the observed data has to be fixed and the precise choice affects the final solution. In the present case the number of generators is not a priori known. Therefore, SVD in the present context seems to be better suited to analyze our data. Earlier SVD analysis of VSD recordings in the turtle visual brain led to discovery of rich dynamics of widespread traveling activity in different low frequency bands b 5, 10, and 20 Hz (Prechtl et al., 1997). Such large-scale timing differences of complex activity patterns across the brain were speculated to reflect processing of diverse visual features by multiple phases of neural activity (Prechtl et al., 1997; Senseman & Robbins, 1999).

In this cortical study, the stimulus was a simple constantly drifting grating. Its orientation was encoded by tonic and highest response amplitudes across cortical neurons located in specific orientation domains. In addition, the gratings' actual motion was represented by stimulus locked propagation of activity for which the relative timing reflected local movement of the gratings' stripes across the retinotopic map independent of orientation domains (Bosking et al., 2002; Buzas et al., 2003; Yu et al., 2005). This observation has an important implication for orientation selectivity: The fact that the propagating waves do not omit orthogonal orientation domains suggests that orientation tuning is influenced by non-orientation specific thalamic or intracortical inputs and consequently, affected through amplitude modulations representing visual space. As a result, non-specific parts of the orientation tuning curve may indeed contain residual retinotopic related activity in response to non-preferred orientations.

2.5.3 Multiplexing of visual information at different spatio-temporal scales

Along the visual pathway multiplexing of oscillatory signals is a general way to increase the amount of information being transmitted (Koepsell et al., 2010). Multiplexed channels acting across various timescales were recently also observed at the subcortical level in thalamic spike trains (Koepsell et al., 2009). The authors found that the firing rate at low frequencies encoded local changes of the stimulus. The second channel, in the gamma frequency range, was coupled to retinal oscillations. It was concluded that these higher frequencies possibly transmit global visual features. However, which frequency band encodes which parameter might critically depend on the stimulus characteristics and on the visual task (Ahissar & Arieli, 2001). For instance, fixational eye movements, which are naturally accompanied by incessant jitter, improve discrimination of local fine spatial details within a visual scene by emphasizing the high spatial frequencies of the stimulus (Henning & Wörgötter, 2007). Importantly, during the course of fixational switches, the change between microscopic tremor and macroscopic eye movements enables a time-varying multiplexing of visual information at different spatial scales (Rucci, 2008). Accordingly, for oscillatory stimulus-locked activity substantially lower than the gamma range, as observed in our study, the system appears flexible as to which channels carry either global or local feature information (Jancke, 2000).

2.5.4 Divergent readout of available information in primary visual cortex

In our study we consider the gratings' retinotopic motion as the source of local signals and the gratings' orientation as the prominent global stimulus feature. Such multiple signals may enable the network to differently affect downstream areas: The fast cortical retinotopic propagation might dominate readout in motion sensitive areas that analyze motion speed and solve the ‘aperture problem’ by integrating motion direction across larger receptive field sizes (Glaser & Barch, 1999; Movshon et al., 1985; Rust et al., 2006; Wallach, 1935; Zhang and Britten, 2006). In parallel, the presence of a steady orientation map signals slowly changing stimulus features (Wiskott & Sejnowski, 2002; Wyss et al., 2006) to enable invariant

representation at higher visual processing stages engaged in form representation (Karklin & Lewicki, 2009). Finally, such complex spatio-temporal activity dynamics could play an important role in cortical memory consolidation particularly, when dealing with high-dimensional natural input (Yao et al., 2007).

2.6 References

Abeles M (1991) Corticonics. Cambridge Univ. Press, Cambridge.

Ahissar E, Arieli A (2001) Figuring space by time. Neuron 32: 185–201.

Ahmed B, Hanazawa A, Undeman C, Eriksson D, Valentiniene S, Roland PE (2008) Cortical dynamics subserving visual apparent motion. Cereb Cortex 18: 2796–2810.

Aubry N, Guyonnet R, Lima R (1991) Spatiotemporal analysis of complex signals: theory and applications. J Stat Phys 64: 683–739.

Abeles M, Goldstein MH (1977) Multispike train analysis. Proc IEEE 65: 762–773.

Basole A, White LE, Fitzpatrick D (2003) Mapping multiple features in the population response of visual cortex. Nature 423: 986–990.

Benucci A, Frazor RA, Carandini M (2007) Standing waves and traveling waves distinguish two circuits in visual cortex. Neuron 55: 103–117.

Blasdel GG, Salama G (1986) Voltage-sensitive dyes reveal a modular organization in the monkey striate cortex. Nature 321: 579–585.

Bonhoeffer T, Grinvald A (1991) Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. Nature 353: 429–431.

Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation selectivity and arrangement of horizontal connections in tree shrew striate cortex. J Neurosci 17: 2112–2127.

Bosking WH, Crowley JC, Fitzpatrick D (2002) Spatial coding of position and orientation in primary visual cortex. Nat Neurosci 5: 874–882.

Buzas P, Volgushev M, Eysel UT, Kisvarday ZF (2003) Independence of visuotopic representation and orientation map in the visual cortex of the cat. Eur J Neurosci 18: 957–968.

Das A, Gilbert CD (1997) Distortions of visuotopic map match orientation singularities in primary visual cortex. Nature 387: 594–598.

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN, (1994) fMRI of human visual cortex. Nature 369: 525.

Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex 7: 181–192.

Glaser EM, Ruchkin DS (1976) Evoked potentials: principal components and varimax analysis. In: Glaser, E.M., Ruchkin, D.S. (Eds.), Principles of Neurobiological Signal Analysis. Academic, New York, pp. 233–290.

Glaser DA, Barch D (1999) Motion detection and characterization by an excitable membrane: the “bow wave” model. Neurocomputing 26–27: 137–146.

Grinvald A, Frostig RD, Lieke EE, Hildesheim R (1988) Optical imaging of neuronal activity. Physiol Rev 68: 1285–1366.

Grinvald A, Lieke EE, Frostig RD, Hildesheim R (1994) Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. J Neurosci 14: 2545–2568.

Grinvald A, Hildesheim R (2004) VSDI: a new era in functional imaging of cortical dynamics. Nat Rev Neurosci 5: 874–885.

Gray CM, König P, Engel A.K, Singer W (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature 338: 334–337.

Henning MH, Wörgötter F (2007) Effects of fixational eye movements on retinal ganglion cell responses: a modelling study. Front Comput Neurosci 1: 2.
Hubel DH, Wiesel TN (1959) Receptive fields of single neurons in the cat's striate cortex. J Physiol 148: 574–591.
Hubel DH, Wiesel TN (1974) Sequence regularity and geometry of orientation columns in the monkey striate cortex. J Comp Neurol 158: 267–293.
Hübener M, Shoham D, Grinvald A, Bonhoeffer T (1997) Spatial relationships among three columnar systems in cat area 17. J Neurosc 17: 9270–9284.
Issa NP, Rosenberg A, Husson TR (2008) Models and measurements of functional maps in V1. J Neurophys 99: 2745–2754.
Jancke D (2000) Orientation formed by a spot's trajectory: a two-dimensional population approach in primary visual cortex. J Neurosc 20: RC86.
Jancke D, Chavane F, Naaman S, Grinvald A (2004) Imaging cortical correlates of illusion in early visual cortex. Nature 428: 423–426.
Jin JZ, Weng C, Yeh CI, Gordon JA, Ruthazer ES, Stryker MP, Swadlow HA, Alonso JM, (2008) On and off domains of geniculate afferents in cat primary visual cortex. Nat. Neurosci. 11: 88–94.
Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2000) Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin Neurophysiol 111: 1745–1758.
Karklin Y, Lewicki MS (2009) Emergence of complex cell properties by learning to generalize in natural scenes. Nature 457: 83–86.
Kayser C, Montemurro MA, Logothetis NK, Panzeri S (2009) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron 61: 597–608.
Koepsell K, Wang X, Vaingankar V, Wie Y, Wang Q, Rathbun DL, Usreym WM, Hirsch JA, Sommer FT (2009) Retinal oscillations carry visual information to cortex. Front Syst Neurosci 3: 4.
Koepsell K, Wang X, Hirsch JA, Sommer FT (2010) Exploring the function of neural oscillations in early sensory systems. Front Neurosci 4: 53–61.
Lee SH, Blake R, Heeger DJ (2005) Traveling waves of activity in primary visual cortex during binocular rivalry. Nat Neurosci 8: 22–23.
Maffei L, Fiorentini A (1973) The visual cortex as a spatial frequency analyser. Vis Res 13: 1255–1267.
Mante V, Carandini M (2003) Visual cortex: seeing motion. Curr Biol 13: R906–R908.
Movshon JA, Thompson ID, Tolhurst DJ (1978a) Receptive field organization of complex cells in the cat's striate cortex. J Physiol 283: 79–99.
Movshon JA, Thompson ID, Tolhurst DJ (1978b) Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's viusal cortex. J Physiol 283: 101–120.
Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1985) The analysis of moving visual patterns. In: Chagas, C., Gattass, R., Gross, C. (Eds.), Pattern recognition mechanisms. Pontifica Academia Scientiarum, Vatican City, pp. 117–151.
Petersen CC, Grinvald A, Sakmann B (2003) Spatiotemporal dynamics of sensory responses in layer 2/3 of rat barrel cortex measured in vivo by voltage-sensitive dye imaging combined with whole-cell voltage recordings and neuron reconstructions. J Neurosc 23: 1298–1309.
Pollen DA, Andrews BW, Feldon SE (1978) Spatial frequency selectivity of periodic complex cells in the visual cortex of the cat. Vis Res 18: 665–682.
Prechtl JC, Cohen LB, Pesaran B, Mitra PP, Kleinfeld D (1997) Visual stimuli induce waves of electrical activity in turtle cortex. Proc Natl Acad Sci USA 94: 7621–7626.

Reynaud A, Takerkart S, Masson GS, Chavane F (2011) Linear model decomposition for voltage-sensitive dye imaging signals: application in awake behaving monkey. Neuroimage 54: 1196–1210.
Richmond BJ, Optican L, Spitzer H (1990) Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. I. Stimulus–response relations. J Neurophysiol 64: 351–369.
Roelfsema PR, Lamme VA, Spekreijse H (2004) Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. Nat Neurosc 7: 982–991.
Ross WN, Salzberg BM, Cohen LB, Grinvald A, Davila HV, Waggoner AS, Wang CH (1977) Changes in absorption, fluorescence, dichroism, and birefringence in stained giant axons: optical measurement of membrane potential. J Membr Biol 33: 141–183.
Rucci M (2008) Fixational eye movements, natural image statistics, and fine spatial vision. Network 19: 253–285.
Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. Nat Neurosc 9: 1421–1431.
Senseman DM, Robbins KA (1999) Modal behavior of cortical neural networks during visual processing. J Neurosc 19: 1–7.
Sharon D, Grinvald A (2002) Dynamics and constancy in cortical spatiotemporal patterns of orientation processing. Science 295: 512–515.
Sharon D, Jancke D, Chavane F, Na'aman S, Grinvald A (2007) Cortical response field dynamics in cat visual cortex. Cereb Cortex 17: 2866–2877.
Shoham D, Hübener M, Schulze S, Grinvald A, Bonhoeffer T (1997) Spatio-temporal frequency domains and their relation to cytochrome oxidase staining in cat visual cortex. Nature 385: 529–533.
Sirovich L (1987) Turbulence and the dynamics of coherent structures, Part I. Coherent structures, Part II. Symmetries and transformations, Part III. Dynamics and scaling. Q Appl Math 45: 561–590.
Sirovich L, Everson R, Kaplan E, Knight BW, O'Brien E, Orbach D (1996) Modeling the functional organization of the visual cortex. Phys D 96: 355–366.
Sornborger A, Sailstad C, Kaplan E, Sirovich L (2003) Spatiotemporal analysis of optical imaging data. Neuroimage 18: 610–621.
Strang G (2007) Computational Science and Engineering. Wellesley-Cambridge Press.
Tusa RJ, Palmer LA, Rosenquist AC (1978) The retinotopic organization of area 17 (striate cortex) in the cat. J Comp Neurol 177: 213–236.
Wallach H (1935) Über visuell wahrgenommene Bewegungrichtung. Psychol Forsch 20: 25–380.
Wiskott L, Sejnowski T (2002) Slow feature analysis: unsupervised learning of invariances. Neural Comp 14: 715–770.
Wyss R, König P, Verschure PF (2006) A model of the ventral visual system based on temporal stability and local memory. PLoS Biol 4: e120.
Xu X, Collins CE, Khaytin I, Kaas JH, Casagrande VA (2006) Unequal representation of cardinal vs. oblique orientations in the middle temporal visual area. Proc Natl Acad Sci USA 103: 17490–17495.
Yacoub E, Harel N, Ugurbil K (2008) High-field fMRI unveils orientation columns in humans. Proc Natl Acad Sci USA 105: 10607–10612.
Yao H, Shi L, Han F, Gao H, Dan Y (2007) Rapid learning in cortical coding of visual scenes. Neuron 10: 772–778.
Yeh C-I, Xing D, Shapley RM (2009) “Black” responses dominate macaque primary visual cortex V1. J Neurosci 29: 11753–11760.

Yu H, Farley BJ, Jin DZ, Sur M (2005) The coordinated mapping of visual space and response features in visual cortex. Neuron 47: 267–280.
Zhang T, Britten KH (2006) The virtue of simplicity. Nat Neurosc 9: 1356–1357.

3. Primary visual cortex represents the difference between past and present (Study 2)

This section is based on the following publication in Cerebral Cortex: N Nortmann[123], S Rekauzke[12], S Onat[3], P König[34], D Jancke[12]* (2013, Epub ahead of print) Primary visual cortex represents the difference between past and present. Cerebral Cortex, doi: 10.1093/cercor/bht318.* *[1]Optical Imaging Group, Institut für Neuroinformatik, Ruhr-University Bochum, Bochum, Germany. [2]Bernstein Group for Computational Neuroscience, Ruhr-University Bochum, Bochum, Germany. [3]Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany. [4]Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. *These authors shared senior authorship.*

3.1 Abstract

The visual system is confronted with rapidly changing stimuli in everyday life. It is not well understood how information in such a stream of input is updated within the brain. We performed voltage-sensitive dye imaging across the primary visual cortex (V1) to capture responses to sequences of natural scene contours. We presented vertically and horizontally filtered natural images, and their superpositions, at 10 or 33 Hz. At low frequency, the encoding was found to represent not the currently presented images, but differences in orientation between consecutive images. This was in sharp contrast to more rapid sequences for which we found an ongoing representation of current input, consistent with earlier studies. Our finding that for slower image sequences, V1 does no longer report actual features but represents their relative difference in time counteracts the view that the first cortical processing stage must always transfer complete information. Instead, we show its capacities for change detection with a new emphasis on the role of automatic computation evolving in the 100-ms range, inevitably affecting information transmission further downstream.

3.2 Introduction

Characterization of stimulus-response relationship is the most fundamental approach to accessing cortical coding behavior. This procedure starts with the assumption that neuronal populations sample sensory input and form faithful internal representations of its actual content. In fact, classical reverse-correlation techniques enable us to determine neuronal tuning properties by backtracking responses to stimulus variations across rapid sequences of presentation (Eckhorn et al., 1993; Ringach et al., 1997). Thus, these techniques build on the idea that neuronal activity is permanently updated by current stimulation, maintaining an ongoing representation of the outer world (Jonides et al., 1982).

To measure the continuous dynamics of cortical population activity, we used voltage-sensitive dye imaging, which reflects gradual changes in membrane potentials across several square millimeters of cortex with an emphasis on superficial layers (Grinvald et al., 1994; Petersen et al., 2003; Jancke et al., 2004; Chen et al., 2006; Roland et al., 2006; Berger et al., 2007; Sit et al., 2009; Grinvald & Hildesheim, 2004 for review). This method does not provide single neuron activity or its dependence on different cortical layers. On the upside, it avoids biased sampling of neurons and it captures population activity irrespective of receptive field locations and preferred feature selectivities (Lee et al., 1988; Vogels, 1990; Jancke et al., 1999; Tsodyks et al., 1999; Jancke, 2000; Dinse & Jancke, 2001; Pouget et al., 2003; Graf et al., 2011; Gilad et al., 2012; Lewis & Lazar, 2013), hence providing the global tuning of the cortex across millions of neurons under different stimulus conditions.

We report, using recordings in cat visual cortex (V1), that ongoing encoding can be found for the representation of briefly presented stimulus sequences (33 Hz) consistent with earlier studies (Ringach et al., 1997; Benucci et al., 2009). However, slower image sequences (10 Hz) reveal an essential addition. Population tuning in the primary visual cortex no longer represents the complete image content, but rather those orientations that were newly added or removed. We propose that such a precise detection of change across sequences of natural scene contours involves the interplay between two well-known neuronal behaviors, adaptation and stimulus off-responses (Movshon & Lennie, 1979; Duysens et al., 1996; Müller et al., 1999; Bair et al., 2002; Felsen et al., 2002; Dragoi et al., 2002) that are important for stimulus transitions (Eriksson et al., 2008; Eriksson et al., 2012) and interact here to encode the difference to past images. In combination with eye movements at different spatiotemporal scales, the observed frequency-dependent encoding of image content might help to remove predictable input correlations in order to emphasize object borders and discontinuities within natural scenes (Rucci et al., 2007; Desbordes & Rucci, 2007; Rucci, 2008; Kuang et al., 2012). We conclude that input timing may entail predictive encoding (Rao & Ballard, 1999; Friston, 2005) at the very first cortical processing stages without the involvement of voluntary or attentional top-down mechanisms.

3.3 Materials and methods

3.3.1 Visual stimuli and presentation

Construction of oriented stimuli (V, H) and superpositions (VH)

We presented sequences of stimuli to anesthetized cats (11 males, 4 females, adult). First, we used 128 natural images (64 urban and 64 nature scenes) in grayscale to construct stimuli with dominant vertical (V) and horizontal (H) orientation. We derived these oriented stimuli by filtering the natural images in Fourier space with real-valued, polar-separable filters. The angular function of these filters was a triangular hat function, symmetric across 180°, with its maxima of 1 at either vertical (for stimulus type V) or horizontal (for stimulus type H) frequency. The half width at half maximum was 45° for the construction of broadly filtered images, and 2.8° for the construction of narrowly filtered images. The radial function (adapted from Simoncelli & Farid, 1996) included a low-pass filter with cut-off at 6.6 c/° visual angle, and the DC component was set to zero. From the first harmonic up to 6.0 c/° (start of low-pass transition range), the function was 1. Thus, in this range, we preserve the relative amplitudes of the original image, including the characteristic 1/f fall-off for natural stimuli.

Next, we constructed superpositions (VH) of vertical (V) and horizontal (H) stimuli by summing them (in image space). We normalized the stimuli such that the global contrast of all superpositions was the same (rms contrast: 0.71); this means that component stimuli (V and H) were scaled with the same factor as their superposition (VH). During this normalization, we had to clip some pixels of images with a large intensity range (in average 2%, at most 15%).

We also presented vertical (V) and horizontal (H) square-wave gratings (0.2 c/°) and their linear superpositions (VH, contrast also 0.71, phases were varied over repetitions). In addition to oriented stimuli and their superpositions we used an isoluminant screen as blank (B) stimulus. All stimuli were gamma-corrected according to the presentation monitor (100 Hz, Sony Triniton GDM-FW900, Japan). Mean luminance of each stimulus, including blank, was 20 cd/m². Stimuli covered a visual field of 31°x31°.

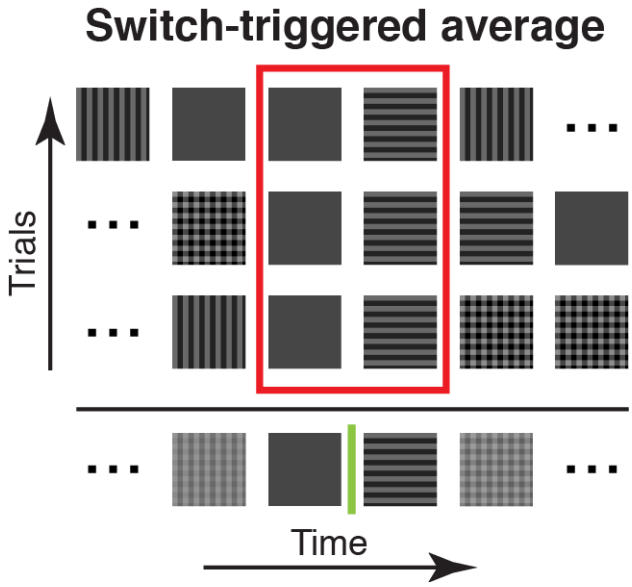


Figure 3.1: Illustration of the switch-triggered average. Three pseudorandom stimulus sequences are exemplified (5 stimuli are shown; in the experiments we used at least 64 different sequences of at least 17 stimuli, see Materials and Methods). Sequences were aligned to a specific switch between a pair of stimuli (blank to horizontal in this example) to illustrate our averaging procedure. Note that when aligning sequences to a particular switch type for averaging, the stimuli before and after were different. For switch-triggered averaging the procedure was used to average the cortical responses to a particular stimulus pair across all sequences shown.

Stimulus sequences

From the four stimulus types: (V) vertical stimulus, (H) horizontal stimulus, (VH) superposition, and (B) blank stimulus, we created pseudorandom sequences including all 16 (4^2) possible transitions, that is, switches, between them, resulting in 17 stimuli per sequence. We constructed 64 different sequences optimizing the following criteria: As response variance across different experimental trials is relatively high, we applied the constraints that all stimulus types (V, H, VH, and B) should occur equally often (4 times within each 10-Hz sequence) and that every switch should occur equally often (once within each 10-Hz sequence). To avoid systematic effects in the responses to particular switch types in dependence of their position in the sequence (start, middle, or end) we randomized positions of both stimulus type and transition type across the 64 sequences (see Fig. 3.1 for a sketch of the paradigm and averaging procedure).

These 64 sequences were repetitively shown with two different presentation frequencies, at 10 or at 33 Hz. For 10-Hz presentation (100-ms presentation of each stimulus), one sequence was shown per trial. Here, stimulus presentation per trial lasted 1700 ms. For 33-Hz presentation (30-ms presentation of each stimulus), we showed three sequences in one trial (1530 ms per trial). Optical data were recorded for 2 s per trial, including a 200 ms baseline, in both cases. The relatively short trial durations were used to avoid dye bleaching effects, photodynamic damage of the cortical tissue, and possible contamination by intrinsic signals (Grinvald & Hildesheim, 2004). The intertrial interval, in which a blank stimulus was shown, was set to a minimum of 5 s. When two stimulus conditions were used (such as narrowly filtered and broadly filtered), these were randomized across trials. In between stimulus trials, we recorded blank conditions, where an isoluminant gray screen was shown for 2 s, to allow correction of breathing and heartbeat artifacts (2 blanks for 16 stimulus trials).

Experimental protocols

In the main body of imaging experiments we used 10-Hz presentation frequency and the two stimulus conditions narrowly filtered and broadly filtered images (Fig. 3.4 and 3.7, 12 experiments, 256–896 repetitions of each switch type). In one additional imaging experiment, we used oriented square-wave gratings for 10-Hz presentation (Figure 3.4 (bottom row) and 5 (imaging trace), 236 repetitions of each switch type). To recapitulate previous optical imaging results (Benucci et al., 2009), in one experiment (Fig. 3.2, 3.3, and 3.4 (upper row), 384

repetitions of each switch type), we used 33-Hz presentation frequency and two different stimulus conditions: gratings and narrowly filtered images. Finally, instead of using long stimulus sequences we presented two stimuli (i.e. isolated single switches) in one experiment. Here, we concentrated on the most informative switch type, superposition to a single orientation (VH to V, and VH to H). Gratings were used in three different timing conditions: The superposition (VH) was shown for 30, 100, and 500 ms before switches to a single orientation (V or H) occurred (Figure 3.6, 80 repetitions).

High-contrast moving gratings

For calibration and mapping of orientation preference, we used moving square-wave gratings (rms-contrast: 1, 0.2 c/°, 6 Hz, mean luminance 53 cd/m² [8 hemispheres], or 38 cd/m² [9 hemispheres]) with four different orientations (0°, 45°, 90°, 135°) and both motion directions. We recorded for 1 s in each trial, including 200-ms prestimulus time. These calibration trials were recorded throughout the entire experiment in between blocks of the main stimulus protocol of each experiment (specified above).

3.3.2 Preparations for optical imaging

All animal experiments were carried out in accordance with the European Union Community Council guidelines and approved by the German Animal Care and Use Committee (application number: AZ 9.93.2.10.32.07.032) in accordance with the Deutsches Tierschutzgesetz (§ 8 Abs. 1) and the NIH guidelines. For further details, see Onat et al. (2011a). In brief, animals were initially anesthetized with ketamine (20 mg per kg i.m.) and xylazine (1 mg per kg i.m.), artificially respirated, continuously anesthetized with 0.8–1.5% isoflurane in a 1:1 mixture of O_2/N_2O, and fed intravenously. Both weak effects on neuronal tuning properties (e.g. Niell and Stryker, 2010) and strong modulations (e.g. Adesnik et al., 2012) have been reported when comparing anesthetized and awake states. This might additionally depend on the type of anesthetics used. Therefore, it is an interesting question, and it remains to be generally tested, in how far results obtained under anesthesia hold in behavioral settings. However, our anesthetized and paralyzed preparation provides the advantage that eyes are fixed and hence allow complete control of the dynamics of the visual input. We administered 0.4 mg/kg dexamethasone i.m. and 0.05 mg/kg atropine sulfate i.m. daily and 20 mg/kg cephazolin twice a day. In few control experiments we used contact lenses with a 3 mm diameter pupil. Heart rate, intratracheal pressure, exhaled CO_2, and body temperature were monitored. The skull was opened above area V1 (A18, occasionally parts of A17), the dura was removed, a chamber was mounted, the cortex was stained for 3 h (and occasionally re-stained) with voltage-sensitive dye (RH-1691), and unbound dye was washed out.

3.3.3 Data acquisition and pre-processing

Optical imaging was conducted with Imager 3001 (Optical Imaging, Inc., Mountainside, NY, USA). The camera was focused ~500 μm below the cortical surface. Data acquisition onset was synchronized with heartbeat and respiration. For detection of changes in fluorescence, the cortex was illuminated with light of wavelength 630 ± 10 nm, and emitted light at wavelengths above 665 nm was collected. The frame rate was set to 100 Hz. We performed normalization by dividing each pixel value by its average 200-ms prestimulus activity; heartbeat and respiration-related artifacts were removed by subtracting the average blank signal. These preprocessing steps lead to a unitless relative signal of fluorescence, denoted by delta F/F. For the main paradigm (10 Hz narrowly/broadly filtered), we excluded five

hemispheres from analysis because of insufficient staining. Data were used when the amplitude of the evoked response in the Fourier power spectrum at 10 Hz (the switch-type unspecific response) was at least three times larger than at surrounding frequencies (± 2 Hz).

3.3.4 Electrophysiology

Electrode recordings served as a control of the voltage-related responses reported by the fluorescent optical signals. The recorded units were collected at a depth between 400 and 700 µm. Before electrophysiology, a vascular map of the brain was captured by illumination with green light (546 nm) from two optic fiber light guides. Additionally, the afterward measured orientation maps were overlaid. This combined map was then used to guide electrode penetrations to orientation-selective domains. Spikes were sorted online by a multiple spike detector, MSD (Alpha Omega Engineering, Ltd., Israel). Cells were selected upon differences in spike-wave forms. Multiple unit activity (MUA, mostly 3-4 cells, occasionally we recorded single units in addition to MUA (5 of 27 sites)) was recorded with tungsten electrodes (0.8-2 MΩ, WPI, Inc., USA). For display purposes (Fig. 3.5A, D), the data were convoluted in a 5 ms window.

For these electrophysiological recordings, we presented oriented square-wave gratings at 10 Hz. We adopted our corresponding imaging paradigm (see above) to use longer trial durations. This allowed us to show sequences of 66 stimuli (6600 ms presentation duration per trial) consisting of the 64 possible triplets of the four stimulus types (V, H, VH, and B). Positions of triplet types within sequences were randomized across trials. Data were recorded in four hemispheres (in two of which we also performed optical imaging).

3.3.5 Analysis

Orientation maps

Maps were computed using data from high-contrast, square-wave moving gratings. After trial-wise pre-processing, we averaged data over repetitions (50-119 repetitions), motion direction, and time. The resulting dataset was spatially band-pass filtered from 1 to 3 c/mm. The vertical-horizontal (VH-) orientation maps (used for correlation analysis in Fig. 3.6E and Suppl. Fig. 3.S1) were obtained by subtracting the horizontal from the vertical map. We also computed additive VH-maps as a control (Fig. 3.6E). Orientation maps, which cover the full range of orientations, were computed based on the four measured orientations using vector summation, and downsampled to 18 bins of 10° each.

Switch-triggered responses

To obtain responses to each of the 16 possible switches (i.e. the pairs of 4 different stimulus types: V, H, VH, and B (blank)), we first removed responses unspecific to switch type by pixel-wise subtraction of the average response across the entire stimulus sequences. We then computed switch-triggered responses for each of the 16 switch types. This was done by aligning responses to a particular switch in time and then averaging over repetitions measured in different sequences. Thereby we averaged over responses to varying stimuli before and after a specific switch. The procedure to compute switch-triggered averages is illustrated in Fig. 3.1 (a comparison to stimulus-triggered averaging is provided in Suppl. Fig. 3.S2). When using filtered natural stimuli the two image categories (urban/nature scenes) were pooled.

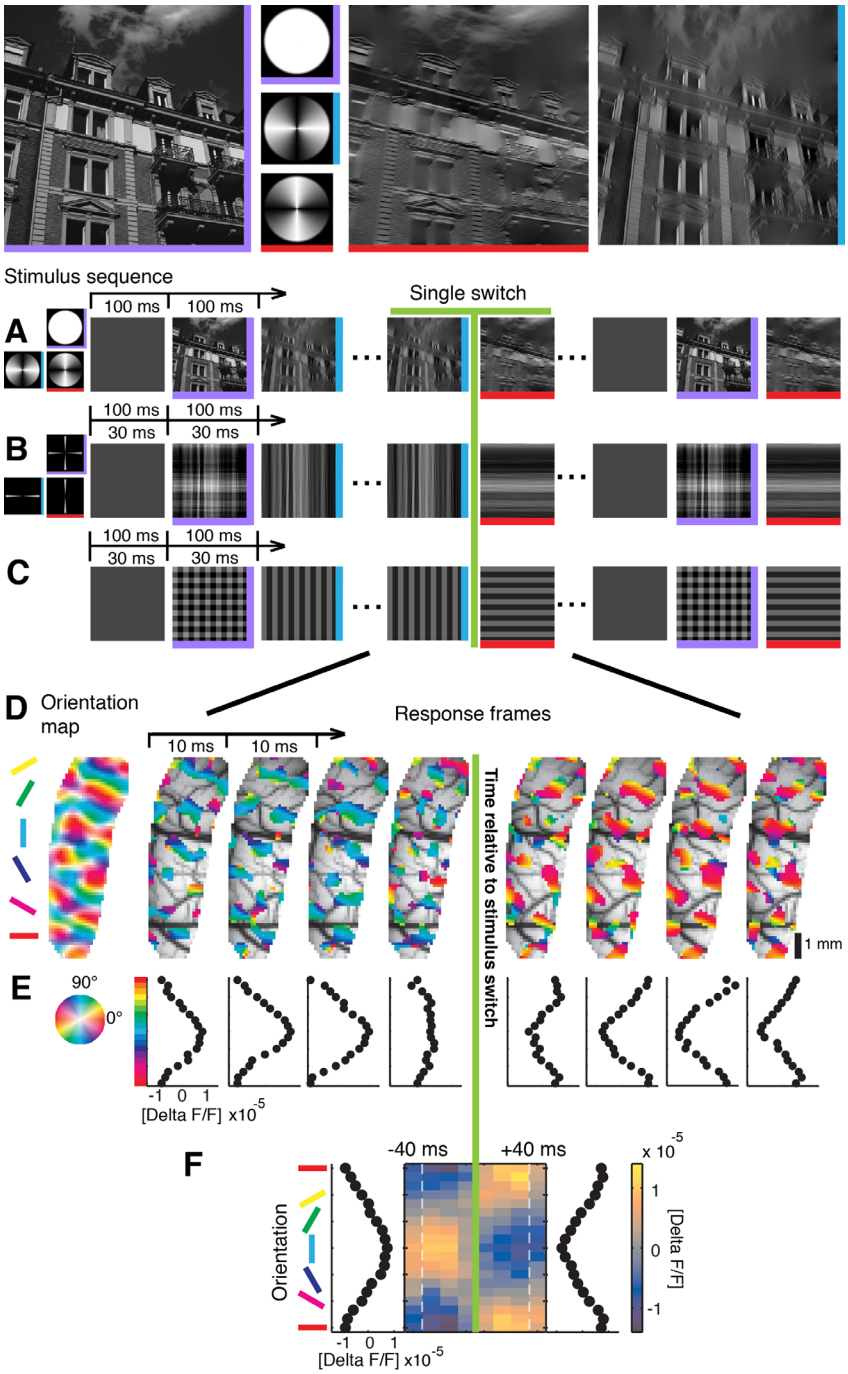


Figure 3.2: Visual stimulation and cortical representation of oriented contours. Top: Example stimulus, a natural image (left, purple), and its horizontally (red) and vertically (blue) filtered versions using broad filters in Fourier space (second column). Here and in subsequent figures colored edges of stimulus icons indicate stimulus type (horizontal stimulus (H): red horizontal edge; vertical stimulus (V): blue vertical edge; superposition (VH): purple horizontal and vertical edges). (**A-C**) Time sequences of stimuli with different orientation content, presented either at 30- or 100-ms stimulus periods (see outlined arrows): broad filters (**A**), narrow filters (**B**), and square-wave grating (**C**). (**D**) Switch from a vertical to a horizontal square-wave grating (30-ms stimulus period): the most active 30% of pixels in each 10-ms camera frame are shown and overlaid on the vascular cortical image (averages over 384 switch repetitions). Here and in subsequent plots, data were shifted by 50 ms to account for average latency; the green line at zero marks relative time to switch. Pixel color denotes preferred orientation derived from a conventional orientation map (most left), which was recorded separately (see Materials and Methods). (**E**) Population tuning curves for each time frame obtained by averaging activities over pixels with the same orientation preference (preferences were binned into 18 classes of 10°). Horizontal orientation bin (0°, red) is displayed twice. Abscissa shows amplitudes of tuned activity (delta F/F, see Materials and Methods). (**F**) Switch-triggered average as compact illustration of the data in **E**. Colorbar shows amplitudes of tuned activity. Orientation tuning curves to the left and right of the central frame illustrate time averages (−40 to 0, and 0 to +40 ms). Dashed faint lines illustrate time shifted (50 ms) positions of transitions from and to other stimuli (before and after the switch), which were averaged across all sequences by switch-triggered averaging (see Fig. 3.1 and Materials and Methods).

Population tuning

The data were spatially band-passed from 1 to 3 c/mm. Population tuning was computed by averaging responses across pixels with the same orientation preference (as determined by the vector-based orientation map). Overall population tuning was then obtained by averaging across experiments (10 Hz narrowly/broadly filtered n = 12, 10 Hz gratings n = 1, 33 Hz narrowly filtered/gratings n = 1) and time (50–90 ms after switch).

Modulation depth and tuning width

Modulation depth was calculated as the average difference between activity at pixels preferring 90±15° and 0±15° (time window 50-90 ms after switch). We tested for differences between broadly and narrowly filtered stimuli, using responses from blank (B) to vertical (V) and horizontal (H) orientations. We first averaged over vertical and horizontal condition, sign-inverting the latter, and subsequently we used a pairwise two-tailed t-test to compare filter conditions within experiments (n = 12). Tuning width was estimated by fitting a Gaussian with three parameters (amplitude, width, and basis level) for each experiment and filter condition, and using a paired two-tailed t-test comparing filter conditions within experiments. We only used experiments for this subsequent test, in which the fitted Gaussian yielded a fit better than 80% of variance explained (the residual variance divided by the variance of the data, subtracted from one), which was the case in 8 of 12 experiments. In general, before conducting t-tests we tested the respective distributions for normality using a Lilliefors test (*alpha* = 0.05).

Statistical evaluation of population tuning

To evaluate population tuning statistically we used a three-step procedure. First, we quantified how much the time-averaged orientation tuning curves deviated from a flat orientation-tuning curve (zero-baseline). To quantify this difference to zero, we computed one $\boldsymbol{X}^2$-test statistic. This allowed us to include the data from all experiments in an overall quantification, while at the same time accounting for differences in sample sizes, variances, and response levels across different experiments. The $\boldsymbol{X}^2$-test statistic is computed across all experiments (experiment e = 1,…,n; where n is the number of experiments), and orientation bins (bin θ = 1,…,18; bins of 10° each), taking the standard error over repetitions of a switch type (repetition r_e = 1,…,m_e; where r_e is a repetition in experiment e, and m_e is the overall number of repetitions in experiment e) in each experiment as a normalization factor:

$$\mathrm{X}^2 = \sum_e \sum_\theta \left(\frac{s_{e,\theta} - \langle d_{e,\theta,r_e} \rangle_{r_e}}{\sigma / \sqrt{m_e}} \right)^2$$

$$\sigma = \sqrt{\frac{1}{m_e - 1} \sum_{r_e} \left(d_{e,\theta,r_e} - \langle d_{e,\theta,r_e} \rangle_{r_e} \right)^2}$$

Here, $\langle \ \rangle_{r_e}$ denotes the average over repetitions in experiment e and d_{e,θ,r_e} the response in experiment e, orientation bin θ, and repetition r_e. In this first step, we wanted to quantify the difference between our data and zero. Thus we used $s_{e,\theta} = 0$ as constant baseline. The degrees of freedom are $df_1 = 18n - 1$, corresponding to 18 orientation bins and n experiments.

In the second step of the evaluation of population tuning, we fitted sinusoids to the population tuning curves and repeated the quantification on the residuals. We used the sinusoid function

$s_{e,\theta} \coloneqq a_e \, cos(10°\theta(2\pi/18))$ with amplitude a_e, which is fitted for each experiment, e, by minimizing ε_e (for each experiment e).

$$\varepsilon_e \coloneqq \sqrt{\sum_{\theta} \sum_{r_e} \left(s_{e,\theta} - d_{e,\theta,r_e}\right)^2}$$

Goodness of fit was then evaluated by computing the $\boldsymbol{X}^2$-test statistic on the residuals as specified above (using the fitted sinusoids in place of the constant baseline).
In this second step, there are fewer degrees of freedom, $df_2 = 18n-1-n$, because we did fit n parameters (10 Hz narrowly/broadly filtered n = 12, 10 Hz gratings n = 1, 33 Hz narrowly filtered/gratings n = 1).

In the third step of the procedure, we investigated the direction of population tuning. It is characterized by the sign of the amplitude of the sinusoid, a_e, which was fitted in the second step. When a_e is positive, the maximum of the sinusoid is at bin 18, which covers orientation preferences 175°-185°, and its minimum is at bin 9, which covers 85°-95°. Thus, positive amplitude a_e indicates horizontal population tuning. Negative amplitude, correspondingly, indicates vertical population tuning.

3.4 Results

To mimic natural viewing conditions (Betsch et al., 2004), stimuli were rapid sequences comprising blank (B), vertical (V), horizontal (H), and superimposed orientations (VH) with different degrees of complexity (Fig. 3.2A–C). We set up pseudorandom sequences from these 4 stimulus types such that each sequence included all 16 possible transitions. Activity was continuously recorded in 10-ms time frames (Fig. 3.2D), and switch-triggered averages were generated for each individual transition.

3.4.1 Representation of current orientations

First we verified that we can reproduce former findings within our settings. Figure 3.2D-F depicts cortical responses to a switch (green line) from vertical to horizontal when square-wave gratings (Fig. 3.2C) were presented within short 30-ms stimulus sequences. Note the rapid change in color (bluish to reddish) that indicates subsequent activation of spatially distinct populations of neurons representing each stimulus orientation briefly before and after the switch. To obtain a compact depiction of overall population tuning over time, we remapped the data onto orientation space (Fig. 3.2E). Figure 3.2F summarizes, in 10-ms time frames, how peak population activity shifted from vertical to horizontal, with a transition phase of typically one to two 10-ms frames (color bar for activity levels, tuning profiles on left and right). Thus, as described in previous electrophysiological studies (Ringach et al., 1997; Gillespie et al., 2001) as well as in a recent work which evaluated both voltage-sensitive dye imaging and extracellular measurements for this paradigm (Benucci et al., 2009), we showed how the primary visual cortex acts as a straightforward "instantaneous decoder" (Benucci et al., 2009) by mapping of currently presented orientations.

Next, we tested whether this coding scheme also holds for complex stimuli that contain more than just a change to a single orientation. To answer this question, we introduced switches from a single grating orientation (V or H) to superposition (VH plaid) and vice versa (Eriksson et al., 2010; Nortmann et al., 2011). Superimposing an orthogonal orientation to the present one resulted in relatively flat distributions, representing the average of the individual orientation patterns (Fig. 3.3, columns #1, stimulus conditions on top; see figure legend),

indicating an unbiased processing of both the sustained and added orientation (MacEvoy et al., 2009; Busse et al., 2009). Importantly, a change back from superposition to a single orientation led to responses tuned to the orientation present after the switch (Fig. 3.3, columns #2, see blue and red arrows). Therefore, current orientations were again directly encoded with a processing delay of ~50 ms, similar to when turned on from blank (Fig. 3.3, compare columns #3). This scheme deviated exclusively for switches from an oriented stimulus back to blank (Fig. 3.3, columns #4). In this case, activity showed persistent tuning after the stimulus was turned off (Coltheart, 1980; Duysens et al., 1985). In summary, our data using 33-Hz square-wave grating sequences confirmed the ongoing encoding of orientation, including the known exception of the tuned response after stimulus offset (Benucci et al., 2009).

Response characteristics obtained with simple artificial stimuli, like gratings of optimal spatial frequency used so far, do not necessarily generalize towards input of ecological relevance (Smyth et al., 2003; David et al., 2004; David & Gallant, 2005; Felsen et al., 2005; Haider et al., 2010; Onat et al., 2011b; Fournier et al., 2011). That is, response behavior to natural input can deviate significantly from predictions based on simple parameterized stimuli, probably due to the extensive spatial context in natural images (see Carandini et al. [2005] and Olshausen & Field [2005] for reviews). Thus, it may be important to validate findings obtained with artificial stimuli using more natural stimulus conditions (Felsen & Dan, 2005). To address this point in a first step, we extracted oriented contours from 128 different natural images (see Materials and Methods) by filtering them along the orientation dimension in Fourier space. In contrast to gratings, these images retain important properties of natural stimuli, such as the phase relationships and the typical 1/f fall-off of amplitudes along the spatial dimension (Simoncelli & Olshausen, 2001; Geisler, 2008; Hyvärinen et al., 2009; see Fig. 3.2B for example). Indeed, population tuning had an overall lower amplitude than for gratings (~30%, cf. scale bars in Figs. 3.3 and 3.4), most likely due to the heterogeneity of local contrasts in the natural stimuli, which may engage widespread population gain control reflected in the dye signal (Sit et al., 2009). Nevertheless, also for these stimuli, actual orientations after the switch were well represented (compare Fig. 3.4 first row with Fig. 3.3, columns #1 and #2).

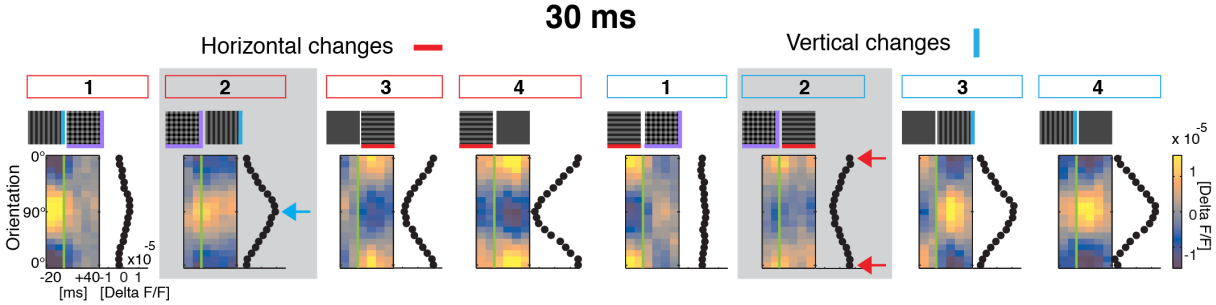


Figure 3.3: Ongoing cortical encoding of current orientation with sequences of short 30-ms stimulus periods. Frames resolve population tuning around stimulus switch-time (green line) in 10-ms steps (−20 to +40 ms, averages over 384 repetitions of each switch, same conventions as in Fig. 3.2F). The icons on top depict switch conditions. Black traces to the right of each frame show time average: 0 to +40 ms. The first four conditions on the left are characterized by a change in horizontal orientation (either turned on or off). Each stimulus has different orientation content after the switch, except for the switch to blank (4th frame). The last four conditions cover corresponding changes in vertical orientation. Note that responses before the switch in 33-Hz sequences can partly include responses to previous stimuli (first dashed line Fig. 3.2F), particularly visible in conditions #3, when the switch occurred from blank that naturally leads to a low amplitude of activity. Tuning curves obtained with voltage-sensitive dye imaging (Sharon & Grinvald, 2002) are generally broader than for spike recordings (Benucci et al., 2009) and hence, superimposed gratings caused per se relatively flat distributions. Also our stimuli were not of highest contrast and evoked lower modulation depth, which in turn produces higher sensitivity to noise (Grabska-Barwinska et al., 2009). Thus, responses to superimposed gratings did not reveal a clear bimodal distribution when averaged over brief time intervals (but see Fig. 3.6 for longer time averages). All switches $p<0.001$ in comparison to flat response; except H to VH: $p = 0.97$; after fitting for all residuals $p>0.99$; except VH to H: res. $p = 0.87$; B to V: res. $p = 0.98$; H to B: res. $p = 0.73$; V to B: res. $p = 0.26$ (see Materials and Methods and Suppl. Table 3.S1–S3 for details).

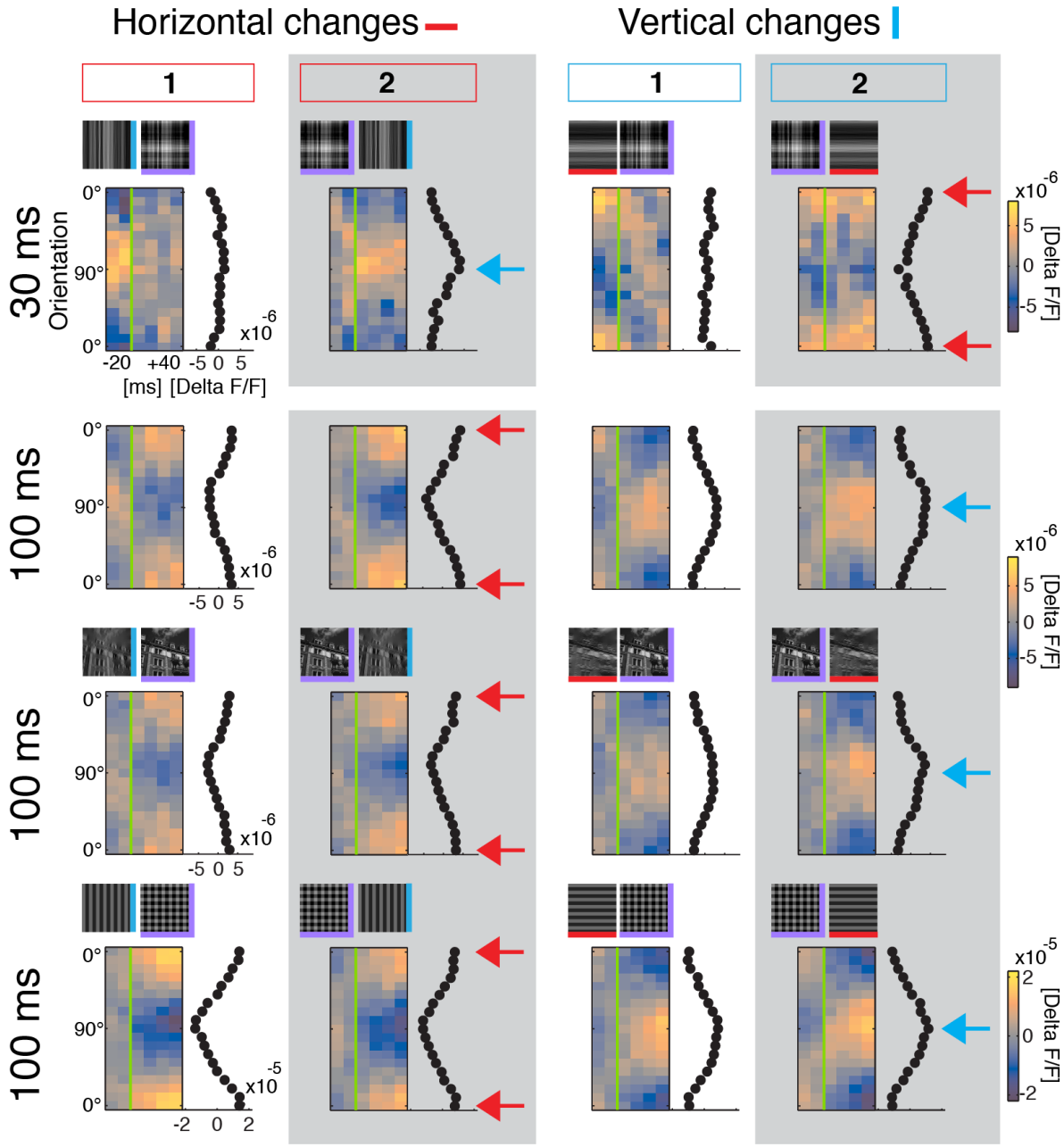


Figure 3.4: Primary visual cortex represents difference in orientation content compared with past. Conventions and layout as in Fig. 3.3 columns #1 and #2. First row is for comparison with short 30-ms stimulus period: Responses to narrow filtered natural stimuli, 1 experiment (384 switch repetitions, V/H to VH: $p>0.99$ no tuning; VH to V: $p<0.01$ tuned [residual $p = 0.99$]; VH to H: $p<0.02$ tuned [res. $p = 0.88$]). Bottom rows: 100-ms stimulus period, narrowly filtered natural stimuli (second row; average modulation depth: 0.6±0.2 $x10^{-5}$), broadly filtered (third row; average modulation depth: 0.5±0.1 $x10^{-5}$), and gratings (last row; average modulation depth: 2.1 $x10^{-5}$). Blue arrows highlight positions of vertical tuning (90°), and red arrows indicate horizontal position (0°). Note that tuning peaked around removed (or added) orientations instead of currently presented orientations (for narrowly and broadly filtered stimuli all displayed switches: $p<0.001$ tuned [residuals: $p>0.99$], 12 experiments [256-896 switch repetitions each], average is shown; for gratings all displayed switches: $p<0.001$ tuned [res.: V to VH: $p = 0.96$, VH to V: $p = 0.69$, H to VH: $p = 0.65$, VH to H: $p>0.99$], 1 experiment, 236 switch repetitions).

3.4.2 Representation of the difference between past and present orientations

During natural vision changes of contour orientation can occur on relatively slow time scales (Gallant et al., 1998; Dragoi et al., 2002; Betsch et al., 2004; Kayser et al., 2004). Hence, in the following we contrast the above scheme of ongoing encoding with the processing characteristics found for slower sequences of natural scene contours, using 100-ms stimulus periods (Fig. 3.4, bottom three rows). Strikingly here, responses to a switch from the vertical orientation to superimposed horizontal and vertical (Fig. 3.4, left column) represented almost exclusively the horizontal orientation, hence the orientation that was added rather than the present superposition. Likewise, switches from superposition to vertical (Fig. 3.4, second column) were followed by responses that were tuned to horizontal, thus representing the removed orientation instead of the remaining vertical orientation (compare first-row blue with bottom red arrows). The same characteristics were found for changes in vertical orientations (Fig. 3.4, two right columns; see Suppl. Fig. 3.S1 for the correlation between population tuning over time and a standard VH-orientation map).

Most remarkably, the mechanism worked precisely for natural scenes (Fig. 3.4, third row) in which the superposition of broadly filtered horizontal and vertical versions was almost identical to original images (see example Fig. 3.2, top). Therefore, even for the most complex stimuli that contained a rich mixture of multiple orientations, we found sensitive cortical tuning for changing orientations rather than for currently presented orientations.

Population tuning to turned-off orientations after a switch from the superposition to a single orientation (VH to V or H, 0.29±0.12 *mad,* $x10^{-5}$ deltaF/F; see Materials and Methods and Suppl. Material Table 3.S1-S3; $n = 48$: medians across 12 experiments, both orientations, and both filter conditions of the natural images) suggests that responses to orientations that were turned-off are stronger than responses to orientations that sustained. When these components were measured directly, tuning amplitudes were indeed higher for turned-off orientations (switches from H or V to blank, 0.58±0.28 *mad,* $x10^{-5}$ deltaF/F) than for sustained orientations (V to V or H to H, 0.14±0.16 *mad,* $x10^{-5}$ deltaF/F; $n = 48$; paired two-tailed sign test $p<0.0001$).

When comparing responses to narrowly and broadly filtered images, we calculated for the latter a decrease in modulation depth of 24±7% (the difference between preferred and orthogonal responses, paired t-test $p<0.02$), while we did not observe differences in tuning width (broadly filtered, HWHM 46±2° *sem*, narrowly filtered, 49±3° *sem*, pairwise difference 3±3° *sem,* $n = 8$ experiments, $p = 0.33$). Because broadly filtered images provide enriched orientation content, a divisive normalization across populations of neurons with different preferred orientations (MacEvoy et al., 2009; Busse et al., 2009) may cause the observed decrease in modulation depth. For a proof of principle, we finally applied sequences of square-wave gratings in an additional experiment (Fig. 3.4 bottom). As expected, modulation depth was large (cf. colorbar) and also for those stimuli we found a dominant representation of orientation change.

Even though VSD imaging may be a powerful tool to measure neuronal population dynamics with high spatiotemporal resolution, the relationship between the imaging signal and spiking activity is not entirely clear. Eriksson et al. (2008) suggest a close relationship between spike rate and the derivative of the VSD response rather than its magnitude. Such behavior might especially apply to the rising phase of the membrane potential after stimulus onsets (Jancke et al., 2004; Sit et al., 2009). Moreover, combined VSD and calcium-sensitive dye imaging suggest that the relationship between spiking activity and the amplitude of the VSD response

depends on stimulus intensity (Berger et al., 2007). Chen et al. [2012] propose that these relationships are well captured by a power function with an exponent of ~4, similarly as observed for the relationship between average membrane potential and spike rates in single V1 neurons (Anderson et al., 2000; Finn et al., 2007). However, dependencies on individual experimental settings, on the particular stimuli used, and on the likely differences between species, are widely unexplored. Hence, to address this issue, and to directly exclude the possibility that the dye signal levels reporting the difference to past orientations would merely reflect subthreshold activity (Petersen et al., 2003; Jancke et al., 2004; Berger et al., 2007; Eriksson et al., 2008), we additionally performed electrophysiological recordings in four hemispheres.

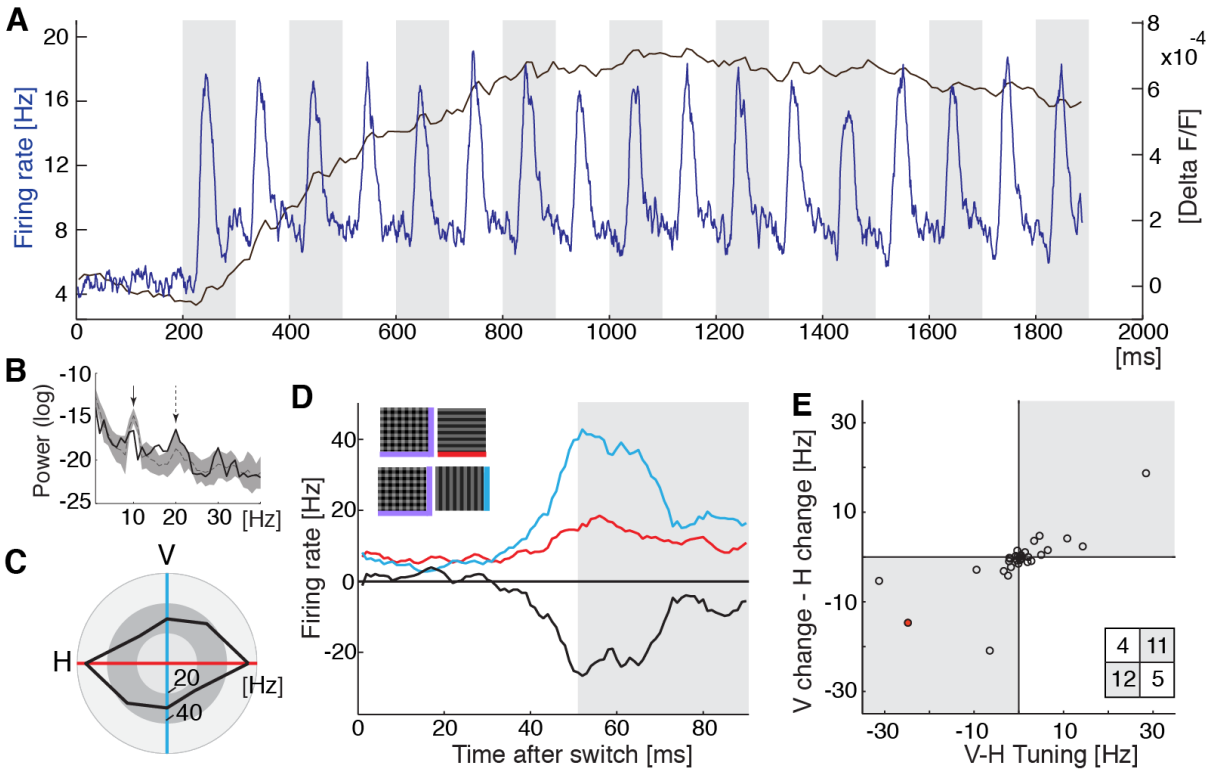


Figure 3.5: Electrophysiological recordings. (**A**) Multiunit activity (MUA) versus optical signal. Blue trace shows averages over 32 recordings (4 different hemispheres) and over randomized 10-Hz sequences of gratings. Black trace outlines the time course of the unfiltered optical signal (average over space and all 10-Hz sequences, 1 experiment). (**B**) Log power spectrum of the evoked dye signal (in the steady phase 900-1900 ms). Spectrum of trace in **A** (black bold line) and the log spectra for filtered natural stimuli of all 12 experiments (narrowly filtered stimuli, gray dotted line shows mean, shaded area shows *std*). First arrow points to stimulus frequency (10 Hz), the dotted arrow indicates first harmonic (20 Hz). (**C**) Example spike recording. Spike responses to moving gratings of four different orientations (temporal average over 800 ms) indicating orientation tuning to horizontal (red axis). (**D**) Responses to either horizontal (red) or vertical (blue) orientation after presentation of their superposition, black trace plots difference. (**E**) Summary: Differences between responses to vertical and horizontal moving gratings (i.e. orientation tuning, cf. black outline in **C**) are plotted against differences in responses to vertical and horizontal after presentation of their superposition (cf. black trace in **D**). Time averages across the same time window as used for the imaging data (see gray shading in **D**). Each circle plots MUA (*n* = 32, red dot represents example shown in **C**-**D**). Spike responses tend to be higher after the preferred orientation was switched-off from the superposition (gray shaded quadrants) than when it was the remaining orientation; Fisher's exact test one-tailed p = 0.016; inset shows count for each quadrant.

Figure 3.5A shows average multiunit activity (MUA, blue trace) recorded after imaging (black trace) in response to grating sequences. Along the entire stimulus sequence small bumps in the dye signal level coincided precisely with the generation of spiking activity, even at low levels during the early phase of the imaged response. Meaningful neuronal signals in VSD recordings can indeed be small in relation to overall activity (cf. Sharon and Grinvald, 2002, for a signature of cross-inhibition suppression). Benucci et al. [2007] showed that small oscillations in the dye signal correlated with the frequency of counterphase oscillating gratings. In Onat et al. [2011a] we showed for the first time that small bumps of activity represented exactly the retinotopic propagation of moving gratings. Very recently, using longer sequences (>10 s) of flashed gratings with different spatial phase, these formerly elusive retinotopic components in the imaging signal were shown in awake monkey (Omer et al., 2013).

For each electrode-recording site we determined the preferred orientation tuning of MUA. Figure 3.5C depicts average spiking responses to moving gratings of different orientations. In this example, the neurons' preferred orientation was horizontal (red axis). Next, in Fig. 3.5D, their time-resolved responses to a switch from superposition to either a horizontal or a vertical orientation are shown (see icons). Despite the fact that these neurons were tuned horizontally (Fig. 3.5C), responses for the switch to the nonpreferred vertical orientation were larger than for the switch to the preferred horizontal orientation (black trace shows the difference). In Fig. 3.5E spiking activity from all recorded units is summarized. The plot indicates that the recorded neurons responded stronger when their preferred stimulus was removed rather than when it was present, similarly as observed in our imaging signals (Fig. 3.4, bottom three rows).

Finally, we used a single-switch paradigm (Fig. 3.6; cf. Eriksson et al., 2010; Eriksson et al., 2012), also to rule-out that stimulus frames other than the switch-pair under analysis (Felsen et al., 2002; see Materials and Methods) may have significantly affected the observed response behavior using continuous sequences. The superposition of two orientations was presented for 30, 100, and 500 ms followed by a sudden switch to a single orientation (Fig. 3.6A). Whereas switches after 30 ms delay produced activity that was tuned to the orientation present after the switch, longer delay times yielded the opposite effect (Fig. 3.6B). Here again, population activity represented the orientation that was removed rather than the sustained orientation (Eriksson et al., 2010; Eriksson et al., 2012). This effect was stable and slightly increased (though not significantly) for 500 ms delays (Fig. 3.6C; in support of these findings see Eriksson et al. [2012] for similar results with a stimulus duration of 250 ms). In Fig. 3.6D the time-courses of global population activity are shown. For each condition, both the initial responses to the superposition as well as the second response to a single orientation were represented by peaks of activity that were separated in time with longer switch delays (dark gray area). For 100 and 500 ms delays, activity to the superposition nearly adapted to baseline levels (occasionally we observed some further oscillations, see gray trace in first plot and black trace in third plot), followed by a strong response after the stimulus changed to a single orientation. In contrast, for the 30-ms switch, activity did not adapt to baseline and responses to both stimuli were merged within a double-peak transient. To demonstrate that the orientation-selective part of the responses was directly traceable across the cortical activation patterns, we show time-averages of the most active pixels in each condition (Fig. 3.6E, see maps). Note the opposing cortical patterns in the maps depending on the switch direction and its delay times. The temporal evolution of the orientation specific patterns is shown as their correlation to a standard VH orientation map over time (black and gray traces in Fig. 3.6E; cf. Suppl. Fig. 3.S1). Thus, by using only a single switch, these measurements confirmed our findings obtained with stimulus sequences: a short 30-ms period of stimulation

resulted in the representation of current orientations (Fig. 3.6E left), whereas longer stimulus periods (100 and 500 ms) caused representation of the difference to past orientations (Fig. 3.6E, second and third graphs).

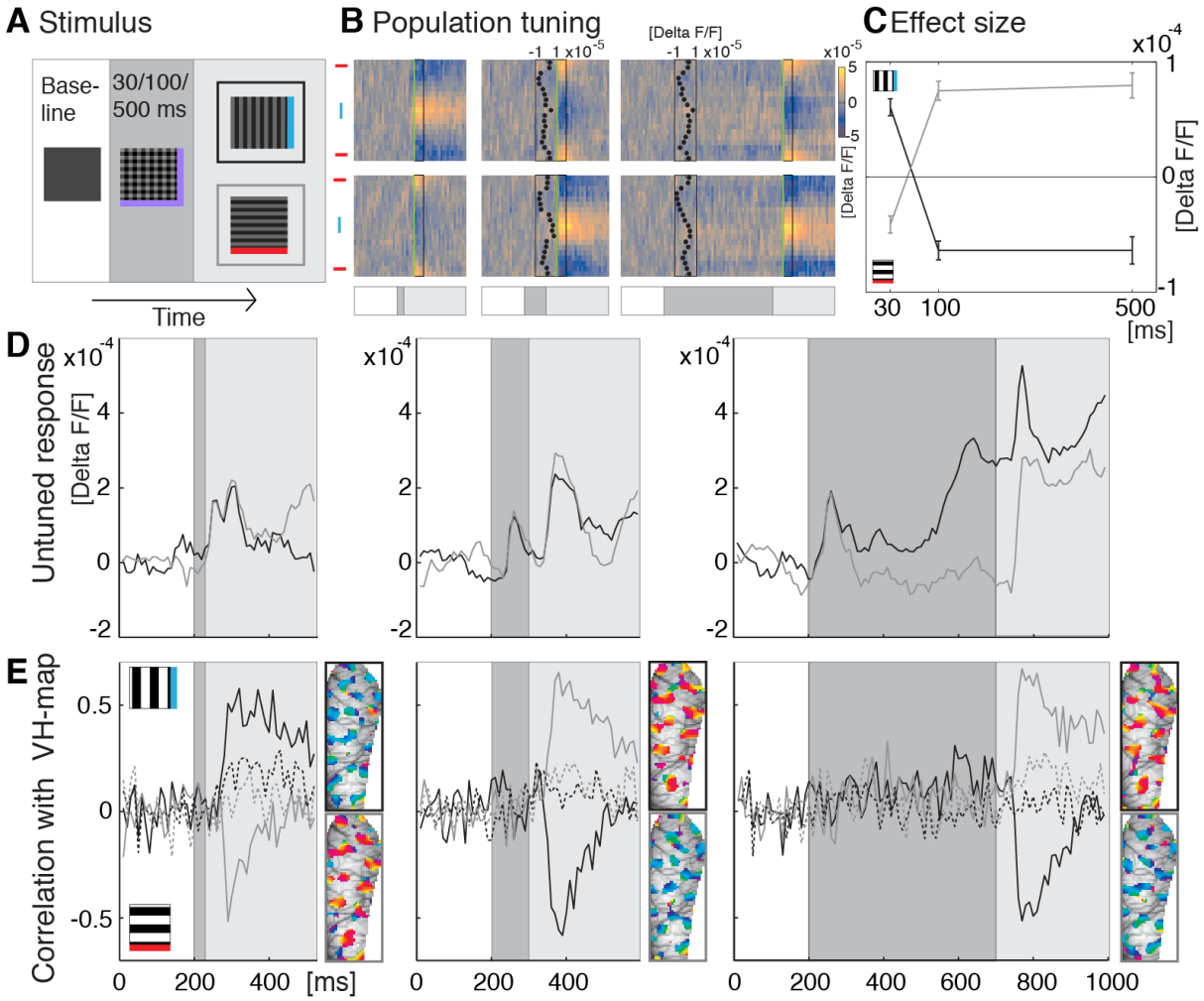


Figure 3.6: Single switches. (**A**) Superimposed gratings were presented for 30, 100, or 500 ms (see dark gray box) before switching to either vertical or horizontal orientation (here and in all subsequent plots these switches are marked in black and gray, respectively). The intertrial interval, in which a blank stimulus was shown, was set to a minimum of 5 s (200 ms were recorded when a new stimulus started to obtain baseline activity, see white boxes). (**B**) Population tuning in response to the 6 stimulus conditions; switches to vertical along top row, horizontal switches at bottom. Same conventions as in previous figures (cf. Figs. 3.2F, 3.3, 3.4). Temporal structure of stimulation is indicated at bottom (cf. **A**). Green lines and boxes mark the time window used to calculate effect size in **C** and maps in **E** (50-90 ms after switch). Insets: Bimodal distribution representing the constituent orientations of the superimposed gratings (time averages 50-150 ms after response onset). (**C**) Effect size. Data points specify average modulation depth (see Materials and Methods) dependent on different presentation times of the superimposed gratings (x-axis). Bars depict standard error (80 stimulus repetitions). Positive values indicate higher activity at vertically tuned pixels (blue icon) and negative values higher activity at horizontal pixels (red icon). When testing for differences between timing conditions, pooling vertical and horizontal conditions (sign inverting the latter), we found significant differences between the 30-ms and the 100-ms condition (pairwise two-tailed t-test: 30 vs. 100 ms $p<0.001$, $n=160$) but no significant differences between the 100-ms and 500-ms condition ($p=0.79$). (**D**) Time courses of global activity (unfiltered spatial averages across imaging frames) in response to the six stimulus conditions. Shaded areas depict the temporal structure of stimulus conditions (cf. **A**). (**E**) Correlation of the cortical activity patterns with a standard VH orientation map (see Materials and Methods and Suppl. Fig. 3.S1). Positive values indicate similarity with activity pattern representing vertical orientation (icon on top), negative values indicate correlations to the horizontal map. On the right of each graph the 30% most active pixels across the imaged cortex are shown for each condition (time averages of 4 imaging frames, 50-90 ms after switches). Colors code preferred orientation at each pixel location (same conventions as in Fig. 3.2D). Upper maps visualize population tuning to a switch to vertical (outlined black), lower maps show switches to horizontal (outlined gray). Dotted lines show correlations with the additive map instead of the standard subtractive VH-orientation map (see Materials and Methods) as control.

3.5 Discussion

Our main result is that for sequences of natural scene contours presented with 10 Hz, activity no longer led to population tuning that represented actual stimulus contours. Instead, when compared with the preceding image, the cortical activity patterns characterized exactly the difference in orientations. Consequently, large amounts of incoming data were relatively suppressed, reminiscent of differencing methods (Fowler et al., 1995) used for video data compression in communication technology. For higher temporal frequency (33 Hz), activity was instead updated linearly, providing an ongoing representation of current stimulus orientation (Ringach et al., 1997; Benucci et al., 2009). Because we opposed two stimulation dynamics selected from a wide range of possible sequence frequencies, the exact time course of the transition between the different encoding schemes remains to be determined. Interestingly, in the same vein, using single squares of light Eriksson et al. [2008] showed in ferrets that V1 responses to rapidly presented stimulus switches (<83 ms) were dominated by stimulus onsets, whereas both VSD imaging and spiking responses to longer stimuli (>133 ms) carried additional prominent information about stimulus offset. We speculate that joint processing at various stimulus temporal frequencies is required to produce a coherent interpretation of a visual scene (Jonides et al., 1982; Rucci et al., 2007; Rucci, 2008; Belitski et al., 2008; Nikolić et al., 2009; Onat et al., 2011a; Jurjut et al., 2011; Eriksson et al., 2012), which might be implemented through differences in coherence between neuronal signals carrying different information at different frequencies, as recently shown for orientation tuning in monkey V1 (Gilad et al., 2012; Womelsdorf et al., 2012).

3.5.1 Interaction between adaptive- and off-response components

Because our results were dependent on stimulus frequency, we provide strong evidence that input history has a decisive effect on cortical orientation tuning. Several time-dependent changes in cortical orientation selectivity could be accounted for by mechanisms of adaptation. Specifically, stimulus-selective adaptation of visual cortical neurons has been shown to reduce responsiveness and causing a shift in tuning curves away from the adapting orientation on short time scales (Movshon & Lennie, 1979; Müller et al., 1999; Felsen et al., 2002). In our experiments, neuronal adaptation mechanisms (Galaretta & Hestrin, 1998; Varela et al., 1999; Sanchez-Vives et al., 2000) and immediate tuned suppression (Nelson, 1991) may decrease activity for the sustained component. Adaptation to a single orientation was shown to also enhance the representation of orthogonal orientations (Dragoi et al., 2002), which in our case would furthermore boost the on-response for the newly added orientation. Thus, for switches within sequences from a single, (V) or (H), orientation to a superimposed (VH) stimulus, adaptation alone would explain the resulting dominant representation of the newly added orientation (instead of both of the current orientations).

These mechanisms, however, cannot entirely explain the prominent representation of the disappeared orientation after exposure to superimposed orientations. The latter provide no bias that could induce adaptation of a particular orientation before the switch. Thus, in addition, responses to the disappeared orientation must be involved (Bair et al., 2002; Sit et al. 2009; Eriksson et al., 2010; Eriksson et al., 2012). Signals following stimulus removal are commonly referred to as visual off-responses, which tend to increase with stimulus duration (Duysens et al., 1996), as similarly found in the somatosensory cortex (Kyrazi et al., 1994). A likely assumption is that off-responses result from post-inhibitory rebound due to sustained hyperpolarization arising from synaptic inhibition (Pernberg et al., 1998, but see Scholl et al., [2010] for auditory cortical neurons), supposedly mediated by tuned push-pull mechanisms within the cortical circuitry (Hirsch et al., 2003). Importantly, such tuned off-responses and

adaptive contributions can be disentangled in our study. A change from superimposed orientations to a single orientation has two underlying constituents: A switch from one orientation to blank (off-component), and an overlaid continuous presentation of the orthogonal orientation (sustained, adaptive component). Fig. 3.7 summarizes such a composition for a switch from superposition (VH) to a single orientation (V or H). Median fits for responses to each of the constituent stimuli are shown (blue/red; see Results), the gray curve outlines their average (i.e. divisive normalization, MacEvoy et al., 2009). The black curve shows the median fit for the responses that we obtained for the composite switch (Suppl. Table 3.S1-S3), indicating tuning to the orientation that was turned off.

Approximating the measured change response for this type of switch (VH to H and VH to V) by a weighted average of the two component responses (MacEvoy et al., 2009, Busse et al., 2009), resulted in a significantly higher average contribution of 60% from the off-component, compared with 40% from the adaptive component (see Suppl. Methods and Table 3.S4 for details). Thus, for a 100-ms stimulus period the off-component is facilitated (see blue arrows in Fig. 3.7) and overrides the response component that undergoes adaptation. Accordingly, due to increased contribution of orientation-selective off-responses, the combination of adaptive- and off-response components results in a representation of the difference between the past and the present image. Taken together, our data suggest mutual interaction across population responses to changing and nonchanging, that is, sustained, features. As a consequence, after periods of longer stimulation, a stimulus-change within stimulus sequences triggers activity to report that a particular feature has disappeared in comparison to what is left.

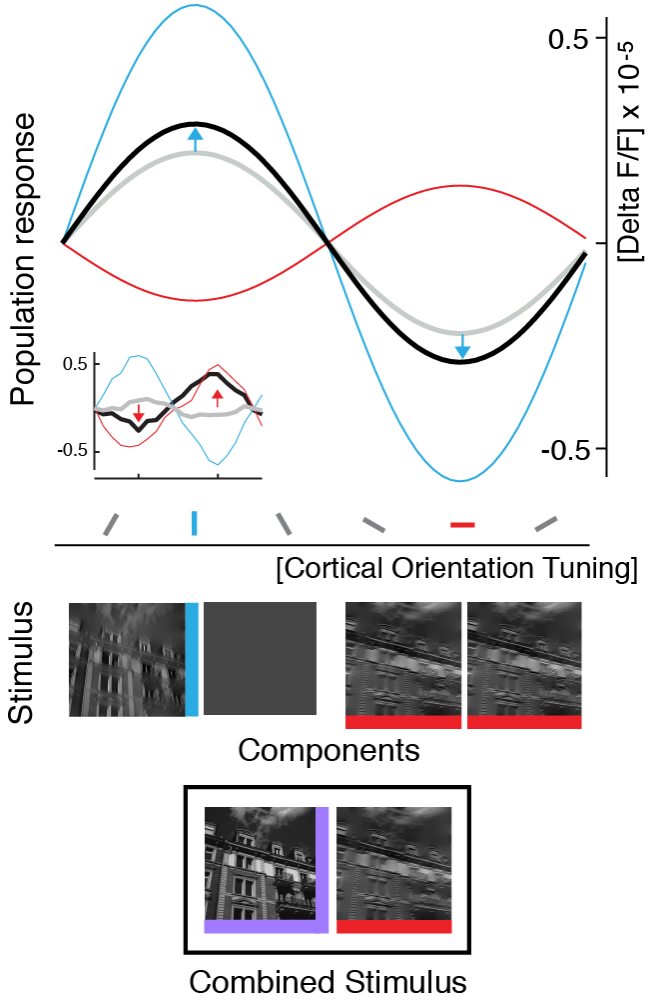


Figure 3.7: Disentangling components that produce change coding within 10-Hz sequences. Schematic of the response to a switch from superimposed orientations to a single orientation (bottom icon, black box) and the responses to the constituents of this switch (middle icons): Median of fitted population tuning curves (see Materials and Methods and Results) for off-component (blue), adaptive component (red), and composite switch (black; *n* = 48, medians across 12 experiments, both orientations, and both filter conditions of the natural images). Note the blue arrows, indicating deviations from the component average (gray, see main text for quantification). Overall, the measured response to the combined stimulus represents the difference between the past stimulus (i.e., superposition, purple icon at bottom) and the current orientation (horizontal, red icon at bottom). This difference is the orientation that was turned off (see peak of black curve at position of vertical orientation tuning, blue). The opposite result was found for 33-Hz sequences (see example in inset; nonfitted data of an experiment obtained with narrowly filtered natural images).

Stimuli usually do not occur in temporal isolation, but within a temporal context where stimuli in the range of a few milliseconds up to seconds appear in succession (cf. Zucker & Regehr, 2002). In an early account, so-called paired pulse stimulation (Allison, 1962) was used, showing that responses to a second stimulus were severely suppressed when compared with the first response dependent on the interstimulus interval. Although the underlying mechanisms are not resolved there is agreement that paired pulse suppression is a cortical phenomenon with a strong GABAergic contribution (Wehr & Zador, 2005). In contrast to paired pulse suppression, which addresses the transient behavior of cortical response properties, continuous, periodic stimulation addresses the response behavior during so-called steady-state conditions (Onat et al., 2011b). Conceivably, our single-switch condition (Fig. 3.6) resembles paired stimulation, while the sequence conditions reflect steady-state stimulation as termed by others. Interestingly, the built-up and the time course of inhibition during steady-state stimulation can be quite different from what can be inferred from paired stimulation (Hickmott, 2010). It is therefore surprising that we obtained equivalent results for both paradigms which may hint on involvement of widespread excitatory-inhibitory mechanisms (Markram et al., 1998). Interestingly, in a recent work by Olsen et al. [2012] it was shown that neurons in cortical layer six have a major impact in controlling the gain of activity in upper layer neurons, the layers imaged here. Most strikingly, gain control occurred without changing orientation tuning. Thus, such mechanism may be the ideal candidate to balance the relative weights of adaptive and off-components dependent on visual input frequency. This might be realized by different gating of activity across two neuronal circuits acting (i.e. "competing", Adesnik and Scanziani, 2010) in parallel: those including adapting neuronal populations and those that produce off-responses.

3.5.2 Coding of stimulus differences between past and present

Representation of difference in the primary visual cortex may lead to attenuation of redundancies (Attneave, 1954) over time and increased sensitivity to dissimilar structures, such as differently oriented borders of objects (Das & Gilbert, 1997; Downar et al., 2000; Dragoi et al., 2002; Desbordes & Rucci, 2007; Rucci et al., 2007; Rucci, 2008; Beste et al., 2011). The proposed activity dynamics of adaptive- and off-response components might be viewed as short-term memory processes (Sperling, 1960; DiLollo, 1977; Coltheart, 1980), which begin with stimulus onset (DiLollo, 1977), trigger recurrent networks (McCormick et al., 2003), and may be coupled with feed-back from higher areas (Rockland & Pandya, 1979; Roland et al., 2006; Golomb et al., 2010; Vetter et al., 2013) to allow prolonged influence of past activity (Coltheart, 1980; Duysens et al., 1985; McCormick et al., 2003) on the processing of current input (Gould, 1967; Jonides et al., 1982; Jancke, 2000; Eagleman et al., 2004; Eriksson et al., 2008; Nikolić et al., 2009; Eriksson et al., 2010; Glasser et al., 2011; Eriksson et al., 2012).

During free viewing of natural scenes intersaccadic durations (in human ~250 ms on average, Kuang et al., 2012; in cat >2000 ms, Moeller et al., 2004) can be even larger than our long stimulus intervals (>100 ms). After the onset of each new fixation (i.e. at low temporal frequencies) difference representation might facilitate cortical encoding of luminance discontinuities and edges at spatial scales larger than covered by retinal ganglion cells and also by neighboring cortical cells with similar orientation tuning (Müller et al., 1999). Specifically, cortical difference representation might compensate the extensive luminance correlations (and thus, reduce redundancies) conveyed by transient activity of retinal ganglion cell populations immediately after a saccade (Desbordes & Rucci, 2007).

High correlation among activity of ganglion cells immediately after a saccade signals long-range correlations in natural images (Kuang et al., 2012). A second regime brought in by

microscopic eye movements, operating at higher temporal frequencies, is proposed to cause whitening of input while decorrelating retinal activity (Kuang et al., 2012). Hence, these retinal signals emphasize small spatial details in visual structures during fixation (Rucci et al., 2007; Desbordes & Rucci, 2007; Rucci, 2008; Kuang et al., 2012), even of contour orientations (Rucci & Desbordes, 2003). The here reported cortical representation of current orientations at higher frequencies (33 Hz) may reflect the transmission of the acquired information further downstream. In conclusion, the two temporal regimes of eye movements may allow complementary contributions (Snodderly et al., 2001) in a motion-based coarse-to-fine processing of visual information (Parker et al., 1992; Jancke, 2000; Ahissar & Arieli, 2001; Geisler, 2001; Henning et al., 2002; Desbordes & Rucci, 2007; Rucci, 2008; Ahissar & Arieli, 2012; Meirovithz et al., 2012) and could be an efficient mechanism for perception of salient structures in the environment. Whether these regimes act linearly at the cortical level cannot ultimately be decided upon our data. Modeling at the retinal level suggests that the two regimes can however be well captured using linear approaches (Desbordes & Rucci, 2007).

More generally, the time-dependent cortical coding of the difference to past events can be interpreted as an early cortical signature of mismatch signals between ongoing stimulation and abrupt stimulus changes, as first described in the auditory domain (Näätänen et al., 1978). Most recently, mismatch signals have been shown in the primary cortex of the mouse to be cooperatively influenced by motor-related input (Keller et al., 2012). We suggest that both adaptation (Jääskelainen et al., 2004) and stimulus off-responses play in important role in generating mismatch signals.

Finally, responses to stimulus differences fit well conceptually with predictive coding principles, proposing that deviations from cortically generated predictions are propagated up the visual hierarchy as error signals (Friston, 2005; Garrido et al., 2009). Given the prediction that contour orientations remain stable over prolonged periods of time, as during periods of fixation, error signals would correspond to representations of change, as measured here. Predictive coding principles were found as early as in the retina (Srinivasan et al., 1982; Hosoya et al., 2005) and have been used in recent modeling frameworks of cortical visual responses (Rao & Ballard, 1999; Friston, 2005; Spratling, 2010; Boerlin & Denève, 2011; Spratling, 2012), also suggesting a combination of stimulus and error-like coding within single neurons (Eriksson et al., 2012). The exact timescales (Thorpe et al., 1996) at which different predictive states may evolve under natural viewing conditions need to be further explored and elaborated in computational models to account for the qualitatively different behaviors of cortical responses—from ongoing representation to representation of difference—that are reported here.

3.6 Supplementary material

3.6.1 Contents of supplementary material

Supplementary figures (3.6.2)

Figure 3.S1: Time course of difference representation.
This Figure shows the correlation between population orientation tuning over time and a standard VH-orientation map. The purpose of this Figure is to also show the time course of the main effect beyond the window used for analysis throughout the paper (50-90 ms after switch). For the single switch paradigm correlation over time is shown in Fig. 3.6.

Figure 3.S2: Stimulus-triggered versus switch-triggered averages.
A. Illustration of the stimulus paradigm and the averaging procedure for analysis. B. The purpose of this Figure is to show how the differences in switch-triggered responses that we find for the different stimulus durations (30 ms vs. 100 ms) lead to differences in stimulus-triggered responses.

Figure 3.S3: Switches to blank, sustained orientation, and from superposition to a single orientation.
The purpose of this Figure is to show responses to a switch from a single orientation to blank, as a supplement to the data provided in Figure 3.4 of the main text.

Supplementary tables (3.6.3)

Table 3.S1: Statistical evaluation of population tuning, step 1.
Table 3.S2: Statistical evaluation of population tuning, step 2.
Table 3.S3: Statistical evaluation of population tuning, step 3, orientation-tuning amplitudes.
The purpose of these three tables is to give a comprehensive report of the statistical results.

Table 3.S4: Response decomposition (results): contributions of adaptive and off-components to change responses.
The purpose of this table is to show the results from the comparison of responses to constituent stimuli (off-switch of one orientation and sustained presentation of the orthogonal orientation) with the response to their composite stimulus (switch from superimposed orientations to a single orientation).

Supplementary methods (3.6.4)

Response decomposition (methods): evaluation and fitting procedure
Here we describe the method that was used to compute the results in Table 3.S4, which are referenced in the Discussion section of the main text. The response to the switch from superposition to single orientations is approximated using responses to the two constituent stimuli.

3.6.2 Supplementary figures

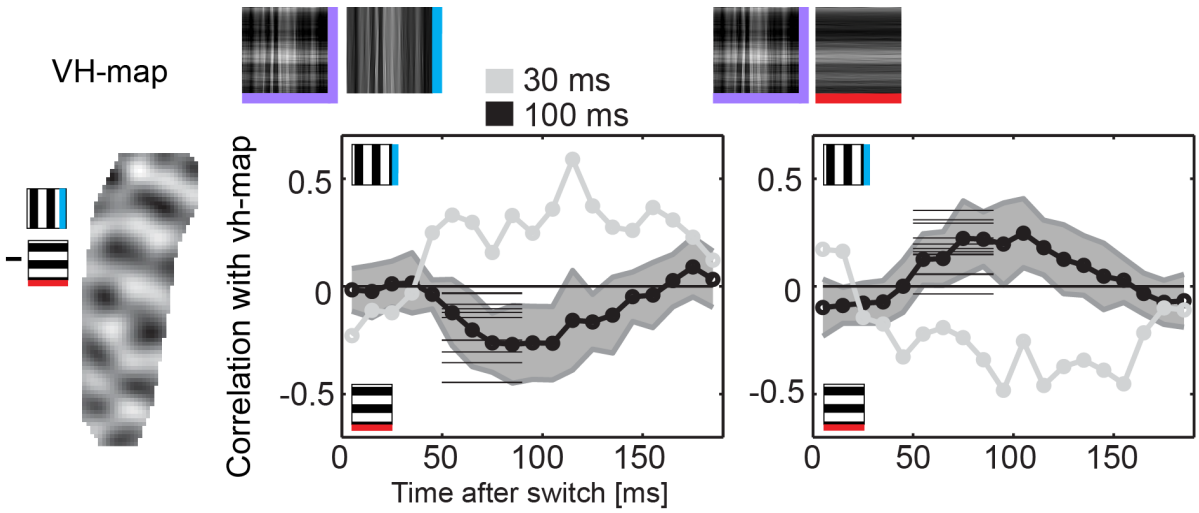


Figure 3.S1: Time course of difference representation. Correlations between responses to a switch from superposition to single orientation (measured within sequences of narrowly filtered natural images) and VH-map (independently measured differential horizontal-vertical activity pattern, see Materials and Methods) over time. Positive values indicate similarity with orientation pattern evoked by a vertical grating (corresponding to white regions in VH-map), and negative correlation indicates similarity with the horizontal pattern (dark regions in VH-map). When presented within 30-ms time sequences, increase in correlation showed tuning for the orientation that was actually present (thin gray traces; average correlation from 50-90 ms after switch, 1 experiment: 0.28 for switch to vertical and -0.25 for switch to horizontal). For 100-ms stimulus time sequences (black traces, gray shaded area mark standard deviation across 12 different experiments; black lines show average values from 50-90 ms for the individual experiments), correlation values reversed, indicating tuning for the orientation that had disappeared (time window 50-90 ms: -0.21±0.15 (*std*), left tailed t-test against zero $p<0.001$ for switch to vertical; 0.17±0.11 (*std*), right tailed t-test $p<0.001$ for switch to horizontal).

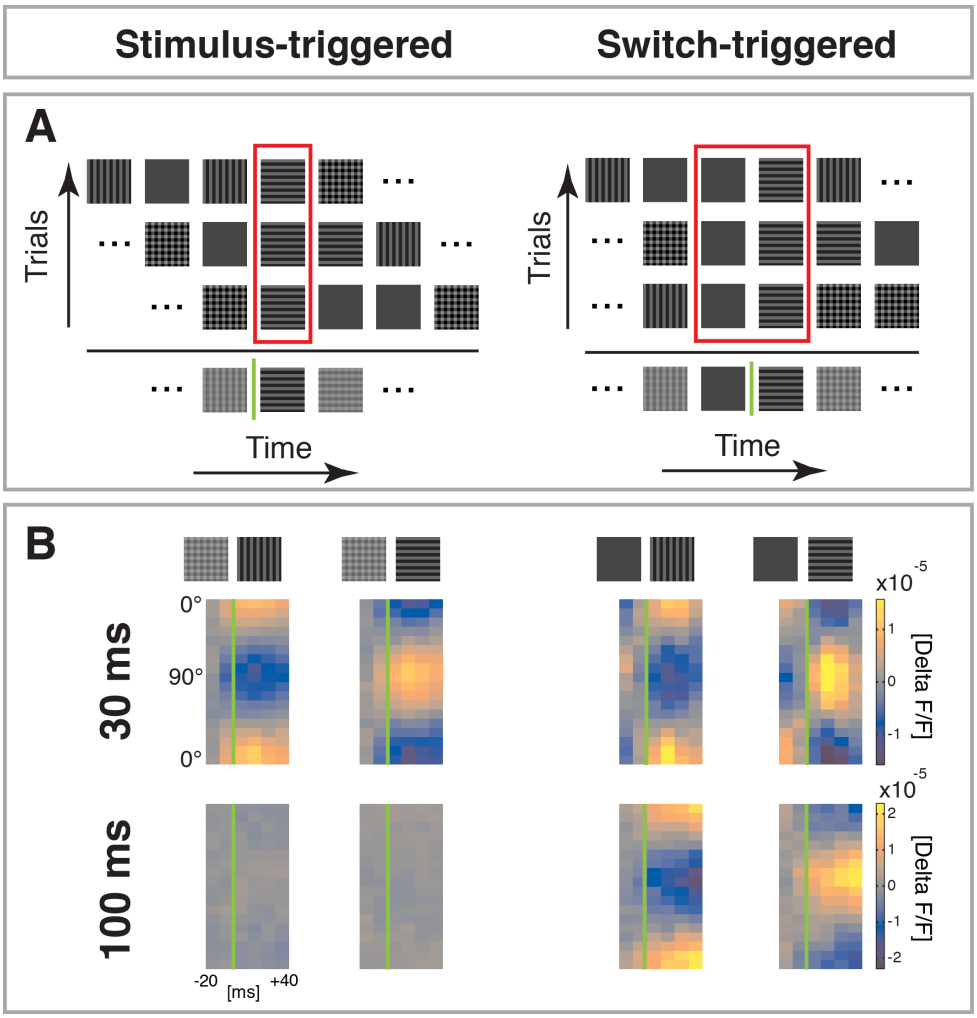


Figure 3.S2: Stimulus-triggered versus switch-triggered averages. A: Illustration of averaging procedures. Exemplified are three different stimulus sequences (5 stimuli from 17-stimulus-long sequences are shown; in the experiments we used at least 64 different sequences of at least 17 images, see Materials and Methods). On the left, sequences are aligned at a specific stimulus (horizontal in this example, red box) and averaged. Note that the different stimuli before and after are averaged. For *stimulus-triggered averaging*, this procedure is used to average the cortical responses to a particular single stimulus across all sequences. On the right, sequences are aligned at switches between specific pairs of stimuli (blank to horizontal in this example) and averaged. In this case the pair is fixed, while different stimuli before and after are averaged. For *switch-triggered averaging*, this procedure is used to average cortical responses to a particular stimulus pair (i.e. switch) across all sequences.
B: Stimulus-triggered versus switch-triggered responses. The left side shows average population responses triggered by either the vertical or horizontal stimulus. Here we averaged over all preceding stimuli (see **A**). This is usually seen as the pure response to the single orientation stimulus. When measured within 30-ms sequences, there is clear population tuning to the currently resented stimulus. However, when measured within 100-ms sequences, the stimulus-triggered response is flat, indicating that the different underlying switch-triggered responses (change responses), comprising switches in which the orientation was removed as well as switches in which is was added, average each other out. The right side shows example switch-triggered responses for switches from blank to a single orientation. Here, we find orientation tuning for both sequence speeds. Altogether this illustrates that at 30-ms presentation duration, population-tuning responses are evoked by individual stimuli in a sequence, while at 100-ms presentation duration responses are highly dependent on specific pairs of stimuli. The window shows 30-90 ms after onset of the oriented stimulus, the green line is at 50 ms (the approximate response latency).

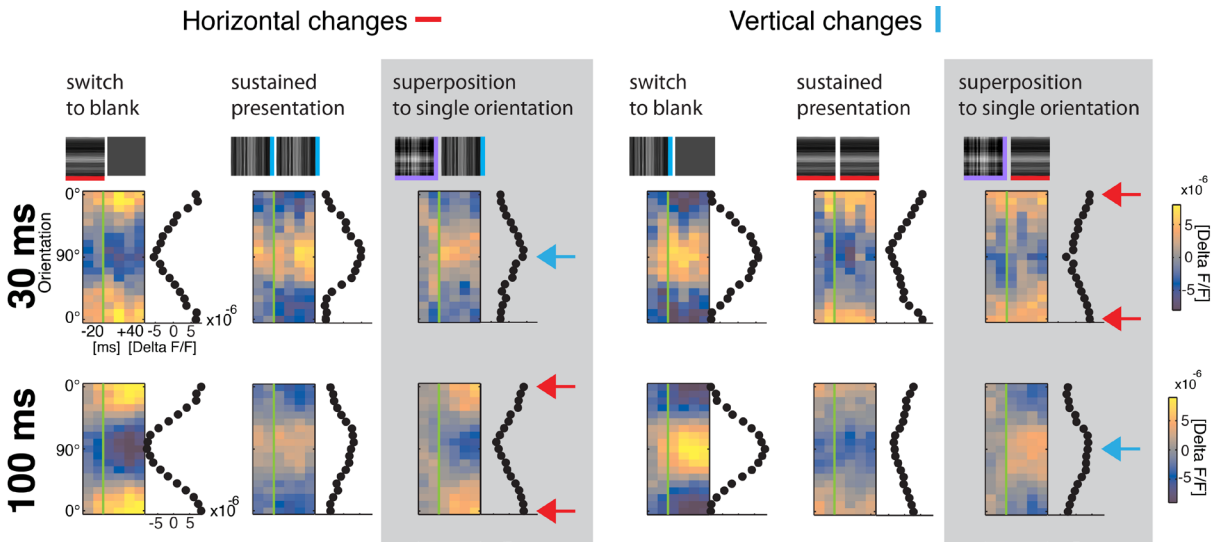


Figure 3.S3: Switches to blank, sustained orientation, and from superposition to a single orientation. We find that population responses were faithfully tuned to both turned-off and sustained orientations (first and second column, respectively) after both 30 ms (first row) and 100 ms (second row) presentation duration. When one orientation was turned-off from superimposed orientations (gray columns), the direction of population tuning was dependent on stimulus timing: Whereas in fast sequences the actual (i.e. the sustained) orientation after the switch was represented, slower sequences resulted in a representation of stimulus difference (i.e. the removed orientation). Conventions as in Fig. 3.4, main text.

3.6.3 Supplementary tables

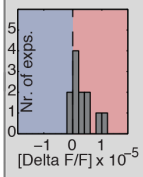
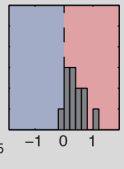
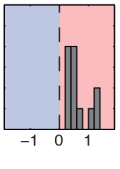
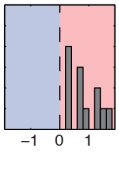
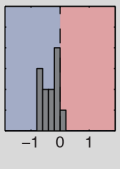
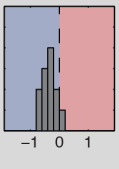
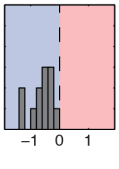
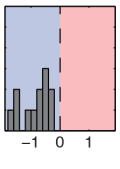
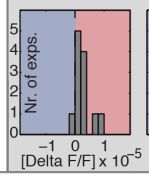
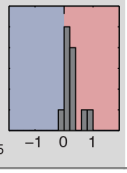
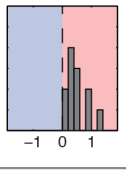
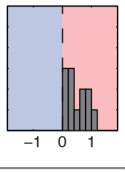
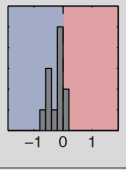
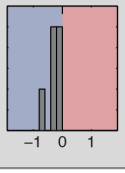
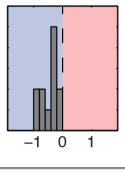
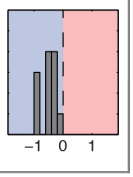
	V to VH	VH to V	B to H	H to B	H to VH	VH to H	B to V	V to B
33-Hz Gratings	56***	339***	350***	727***	8 (p=0.97)	123***	471***	618***
33-Hz Narrow	6 (p>0.99)	37**	34**	114***	5 (p>0.99)	33*	47***	151***
10-Hz Narrow	563***	671***	1180***	2144***	422***	443***	1463***	2663***
10-Hz Broad	388***	370***	704***	810***	324***	306***	728***	1051***
10-Hz Gratings	280***	184***	372***	450***	149***	179***	276***	518***

Table 3.S1: Statistical evaluation of population tuning, step 1. In the first step, we evaluated the difference to zero (or flat response). Shown are the X^2-statistic and p-values (*** = p<0.001, ** = p<0.01, * = p<0.02). For the 33-Hz data, we find significant differences to zero for switches with a single orientation present after the switch (columns 2, 3, 6, and 7) and off-switches from a single orientation to the gray screen (columns 4 and 8). For the transition from a single orientation to the superposition, we find no significant deviation (columns 1 and 5), as is expected from a representation of both orientations present after the switch, with the exception of one switch (V to VH, column 1, row 1). For the 10-Hz data, the population tuning curves deviate significantly from zero in all cases (lower three rows).

	V to VH	VH to V	B to H	H to B	H to VH	VH to H	B to V	V to B
33-Hz Gratings	4 (p>0.99)	5 (p>0.99)	5 (p>0.99)	12 (p=0.73)	3 (p>0.99)	10 (p=0.87)	6 (p=0.98)	19 (p=0.26)
33-Hz Narrow	3 (p>0.99)	6 (p=0.99)	4 (p>0.99)	7 (p=0.97)	5 (p>0.99)	10 (p=0.88)	5 (p>0.99)	5 (p>0.99)
10-Hz Narrow	101 (p>0.99)	139 (p>0.99)	119 (p>0.99)	143 (p>0.99)	52 (p>0.99)	126 (p>0.99)	105 (p>0.99)	120 (p>0.99)
10-Hz Broad	76 (p>0.99)	116 (p>0.99)	97 (p>0.99)	86 (p>0.99)	96 (p>0.99)	107 (p>0.99)	66 (p>0.99)	85 (p>0.99)
10-Hz Gratings	7 (p=0.96)	13 (p=0.69)	9 (p=0.93)	21 (p=0.16)	13 (p=0.65)	3 (p>0.99)	33 (p<0.01)	14 (p=0.62)

Table 3.S2: Statistical evaluation of population tuning, step 2. In the second step, we fit sinusoids to the population tuning curves of each experiment and repeat the test against zero on the residuals. Shown are X^2-statistics and p-values. Now there are no significant differences, except in one case (10-Hz-Gratings B to V). In all cases the X^2-values decreased substantially, thus a large part of the variance in the data can be explained by the fitted function.

	V to VH	VH to V	B to H	H to B	H to VH	VH to H	B to V	V to B
33-Hz Gratings	-0.33	-0.88	0.89	1.32	-0.10	0.53	-1.02	-1.21
33-Hz Narrow	-0.08	-0.28	0.27	0.53	0.01	0.26	-0.33	-0.62
10-Hz Gratings	1.29	0.99	1.46	1.62	-0.90	-1.05	-1.19	-1.79
10-Hz Narrow	0.30 (0.21*)	0.39 (0.34**)	0.61 (0.45***)	0.84 (0.77***)	-0.32 (-0.32**)	-0.34 (-0.32**)	-0.62 (-0.56***)	-0.84 (-0.67***)
10-Hz Broad	0.25 (0.20**)	0.30 (0.22**)	0.48 (0.43***)	0.49 (0.35***)	-0.24 (-0.14*)	-0.31 (-0.25***)	-0.45 (-0.37***)	-0.52 (-0.49***)

	V to VH	VH to V	B to H	H to B	H to VH	VH to H	B to V	V to B
10-Hz Narrow								
10-Hz Broad								

All values: x10^{-5} DeltaF/F; Sign test: ***p<0.001, **p<0.01, *p<0.05

Table 3.S3: Statistical evaluation of population tuning, step 3, orientation-tuning amplitudes.

Black values indicate switches where population-tuning curves did not significantly deviate from zero. Amplitudes for switches were population-tuning curves deviated significantly are indicated in either red or blue, depending on the sign of the fitted functions. Positive values indicate horizontal tuning and are indicated in **red**; negative values indicate vertical tuning and are indicated in **blue**. Values were fitted for each experiment individually. For the 33-Hz data and the 10-Hz grating data, the fitted values are shown directly (n = 1 in both cases). For the 10-Hz data (derived from narrowly and broadly filtered natural images), averages and medians (in brackets) over experiments (n = 12) are shown, asterisks indicate results from a sign test against zero over n = 12 values. Histograms detail the individual values for the 12 experiments (blue side negative, red side positive amplitudes), corresponding to the last two rows of the above table.

33-Hz data: Amplitudes are negative (indicating vertical tuning) when vertical orientation is presented after the switch (columns 2 and 7) and positive (indicating horizontal tuning) when the orthogonal (horizontal) orientation is presented after the switch (columns 3 and 6). We confirm persistent tuning to previous orientation after a switch from a single orientation to blank for horizontal (column 4) and vertical orientation (column 8). When we examine the case where transition to the superposition led to significant differences from zero before the fitting, we find that the population tuning represents the vertical orientation presented before the switch and not the horizontal orientation that was turned on (column 1, row 1), most likely reflecting a delayed response during transition.

10-Hz data: Responses were tuned to a single orientation switched on from blank for both horizontal (column 3) and vertical orientation (column 7). Corresponding to the 33-Hz data, we find a persistent response to the previous orientation when switched-off to blank (off-component), for both horizontal (column 4) and vertical orientation (column 8). For switches from a single orientation to the superposition, positive amplitudes indicate horizontal population tuning when the horizontal

orientation was added (column 1), and negative amplitudes indicate vertical population tuning when the vertical orientation was added (column 5). For switches from the superposition to a single orientation, amplitudes indicate population tuning to the removed orientation when horizontal is turned-off (column 2) and when vertical is turned off (column 6). In general, we find that the sign of the fitted amplitude is consistent with population tuning to the orientation present after the switch for the 33-Hz data and representative of the difference between two successive stimuli for the 10-Hz data.

#	Description	Adaptive comp.	Off-comp.	df	X^2	X^2/ X^2 of Baseline	p-value
0	Baseline	0	0	863	1789	100%	<0.0001
1	Adaptive comp. alone	1	0	863	5101	285%	<0.0001
2	Off comp. alone	0	1	863	3283	183%	<0.0001
3	50-50-Average	0.5	0.5	863	1077	60%	<0.0001
4	Weighted Average	0.40± 0.03	0.60	839	909	51%	0.047

Table 3.S4: Response decomposition (results): contribution of adaptive and off-component to change response.

Results for 10-Hz sequences, 12 experiments. Columns describe the models, which are characterized by the respective weights given to adaptive component (i.e. response to sustained orientation, third column) and off-component (response to orientation that is turned-off, fourth column). For model #4, mean and *sem* across fitted weights for experiments and both filter conditions (broad and narrow) is given. The fifth column shows degrees of freedom, the sixth column the $\boldsymbol{X^2}$-statistic (see Materials and Methods). The seventh column shows the ratio of $\boldsymbol{X^2}$-values between the respective model and the baseline. The eighth column shows the p-value for data-model comparison.

Neither of the constituent responses alone (#1 and #2) can explain the data better than the baseline model ($\boldsymbol{X}^2$-values in #1 and #2 are higher than in #0), albeit the off-component alone is closer to the measured data than the adaptive component alone ($\boldsymbol{X}^2$-values in #2 are smaller than in #1). The 50-50-average of both components (#3) provides a better fit than either of the components alone (#1 and #2). When constituent responses are combined as weighted average, we obtain significantly higher weights for the off-component (than for the adaptive component (see #4; two-tailed t-test against 0.5, $p<0.01$). The prediction resulting from such a combination reduces the $\boldsymbol{X}^2$-values by 49%.

3.6.4 Supplementary methods

Response decomposition (methods): evaluation and fitting procedure

To investigate the response to the switch from the superposition to a single orientation (VH to V/H) at 10 Hz in more detail, we approximated response to this complex switch through the combination of the responses to its constituents, a sustained presentation of a single orientation (V to V or H to H), and the off-switch of the other orientation (H to blank or V to blank). The data used in our models are time-averaged (50–90 ms after switch) population tuning curves.

To statistically evaluate the goodness of the fit, we first tested the response from the complex switch (VH to H/V) against a flat response at zero to obtain a baseline. In the next step, we repeated the test on the residuals between the predictions of models 1 to 4 and the data.

As test statistics, we computed one $\boldsymbol{X}^2$-value across all experiments (e = 1,…,n; where n is the number of experiments), stimulus types (f = narrowly filter, broadly filtered naturals), orientation bins (θ = 1,…,18; bins of 10° each), and types of switches (c = 1,2; corresponding to VH to V and VH to H), taking the standard error over repetitions of a switch type (repetitions r_e = 1,…,m_e; where r_e is a repetition in experiment e, and m_e is the overall number of repetitions in experiment e), as a normalization factor:

$$\mathrm{X}^2 = \sum_e \sum_f \sum_\theta \sum_c \left(\frac{a\langle u_{e,f,\theta,c,r_e}\rangle_{r_e} + b\langle v_{e,f,\theta,c,r_e}\rangle_{r_e} - \langle d_{e,f,\theta,c,r_e}\rangle_{r_e}}{\frac{1}{\sqrt{m_e}}\sigma} \right)^2$$

$$\sigma = \sqrt{\frac{1}{m_e - 1} \sum_{r_e} \left(d_{e,f,\theta,c,r_e} - \langle d_{e,f,\theta,c,r_e}\rangle_{r_e} \right)^2}$$

Here, $\langle \quad \rangle_{r_e}$ denotes the average response over repetitions in experiment e, and d_{e,f,θ,c,r_e} is the response that we wanted to model in experiment e, filter condition f, orientation bin θ, switch type c, and repetition r_e. Respectively, $\langle u_{e,f,\theta,c,r_e}\rangle_{r_e}$ is the average of responses to the sustained orientation, and $\langle v_{e,f,\theta,c,r_e}\rangle_{r_e}$ is the average of off-response over all repetitions in experiment e, in filter condition f, orientation bin θ, for switch type c.

We investigated three different models and the baseline response (model 0). The models differ in their weights for the sustained response, denoted with a, and the weight for the off-response, denoted with b. For the baseline (test against constant zero), m_0: $a = b = 0$. In model 1 (only sustained response), m_1: $a = 1$; $b = 0$, in model 2 (only off-response), m_2: $a = 0$; $b = 1$, in model 3 (50-50-average) m_3: $a = 0.5$; $b = 0.5$. In model 4 parameters were fitted to the data using the constraint $a + b = 1$. We fitted a_{ef} for each experiment e and filter condition f by minimizing all ε_{ef}:

$$\varepsilon_{ef} \coloneqq \sqrt{\sum_\theta \sum_c \sum_{r_e} \left(a_{ef}\langle u_{e,f,\theta,c,r_e}\rangle_{r_e} + (1 - a_{ef})\langle v_{e,f,\theta,c,r_e}\rangle_{r_e} - d_{e,f,\theta,c,r_e} \right)^2}$$

This way, one set of weights was fitted to each experiment and filter condition, with data-model comparison at 864 points: 12 experiments x 2 filter conditions x 18 bins of orientation

preference x 2 versions of the switch. The models have different degrees of freedom, depending on the number of fitted parameters: $df_{nr} = 864 - 1 - p_{nr}$, with fitted parameters (p_{rn}): $p_0 = p_1 = p_2 = 0$, $p_3 = 24$.

3.7 References

Adesnik H, Scanziani M (2010) Lateral competition for cortical space by layer-specific horizontal circuits. Nature 464: 1155–1160.

Adesnik H, Bruns W, Taniguchi H, Huang ZJ, Scanziani M (2012) A neural circuit for spatial summation in visual cortex. Nature 490: 226–231.

Ahissar E, Arieli A (2001) Figuring space by time. Neuron 32: 185–201.

Ahissar E, Arieli A (2012) Seeing via miniature eye movements: a dynamic hypothesis for vision. Front Comput Neurosci 6: 89.

Allison T (1962) Recovery functions of somatosensory evoked responses in man. Electroencephalogr Clin Neurophysiol 14: 331–343.

Anderson JS, Lampl I, Gillespie D, Ferster D (2000) The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. Science 290: 1968–1971.

Attneave F (1954) Some informational aspects of visual perception. Psychol Rev 61: 183–193.

Bair W, Cavanaugh JR, Smith MA, Movshon JA (2002) The timing of response onset and offset in macaque visual neurons. J Neurosci 22: 3189–3205.

Belitski A, Gretton A, Magri C, Murayama Y, Montemurro MA, Logothetis NK, Panzeri S. (2008) Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. J Neurosci 28: 5696–5709.

Benucci A, Frazor RA, Carandini M (2007) Standing waves and traveling waves distinguish two circuits in visual cortex. Neuron 55:103–117.

Benucci A, Ringach DL, Carandini M (2009) Coding of stimulus sequences by population responses in visual cortex. Nat Neurosci 12: 1317–1324.

Berger T, Borgdorff A. Crochet S, Neubauer FB, Lefort S, Fauvet B, Ferezou I, Carleton A, Luscher H-R, Petersen CCH (2007) Combined voltage and calcium epifluorescence imaging in vitro and in vivo reveals subthreshold and suprathreshold dynamics of mouse barrel cortex. J Neurophysiol 97: 3751–3762.

Beste C, Wascher E, Güntürkün O, Dinse HR (2011) Improvement and impairment of visually guided behavior through LTP- and LTD-like exposure-based visual learning. Curr Biol. 21: 876–882.

Betsch B, Einhäuser W, Körding KP, König P (2004) The world from a cat's perspective – statistics of natural videos. Biol Cybern 90: 41–50.

Boerlin M, Dèneve S (2011) Spike-Based Population Coding and Working Memory. PLoS Comp Biol 7: 1–18.

Busse L, Wade AR, Carandini M (2009) Representation of concurrent stimuli by population activity in visual cortex. Neuron 64: 931–942.

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC (2005) Do we know what the early visual system does? J Neurosci 25: 10577–10597.

Chen Y, Geisler WS, Seidemann E (2006) Optimal decoding of correlated neural population responses in the primate visual cortex. Nat Neurosci 9: 1412–1420.

Chen Y, Palmer CR, Seidemann E (2012) The relationship between voltage-sensitive dye imaging signals and spiking activity of neural populations in primate V1. J Neurophysiol 107: 3281–3295.

Coltheart M (1980) Iconic memory and visible persistence. Percept Psychophys 27: 183–228.

Das A, Gilbert CD (1997) Distortions of visuotopic map match orientation singularities in primary visual cortex. Nature 387: 594–598.

David SV, Vinje WE, Gallant JL (2004) Natural stimulus statistics alter the receptive field structure of v1 neurons. J Neurosci 24: 6991–7006.
David SV, Gallant JL (2005) Predicting neuronal responses during natural vision. Network. 16: 239–260.
Desbordes G, Rucci M (2007) A model of the dynamics of retinal activity during natural visual fixation. Vis Neurosci 24: 217–230.
DiLollo V (1977) Temporal characteristics of iconic memory. Nature 267: 241–243.
Dinse HR, Jancke D (2001) Time-variant processing in V1: From microscopic (single cell) to mesoscopic (population) levels. Trends in Neurosci 24: 203–205.
Downar J, Crawley AP, Mikulis DJ, Davis KD (2000) A multimodal cortical network for the detection of changes in the sensory environment. Nat Neurosci 3: 277–283.
Dragoi V. Sharma J, Miller EK, Sur M (2002) Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. Nat Neurosci 5: 883–891.
Duysens J, Orban GA, Cremieux J, Maes H (1985) Visual cortical correlates of visible persistence. Vision Res 25: 171–178.
Duysens J, Schaafsma SJ, Orban GA (1996) Cortical off response tuning for stimulus duration. Vision Res. 36: 3243–3251.
Eagleman DM, Jacobson JE, Sejnowski TJ (2004) Perceived luminance depends on temporal context. Nature. 428: 854–856.
Eckhorn R, Krause F, Nelson JI (1993) The RF-cinematogram. A cross-correlation technique for mapping several visual receptive fields at once. Biol Cybern 69: 37–55.
Eriksson D, Tompa T, Roland PE (2008) Non-Linear Population Firing Rates and Voltage Sensitive Dye Signals in Visual Areas 17 and 18 to Short Duration Stimuli. PLoS ONE 3: e2673.
Eriksson D, Valentiniene S, Papaioannou S (2010) Relating Information, Encoding and Adaptation: Decoding the Population Firing Rate in Visual Areas 17/18 in Response to a Stimulus Transition. PLoS ONE 5: e10327.
Eriksson D, Wunderle T, Schmidt KE (2012) Visual cortex combines a stimulus and an error-like signal with a proportion that is dependent on time, space, and stimulus contrast. Front Syst Neurosci 6: 26.
Felsen G, Shen YS, Yao H, Spor G, Li C, Dan Y (2002) Dynamic modification of cortical orientation tuning mediated by recurrent connections. Neuron 36: 945–954.
Felsen G, Dan Y (2005) A natural approach to studying vision. Nat Neurosci 8: 1643–1646.
Felsen G, Touryan J, Han F, Dan Y (2005) Cortical sensitivity to visual features in natural scenes. PLoS Biol 3: e342.
Finn IM, Priebe NJ, Ferster D (2007) The emergence of contrastinvariant orientation tuning in simple cells of cat visual cortex. Neuron 54: 137–152.
Fournier J, Monier C, Pananceau M, Frégnac Y (2011) Adaptation of the simple or complex nature of V1 receptive fields to visual statistics. Nature Neurosci 14: 1053–1060.
Fowler G, Korn D, North S, Rao H, Vo K (1995) vdelta: Differencing and Compression. In: Krishnamurthy B, editor. Practical Reusable Unix Software. New York: John Wiley & Sons, Inc.
Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360: 815–836.
Gallant JL, Connor CE, Van Essen DC (1998) Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing. Neuroreport 9: 2153–2158.
Galaretta M, Hestrin S (1998) Frequency dependent synaptic depression and the balance of excitation and inhibition in the neocortex. Nat Neurosci 1: 587–594.
Garrido MI, Kilner J, Stephan KE, Friston KJ (2009) The mismatch negativity: A review of underlying mechanisms. Clin Neurophysiol. 120: 543–463.

Glasser DM, Tsui JMG, Pack CC, Tadin D (2011) Perceptual and neural consequences of rapid motion adaptation. Proc Natl Acad Sci USA 108: E1080–1088.
Geisler WS, Albrecht DG, Crane AM, Stern L (2001) Motion direction signals in the primary visual cortex of cat and monkey. Vis Neurosci 18: 501–516.
Geisler WS (2008) Visual Perception and the Statistical Properties of Natural Scenes. Annu Rev Psychol 59: 167–92.
Gilad A, Meirovithz E, Leshem A, Arieli A, Slovin H (2012) Collinear stimuli induce local and cross-areal coherence in the visual cortex of behaving monkeys. PLoS One 7: e49391.
Gillespie DC, Lampl I, Anderson JS, Ferster D (2001) Dynamics of the orientation-tuned membrane potential response in cat primary visual cortex. Nature Neurosci 4: 1014–1019.
Golomb JD, Nguyen-Phuc AY, Mazer JA, McCarthy G, Chun MM (2010) Attentional facilitation throughout human visual cortex lingers in retinotopic coordinates after eye movements. J Neurosci 30: 10493–10506.
Gould JD (1967) Pattern recognition and eye-movement parameters. Perception 2: 399–407.
Grabska-Barwinska A, Distler C, Hoffman K-P, Jancke D (2009) Contrast independence of cardinal preference: stable oblique effect in orientation maps of ferret visual cortex. Eur J Neurosci 29: 1258–1270.
Graf ABA, Kohn A, Jazayeri M, Movshon JA (2011) Decoding the activity of neuronal populations in macaque primary visual cortex. Nat Neurosci 14: 239–245.
Grinvald A, Lieke EE, Frostig RD, Hildesheim R (1994) Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. J Neurosci 14: 2545–2568.
Grinvald A, Hildesheim R (2004) VSDI: A new era in functional imaging of cortical dynamics. Nat Rev Neurosci 5: 874–885.
Haider B, Krause MR, Duque A, Yu Y, Touryan J, Mazer JA, McCormick DA (2010) Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. Neuron 65: 107–121.
Henning MH, Kerscher NJ, Funke K, Wörgötter F (2002) Stochastik resonance in visual cortical neurons: does the eye-tremor actually improve visual acuity? Neurocomp 44–46: 115–120.
Hickmott PW (2010) Synapses of horizontal connections in adult rat somatosensory cortex have different properties depending on the source of their axons. Cereb Cortex 20: 591–601.
Hirsch JA, Martinez LM, Pillai C, Alonso JM, Wang Q, Sommer FT (2003) Functionally distinct inhibitory neurons at the first stage of visual cortical processing. Nat Neurosci 6: 1300–1308.
Hosoya T, Baccus SA, Meister M (2005) Dynamic predictive coding by the retina. Nature 436: 71–77.
Hyvärinen A, Hurri J, Hoyer PO (2009) Natural Image Statistics. A probabilistic approach to early computational vision. Book published by Springer-Verlag.
Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin F-H, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW (2004) Human posterior auditory cortex gates novel sounds to consciousness. Proc Natl Acad Sci USA 101: 6809–6814.
Jancke D, Erlhagen W, Dinse HR, Akhavan M, Giese M, Steinhage A, Schöner G (1999) Parametric representation of retinal location: Neural population dynamics and interaction in cat visual cortex. J Neurosci 19: 9016–9028.
Jancke D (2000) Orientation formed by a spot's trajectory: a two-dimensional population approach in primary visual cortex. J Neurosci 20: 1–6.

Jancke D, Chavane F, Naaman S, Grinvald A (2004) Imaging cortical correlates of illusion in early visual cortex. Nature. 428: 423–426.
Jonides J, Irwin DE, Yantis S (1982) Integrating visual information from successive fixations. Science 215: 192–194.
Jurjut OF, Nikolic D. Singer W, Yu S, Havenith MN, Muresan RC (2011) Timescales of multineuronal activity patterns reflect temporal structure of visual stimuli. PLoS ONE 6: 1–15.
Kayser C, Einhäuser W, König P (2004) Temporal correlations of orientations in natural scenes. Neurocomputing. 52–54: 117–123.
Keller GB, Bonhoeffer T, Hübener M (2012) Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. Neuron. 74: 809–815.
Kuang X, Poletti M, Victor JD, Rucci M (2012) Temporal encoding of spatial information during active visual fixation. Current Biol. 22: 510–514.
Kyrazi HT, Carvell GE, Simons DJ (1994) Off response transformations in the whisker/barrel system. J Neurophysiol 72: 392–401.
Lee C, Rohrer WH, Sparks DL (1988) Population coding of saccadic eye movements by neurons in the superior colliculus. Nature 332: 357–360.
Lewis CM, Lazar AE (2013) Orienting towards ensembles: From single cells to neural populations. J Neurosci 33: 2–3.
MacEvoy SP, Tucker TR, Fitzpatrick D (2009) A precise form of divisive suppression supports population coding in the primary visual cortex. Nat Neurosci 12: 637–645.
Markram H, Wang Y, Tsodyks M (1998) Differential signaling via the same axon of neocortical pyramidal neurons. Proc Natl Acad Sci USA 95: 5323–5328.
McCormick DA, Shu Y, Hasenstaub A, Sanchez-Vives M, Badoual M, Bal T (2003) Persistent cortical activity: mechanisms of generation and effects on neuronal excitability. Cereb Cortex 13: 1219–1231.
Meirovithz E, Ayzenshtat I, Werner-Reiss U, Shamir I, Slovin H (2012) Spatiotemporal effects of microsaccades on population activity in the visual cortex of monkeys during fixation. Cereb Cortex 22: 294–307.
Moeller GU, Kayser C, Knecht F, König P (2004) Interactions between eye movement systems in cats and humans. Exp Brain Res 157: 215–224.
Movshon JA, Lennie P (1979) Pattern-selective adaptation in visual cortical neurones. Nature 278: 850–852.
Müller JR, Metha AB, Krauskopf J, Lennie P (1999) Rapid adaptation in visual cortex to the structure of images. Science 285: 1405–1408.
Näätänen R, Gaillard AWK, Mäntysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. Acta Psychol 42: 313–329.
Nelson SB (1991) Temporal interactions in the cat visual system. I. Orientation-selective suppression in visual cortex. J Neurosci 11: 344–356.
Niell CM, Stryker MP (2010) Modulation of visual responses by behavioral state in mouse visual cortex. Neuron 65: 472–479.
Nikolić D, Häusler S, Singer W, Maass W (2009) Distributed fading memory for stimulus properties in the primary visual cortex. PLoS Biol 7: e1000260.
Olsen SR, Bortone DS, Adesnik H, Scanziani M (2012) Gain control by layer six in cortical circuits of vision. Nature 483: 47–52.
Olshausen BA, Field DJ (2005) How close are we to understanding V1? Neural Comput 17: 1665–1699.
Omer DB, Hildesheim R, Grinvald A (2013) Temporally-structured acquisition of multidimensional optical imaging facilitates visualization of elusive cortical representations in the behaving monkey. Neuroimage 82: 237–251.

Onat S, Nortmann N, Rekauzke S, König P, Jancke D (2011a) Independent encoding of grating motion across stationary feature maps in primary visual cortex visualized with voltage-sensitive dye imaging. Neuroimage 55: 1763–1770.
Onat S, König P, Jancke D (2011b) Natural scene evoked population dynamics across cat primary visual cortex captured with voltage-sensitive dye imaging. Cereb Cortex 21: 2542–2554.
Parker DM, Lishman JR, Hughes J (1992) Temporal integration of spatially filtered visual images. Perception 21: 147–160.
Pernberg J, Jirmann K, Eysel UT (1998) Structure and dynamics of receptive fields in the visual cortex of the cat (area 18) and the influence of GABAergic inhibition. Eur J Neurosci 10: 3596–3606.
Petersen CC, Grinvald A, Sakmann B (2003) Spatiotemporal dynamics of sensory responses in layer 2/3 of rat barrel cortex measured in vivo by voltage-sensitive dye imaging combined with whole-cell voltage recordings and neuron reconstructions. J Neurosci 23: 1298–1309.
Pouget A, Dayan P, Zemel RS (2003) Inference and computation with population codes. Annu Rev Neurosci 26: 381–410.
Rao RPN, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. Nat Neurosci. 2:79-87.
Ringach DL, Hawken MJ, Shapley R (19979 Dynamics of orientation tuning in macaque primary visual cortex. Nature 387: 281–284.
Rockland KS, Pandya DN (1979) Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. Brain Res 179: 3–20.
Roland PE, Hanazawa A, Undeman C, Eriksson D, Tompa T, Nakamura H, Valentiniene S, Ahmed B (2006) Cortical feedback depolarization waves: a mechanism of top-down influence on early visual areas. Proc Natl Acad Sci USA 103: 12586–12591.
Rucci M, Desbordes G (2003) Contributions of fixational eye movements tot he discrimination of briefly presented stimuli. J Vis 3: 852–864.
Rucci M, Iovin R, Poletti M, Santini F (2007) Miniature eye movements enhance fine spatial detail. Nature 447: 852–855.
Rucci M (2008) Fixational eye movements, natural image statistics, and fine spatial vision. Network 19: 253–285.
Sanchez-Vives MV, Nowak LG, McCormick DA (2000) Cellular mechanisms of long-lasting adaptation in visual cortical neurons. J Neurosci 20: 4286–4299.
Scholl B, Gao X, Wehr M (2010) Nonoverlapping sets of synapses drive on responses and off responses in auditory cortex. Neuron 65: 412–421.
Sharon D, Grinvald A (2002) Dynamics and constancy in cortical spatiotemporal patterns of orientation processing. Science 295: 512.
Simoncelli EP, Farid H (1996) Steerable wedge filters for local orientation analysis. IEEE Trans. Image Proc 5: 1377–1382.
Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24: 1193–1216.
Sit YF, Chen Y, Geisler WS, Mikkulainen R, Seidemann E (2009) Complex dynamics of V1 population responses explained by a simple gain-control model. Neuron 24: 943–956.
Smyth D, Willmore B, Baker GE, Thompson ID, Tolhurst DJ (2003) The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. J Neurosci 23: 4746–4759.
Snodderly DM, Kagan I, Gur M (2001) Selective activation of visual cortex neurons by fixational eye movements: Implications for neural coding. Vis Neurosci 18: 259–277.
Sperling G (1960) The information available in brief visual presentations. Psychol Monogr 74: 1–29.

Spratling MW (2010) Predictive coding as a model of response properties in cortical area V1. J. Neurosci 30: 3531–3543.
Spratling MW (2012) Predicitve coding as a model of the V1 saliency map hypothesis. Neural Networks 26: 7–28.
Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding: a fresh view of inhibition in the retina. Proc R Soc Lond B 216: 427–459.
Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381: 520–522.
Tsodyks M, Kenet T, Grinvald A, Arieli A (1999) Linking spontaneous activity of single cortical neurons and the underlying functional architecture. Science 286: 1943–1946.
Varela JA, Song S, Turrigiano GG, Nelson SB (1999) Differential depression at excitatory and inhibitory synapses in visual cortex. J Neurosci 19: 4293–4304.
Vetter P, Grosbras MH, Muckli L (2013) TMS over V5 disrupts motion prediction. Cereb Cortex. Epub ahead of print. doi:10.1093/cercor/bht297.
Vogels R (1990) Population coding of stimulus orientation by striate cortical cells. Biol Cybern 64: 24–31.
Wehr M, Zador AM (2005) Synaptic mechanisms of forward suppression in rat auditory cortex. Neuron 47: 437–445.
Womelsdorf T, Lima B, Vinck M, Oostenveld R, Singer W, Neuenschwander S, Fries P. (2012) Orientation selectivity and noise correlation in awake monkey area V1 are modulated by the gamma cycle. Proc Natl Acad Sci USA 109: 4302–4307.
Zucker RS, Regehr WG (2002) Short-term synaptic plasticity. Annu Rev Physiol 64: 355–405.

4. Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention (Study 3)

This section is based on the following publication in PLoS Compuational Biology: S Kollmorgen[12], N Nortmann[1*], S Schröder[12*], P König[1] (2010) Influence of Low-Level Stimulus Features, Task Dependent Factors, and Spatial Biases on Overt Visual Attention. PLoS Comp Biol 6: 1-20.* *[1]Institute of Neurobiopsychology, University of Osnabrück, Osnabrück, Germany. [2]Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich, Switzerland. *These authors contributed equally to this work.*

4.1 Abstract

Visual attention is thought to be driven by the interplay between low-level visual features and task dependent information content of local image regions, as well as by spatial viewing biases. Though dependent on experimental paradigms and model assumptions, this idea has given rise to varying claims that either bottom-up or top-down mechanisms dominate visual attention. To contribute toward a resolution of this discussion, here we quantify the influence of these factors and their relative importance in a set of classification tasks. Our stimuli consist of individual image patches (bubbles). For each bubble we derive three measures: a measure of salience based on low-level stimulus features, a measure of salience based on the task dependent information content derived from our subjects' classification responses and a measure of salience based on spatial viewing biases. Furthermore, we measure the empirical salience of each bubble based on our subjects' measured eye gazes thus characterizing the overt visual attention each bubble receives. A multivariate linear model relates the three salience measures to overt visual attention. It reveals that all three salience measures contribute significantly. The effect of spatial viewing biases is highest and rather constant in different tasks. The contribution of task dependent information is a close runner-up. Specifically, in a standardized task of judging facial expressions it scores highly. The contribution of low-level features is, on average, somewhat lower. However, in a prototypical search task, without an available template, it makes a strong contribution on par with the two other measures. Finally, the contributions of the three factors are only slightly redundant, and the semi-partial correlation coefficients are only slightly lower than the coefficients for full correlations. These data provide evidence that all three measures make significant and independent contributions and that none can be neglected in a model of human overt visual attention.

4.2 Introduction

In daily life, eye movements center parts of a scene on the human fovea several times a second (Buswell, 1935). The part of the visual field falling onto the fovea is represented with the highest spatial acuity and, compared to the periphery, receives disproportionately more cortical processing resources (Tootell et al., 1982). The selection process is an important aspect of attention, and it has a profound impact on our perception (Rizzolatti et al., 1987). The selection of fixation points is governed by several factors. First, goal-driven, top-down mechanisms adapt eye movements to the specific task (Yarbus, 1967; Land et al., 1999). Second, bottom-up mechanisms that consider only sensory-driven aspects, such as local image features (Koch & Ullman, 1985), contribute to the fixation selection process. Third, characteristics inherent to the visual apparatus, such as the spatial bias to the center region (Tatler, 2007) and geometric properties of saccades (Brockmann & Geisel, 1999), are widely acknowledged to influence the selection of fixation points. However, the relative roles and the

interaction of these mechanisms are not understood, and a quantitative understanding of the principles of fixation selection is still lacking.

Attention models designed to cope with the complexities of natural conditions are usually based on a so-called salience map (Koch & Ullman, 1985). Filtering the input image with kernels reminiscent of early visual processing generates feature maps at various spatial scales. These are then combined into a single salience map, which encodes the probability that an image region will be attended (Itti & Koch, 2001). In principle, the selection of features for such models is unconstrained. First implementations were designed to explain covert attention in experiments involving artificial stimuli and based on a repertoire of simple features. Present models slowly move towards a more complete list of relevant features (Wolfe & Horowitz, 2004) and include more and more features (Betz et al., 2010). Furthermore, they introduce probabilistic and decision theoretic concepts (Zhang et al., 2008; Gao & Vasconcelos, 2009; Itti & Baldi, 2009). Salience maps predict, to some extent, fixations in complex scenes (Privitera & Stark, 2000; Li, 2002; Parkhurst et al., 2002; Peters et al., 2005; Tatler et al., 2005) for humans as well as for monkeys (Einhäuser et al., 2006). The critical phrase "to some extent" is at the center of an intense debate. Is it possible to refine models based on stimulus dependent salience to model overt attention as well as intersubject variability allows?

A major concern is that even if features of the salience map, such as luminance contrast, are good correlates of fixation probability, they do not necessarily drive attention causally (Einhäuser & König, 2003; Carmi & Itti, 2006; Tatler, 2007), but are contingent on higher-order statistics (Einhäuser et al., 2006). These issues have raised considerable skepticism regarding models based purely on low-level image features.

For these reasons, there is consensus that viable models of human attention should not rely solely on stimulus properties. Specifically, eye movements are influenced by spatial constraints and properties of the oculomotor system. A wide range of studies has demonstrated a preponderance of small amplitude saccades (Bahill et al., 1975). Furthermore, under typical lab conditions observers have a central bias—i.e., a tendency to fixate preferentially close to the center of photographs of natural scenes, in excess of behavior under truly natural conditions (Tatler et al., 2006; Schuhmann et al., 2008). Furthermore, the recent years have seen a major advance in our understanding of scene layout. Including such information, which was automatically generated by machine learning algorithms, leads to a very high prediction accuracy in a search task for pedestrians (Ehinger et al., 2009). Furthermore, recent work demonstrates that spatial properties might have a large influence on overt attention (Tatler & Vincent, 2009). While it is clear that these spatial factors contribute to the selection of fixation points, there is as yet no quantification of their general influence.

That the task context influences eye movements has long been observed (Buswell, 1935; Yarbus, 1967). In a study utilizing a variety of tasks—including abstract interpretations, such as the judgment of social status—the task was found to strikingly modify observed fixation patterns (Yarbus, 1967). Also the complex activities of daily living reveal the task dependence of human eye movements (Land & Hayhoe, 2001). Models for visual attention based solely on low-level visual features fail to capture the effects of the task context. Several extensions to existing and also new models have been proposed to address that shortcoming (Tsotsos et al., 1995; Torralba, 2003; Navalpakkam & Itti, 2007). An elegant information theoretic approach combines visual appearance, spatial information and high-level information further improving prediction accuracy (Kanan & Cottrell, 2010).

It was suggested early on that in a specific task context, the information content of an image patch defines its salience (Antes, 1974; Mackworth & Morandi, 1967). This proposal has triggered bottom-up driven models of attention incorporating a decision theoretic approach (Zhang et al., 2008; Gao & Vasconcelos, 2009; Itti & Baldi, 2009). Hence the information content of a patch may be viewed as a task dependent, high-level feature. This view is suited to reconciling stimulus-driven models and task-centered models. Recent studies emphasizing the importance of objects in overt attention are compatible with this view (Einhäuser et al., 2008). However, information content is determined either intuitively or based on a direct subjective rating. Furthermore, there is presently no general algorithm available that would reliably label objects in a visual scene. Instead studies rely again on ratings of human subjects (http://labelme.csail.mit.edu/). An explicit quantification of the contribution of task dependent factors relative to feature-based factors is still missing.

In summary, it is widely acknowledged that image features, the task of the observer, and the properties of the oculomotor system contribute to the selection of fixation points. Still, in the absence of quantitative data on the relative weight of the different factors, settling the issue of how exactly each of these contributes towards overt attention is not possible. In the present study, we attempt a step in this direction: we quantify the relative contribution of stimulus properties, task dependence, and oculomotor constraints to the selection of fixation points. We capture stimulus dependent properties by an analysis of low-level image features. Task dependent factors are captured by the information content of discrete parts of the stimulus in well defined tasks. The influence of oculomotor constraints is taken into account by a generative model including typical saccadic parameters and the central bias. With this approach we obtain scalars for each of these three factors for each image region, allowing us to quantify their independent contributions to human eye movements.

To quantify the three different types of influences we sample non-overlapping image patches (bubbles) from forest scenes and face images. These isolated patches are shown in different configurations in combination with a classification task. This design is inspired by Gosselin and Schyns, who have introduced the bubble paradigm to measure which parts of an image are used by the observer to solve a classification task (Gosselin & Schyns, 2001; Gosselin & Schyns, 2002; Schyns et al., 2002; Vinette et al., 2004). The technique applies two-dimensional Gaussian filters to isolate locations of visual cues, which are called bubbles. These are then presented in varying combinations, revealing only a limited controlled subset of the image content in combination with a classification task. Based on the observers' responses, Gosselin and Schyns derived a map revealing the regions of an image that are relevant for a specific classification task (Gosselin & Schyns, 2001). We use the bubble technique in combination with an eye-tracking experiment to obtain measures of different contributions to overt attention. Each bubble is treated as an independent unit. Utilizing recorded eye movements, responses in the classification task, feature analysis of the image patches, and baseline data taken from a free viewing eye-tracking study, we compute four measures: the stimulus dependent measure captures low-level feature contrast and is based on the luminance and texture distribution within each bubble. The task-related measure ignores image features, but quantifies how much information a bubble contains in the context of a specific classification task. Additional high-level factors, e.g. emotional and attentional state, might be relevant. We tried to keep these constant as much as possible. This quantification does of course not capture all possible top-down effects discussed in the literature as a classification task provides a particular context. The third measure, describing the spatial characteristics of eye movements, builds on a baseline study and takes into consideration the global fixation bias and geometrical properties of saccades. By evaluating the eye-tracking

data of the main study, we obtain the fourth measure that captures the empirical salience of each bubble.

In comparison to full field stimuli, our bubble stimuli consist of a manageable number of discrete perceptual units. Using discrete units allows us to assign a single value for each of the measures to each bubble. In particular, describing the task dependent information of a bubble using the degree of agreement between subjects with respect to a classification task requires individual pieces of visual information. It is not clear how a measurement of local information content could be achieved using full field stimuli only. Accordingly, the problem of measuring local information is exactly the one addressed when the bubbles technique was first established (Gosselin & Schyns, 2001).

Having acquired the four measures for each bubble, we finally use linear multivariate regression to quantify the overall and the individual, i.e., non-redundant, contributions of the task-dependent, feature-based, and spatial-based factors influencing attention.

4.3 Results

4.3.1 Context

In this study, subjects had to classify visual stimuli based either on face images or on forest scenes. We employed a total of four different tasks. Face stimuli had to be classified either according to ***gender*** or according to facial ***expression***, with the stimulus **classes** *happy*, *sad*, *fearful*, or *disgusted*. For the stimuli based on forest scenes, one task (***space***) was to decide whether the scenery was *close and narrow* – when the image was a close-up or displayed a closed environment – or whether it was *wide and open*. The other task was to judge the presence of indicators of human ***influence*** such as houses, roads, paths, trunks of trees. Stimulus presentations lasted for three seconds during which the subjects' eye movements were recorded. The majority of the stimuli were composed of 1 to 5 bubbles placed on a gray background. Half of the stimuli consisted of bubbles originating from the same full field image (condition ***same***), whereas 15% of the stimuli combined bubbles from different full field images belonging to the same stimulus class (condition ***congruent***). Another 15% of the stimuli were composed of bubbles originating from full field images of different classes (condition ***incongruent***). To control for position effects, we also showed stimuli (16%) in which the positions of the bubbles are shuffled (condition ***permuted***). The remaining 4% of the stimuli were full field images, which we used to confirm that subjects agreed on the classes of the images underlying the bubble stimuli. The bubbles themselves were constructed from square image patches with a side length of 6 visual degrees. To each patch, we applied a space-variant filter to imitate the retinal resolution when fixating the center of the bubble and a Gaussian mask to avoid visible edges.

75 subjects took part in this study, each performed 280 trials. We used a total of 2061 gray-scale stimuli for all subjects. This resulted in 21000 trials, recorded with 131935 fixations.

4.3.2 Bubbles are treated as units

In a first step, we investigated viewing behavior relative to bubbles. Subjects made, on average, 6.2 fixations in each trial where bubbles were presented. Of these, 94% were no more than 3 visual degrees distant from the closest bubble center and thus were located well inside a bubble. Three percent were located at the screen center and can be attributed to anticipation of the decision screen that followed each trial. The remaining 3% were scattered

across the screen. Hence, the fixations were rare in the space between bubbles and were clearly targeted at bubbles.

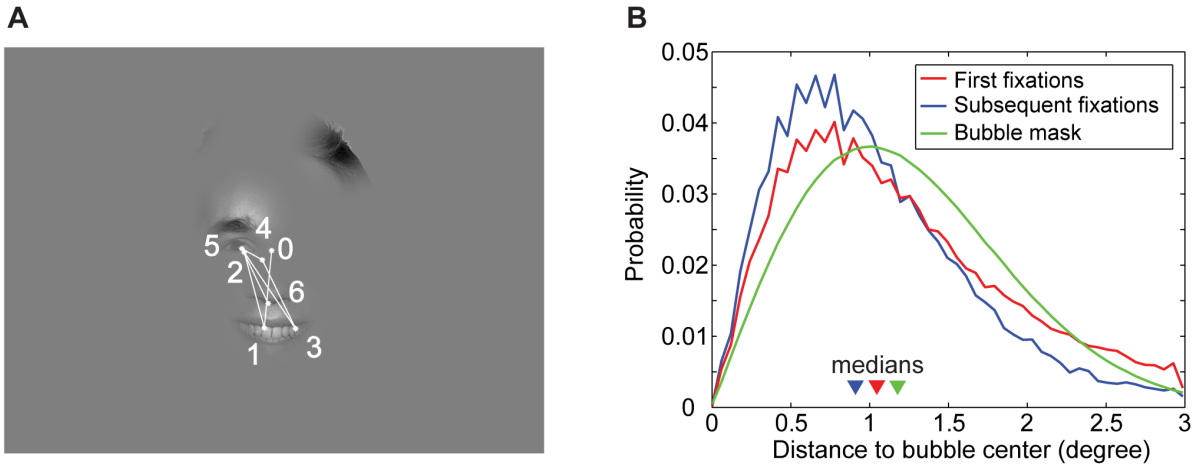


Figure 4.1 Fixations on bubbles.
(A) An example trajectory recorded during the experiment. The fixation labeled with zero is the first fixation in that trial, which was excluded from analyses. (B) Distributions of distances between fixation locations and the closest bubble center for "first fixations" into a bubble (median 1.05°) and "subsequent fixations" within the same bubble (median 0.91°). For comparison, the distribution that would result if all fixations were sampled from the Gaussian window used to construct the bubbles (median 1.18°) is also given.

We designed the bubbles in such a way that maximal and complete information is available when subjects fixated the center of the bubble (see Methods). Hence subjects would not gain anything by scanning different positions within the same bubble. This, however, does not necessarily prevent them from doing so. We tested this to confirm that bubbles were indeed treated as perceptual units. Of the total number of fixations that targeted bubbles, 60% originated from outside the respective bubble ("first fixations"). The remaining 40% were due to saccades within a bubble ("subsequent fixations"). The distributions of distances to bubble centers for these two groups of fixations were significantly different ($p<0.01$, KS-test, Fig. 4.1). The median of the distances to the closest bubble center was 1.05° for "first fixations" and larger than the median of 0.91° for "subsequent fixations". For pairs of first and subsequent fixations, the subsequent fixation was, on average, 0.16° closer to the bubble center. Additionally, both distributions were more sharply peaked than the distribution that would have resulted if fixations had been sampled from the Gaussian mask used for bubble construction ($p<0.01$ in both cases, KS-test, see Fig. 4.1). Altogether, the fixation data do not indicate that individual bubbles were scanned for information, but suggest that participants targeted bubble centers and made small corrective saccades towards bubble centers when landing off-center. The data is hence consistent with the assumption that bubbles were treated as perceptual units.

Building on the property that bubbles are treated as units, we derive a measure characterizing the empirical salience of a complete bubble. It is based on the fixation counts of a bubble in specific stimulus configurations (see below). In the above example of Fig. 4.1 these fixation counts amount to 3, 3, 0, and 0 for bubbles A, B, C, and D, respectively. These counts are then averaged over subjects.

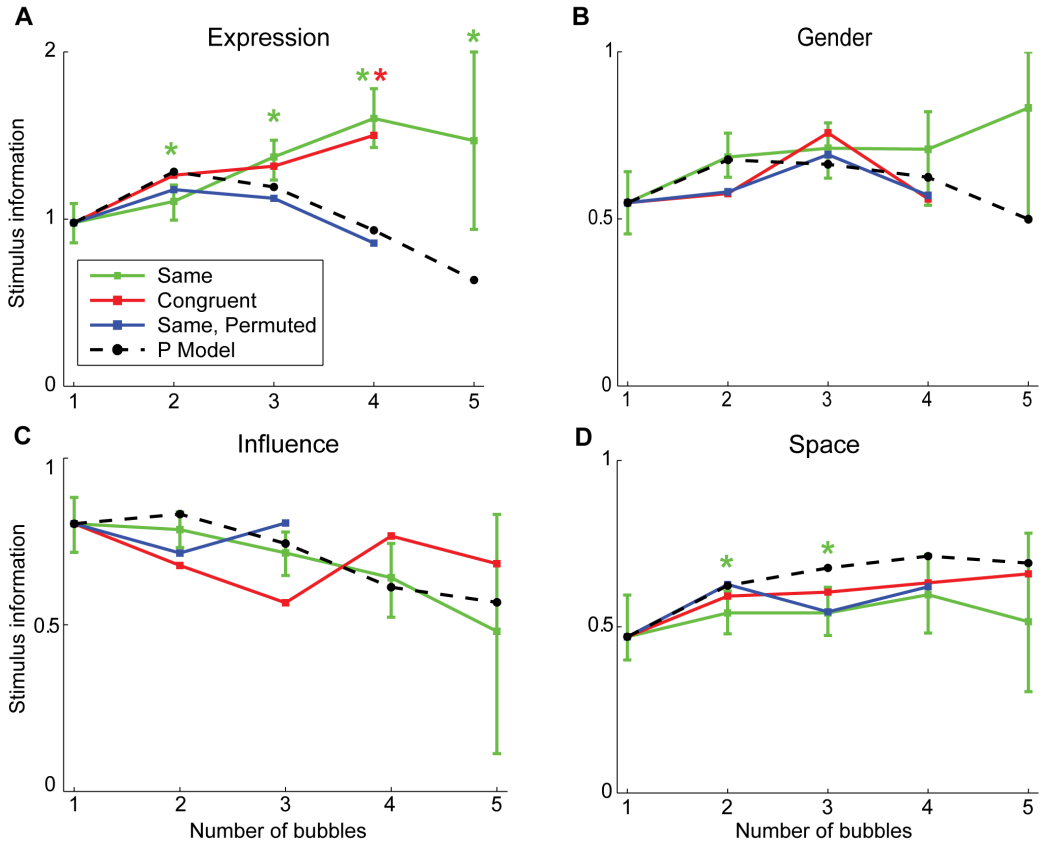


Figure 4.2 Stimulus information versus number of bubbles for the four tasks.
Stimulus information estimated using the p-model is plotted for all four tasks (black dashed line). This is contrasted with the measured stimulus information in the *same* condition (green line) and in the *congruent* condition (red line). The blue line marks the measurements that result if the positions of bubble stimuli of the *same* condition are shuffled (*same, permuted*). The colored stars mark significant differences ($p<0.05$, bootstrapped confidence intervals) between the curve belonging to the respective condition and the p-model estimate. For visibility, the 95% confidence interval is marked by error bars only for condition *same*.

4.3.3 Information of different bubbles is integrated

The average classification performance for the original images (full field stimuli) was 94%, as measured by the fraction of responses that were correct with respect to the image class established in pre-experiments. For the four tasks average performance was 87%, 94%, 99% and 94% (*expression*, *gender*, *influence,* and *space*). When comparing the different tasks, please note that in *expression* the chance level is 25% and in all others 50%. The high-level of performance indicates that participants understood the tasks and had a shared interpretation of stimulus classes.

In order to be independent of predefined "correct" responses in the following analyses, we use the more general measure of *stimulus information*. It is defined by the maximal possible entropy of the distribution of responses minus the entropy of the actual response distribution (see Methods). When all subjects agree on the classification of a stimulus, that stimulus contains maximal information with respect to the classification tasks. When their response distribution is flat, the stimulus contains no information. In the case of *expression,* with 4 choices the *stimulus information* ranges from 0 to 2 bit, in the other tasks from 0 to 1 bit. *Stimulus information* thus captures the degree of consensus from the subjects classifying the stimuli.

Next we investigated stimulus information in the reduced stimuli composed of bubbles. Presenting bubble stimuli composed of bubbles taken from the same base image (condition *same*) yielded average stimulus information of 1.18 bit, 0.66 bit, 0.74 bit, and 0.54 bit for the four tasks *expression, gender*, *influence,* and *space*, respectively. Presenting stimuli composed of bubbles taken from different images of the same response class (condition *congruent*) average stimulus information changed to 1.31 bit, 0.62 bit, 0.67 bit, and 0.61 bit (*expression, gender*, *influence,* and *space*). In contrast, in presenting stimuli composed of bubbles taken from images of different response classes (condition *incongruent*) it dropped to 1.12 bit, 0.55 bit, 0.58 bit, and 0.34 bit. These data demonstrate that stimulus information is far from complete and that no ceiling effects are to be expected.

To address the integration of information we analyze stimulus information as a function of the number of bubbles (Fig. 4.2). First, we compare measured stimulus information in the *same* condition with estimates of a probabilistic model of information integration (see Methods). The model, which we denote as **p-model**, integrates the response distributions of individual bubbles to estimate the stimulus response distribution and is presented here as a hypothesis. In the following, we only test plausibility of the p-model; we give a more detailed account in the Discussion. Stimulus information is computed from the entropy of the stimulus response distribution as described above. The p-model assumes independence of the information in different bubbles and integrates the information optimally. To predict stimulus information as a function of the number of bubbles, a sample of the bubbles, which were presented on their own, is selected. Then the respective response distributions of these stimuli are integrated using the p-model. This procedure is repeated 1000 times for each number of bubbles and each task. The resulting average information values are compared to the empirically found information values of the stimuli containing the respective numbers of bubbles (Fig. 4.2). The selection of single bubble stimuli for integration is done independent of image class. In the *expression* task, which uses face stimuli, we observe a pronounced surplus of experimentally observed average stimulus information (green line) compared to the prediction of the p-model (dashed black line). This higher-than-expected stimulus information indicates a violation of the assumption of independence of the information in different bubbles and is investigated below. In the gender task, which also uses face stimuli, at four and five bubbles a surplus of measured information is observed as well. Due to the larger variance of these two data points it does not reach significance. Stimulus information in the *influence* task, which uses natural scenes, is well predicted by the p-model, and no significant deviation of estimate and data could be detected ($p>0.05$, bootstrapped confidence intervals). For *space*, stimulus information is a little, but significantly, smaller than predicted by the p-model ($p<0.05$, bootstrapped confidence intervals). In this condition, the integration strategy of the subjects does not quite reach optimal performance. These data suggest that the p-model provides a reasonable description of the information integration. The mentioned deviations are further investigated below.

Now, we investigate the integration of information for the different conditions. We compare stimulus conditions *same* and *congruent*. In contrast to the former, the latter is composed of bubbles that originate from different full field images of the same response class. Data obtained in *same* and *congruent* conditions give rise to nearly identical values of stimulus information in all tasks, and their difference is never significant ($p>0.05$, bootstrapped confidence intervals, Fig. 4.2 green and red lines). Specifically, this includes the large deviation from the prediction of the p-model in the *expression* task. This indicates that the information of bubbles is integrated in the same way, irrespective of whether the bubbles originate from the *same* or different *congruent* full field stimuli.

To further elucidate the cause for the deviations of the data from the p-model estimates, we consider the interaction of bubble information and spatial location. For this purpose, we employ permuted stimuli. These are composed of bubbles placed at positions not matching their location in the respective full field stimuli (see Methods). In all tasks, including the *expression* task, the stimulus information in this *permuted* condition is well predicted by the p-model (Fig. 4.2A, blue line). For the face stimuli, this, together with the large differences between the p-model and the *same* and the *congruent* condition for high numbers of bubbles, demonstrates that the subjects' integration of information is influenced by bubble locations. This can be understood intuitively if one assumes that bubbles at certain locations (e.g., mouth) are given more weight, irrespective of the actual content of the bubble (e.g., smile). Indeed, the main result for permuted stimuli is an improved fit by the p-model. On the other hand, position effects are not a likely cause for the deviations in the space task. There, the *permuted* and *same* conditions show no pronounced differences. The stimulus information for both is slightly below that of the p-model.

To test more directly whether bubble position and arrangement have an influence on information integration in the tasks *gender, influence,* and *space*, we performed an additional test and considered the differences between the response distributions of normal stimuli and their permuted versions. To specify whether these differences reflect a significant effect of permutation, we investigate whether the differences are consistent with the assumption that the responses for normal and permuted stimuli are sampled from the same stimulus answer distribution, independent of bubble arrangement. As the overwhelming majority (98.6%) of the differences between permuted and non permuted stimuli is located within the 95% confidence region of the zero hypothesis, no significant effect of permutation could be detected. It must be noted, however, that the test power is limited by the small number of trials using permuted stimuli.

We arrive at the conclusion that the p-model provides a good description of integration of information for face stimuli in the permuted condition and for forest scenes in all conditions. In the *same* and the *congruent* condition, face stimuli consisting of many bubbles are processed using additional configural information (Calder et al., 2000).

4.3.4 Three different saliences of bubbles and their relation to fixation behavior

Now we address the relative contributions to fixation behavior of the stimulus dependent salience, task dependent information, and geometric properties of the stimuli. First, to address the stimulus dependent salience, we consider the low-level visual information of luminance contrast and texture contrast. These features are presumably processed in a bottom-up manner and have been used in other studies before. Second, to address the task dependent information, we consider the measure of bubble information, which captures the contribution of a bubble to the classification responses of subjects (see Methods). Third, to address the geometric properties, we investigate whether a simple generative model of fixation behavior that is based on the spatial arrangement of bubbles, central fixation bias, and geometrical constraints of average saccadic length and direction is informative with respect to the frequencies of fixation of different bubbles. Finally, these three components are used to explain the empirical distribution of fixations on bubbles, represented by empirical saliences. The measure of empirical salience is a context independent measure that represents how often a bubble is fixated relative to any other bubble. To obtain a measure which is independent of the specific stimulus context (instead of values for each stimulus) we combined the data from all stimuli and computed the best linear fit (see Methods). With this measure in turn the actual averaged fixation counts on the individual stimuli can be reconstructed with an average accuracy of 94.4%. Hence the empirical salience gives a faithful description of the fixation

probability of a bubble in all stimulus configurations. The three former components and their relation with empirical salience are now considered in turn.

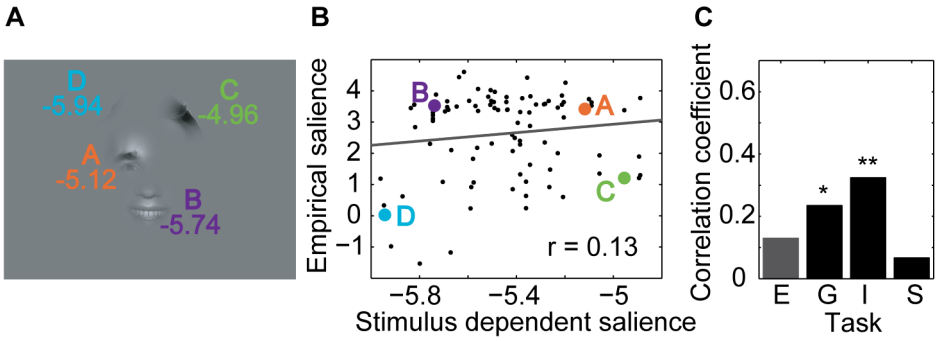


Figure 4.3 Relationship between stimulus dependent and empirical salience.
(A) Example stimulus from the expression task with bubbles labeled by their stimulus dependent salience. (B) Scatter plot of stimulus dependent vs. empirical salience for the *expression* task. The positions of the bubbles from the example stimulus are marked by colored dots. The correlation coefficient r is given as a figure inset. (C) Correlation coefficients for all four tasks (E – *expression*, G – *gender*, I – *influence*, S – *space*). One star marks a significant correlation ($p<0.05$, t-test); two stars mark a highly significant correlation ($p<0.01$, t-test).

4.3.5 Correlation of low-level stimulus features with empirical salience

In agreement with a large body of previous research (Reinagel & Zador, 1999; Einhäuser & König, 2003; Parkhurst & Niebur, 2004; Zhang et al., 2008; Açık et al., 2009; Gao & Vasconcelos, 2009), we characterize low-level visual information contained in a bubble by its luminance and texture contrast. We estimate the contribution to fixation behavior by considering fixation probability conditioned on these feature contrasts. This allows determining the correlation of local features, as used in common stimulus-driven models of overt attention, with the empirical salience of bubbles.

The luminance and texture contrast of bubbles are determined by standard procedures (see Methods). To infer the conditional fixation probability, we recur to a previous study where gaze movements on full field images have been recorded, and the conditional probability to fixate a location given its feature values was determined empirically (Schumann et al., 2007). Here we use the same procedure and the results of the previous study to convert both luminance contrasts and texture contrasts into fixation probabilities. To obtain a model that incorporates both, we combine the resulting probabilities, assuming independence of the contributions of the two feature contrasts. Fig. 4.3A shows an example stimulus from the *expression* task with the individual bubbles labeled with their stimulus dependent salience. Bubble A, located on the right eye and eyebrow, contains high luminance and texture contrasts. This is mapped to a high value of the stimulus dependent salience (see Methods). Relative to the other bubbles of the *expression* task, bubble A has a high stimulus dependent salience and a high empirical salience, placing it in the upper right-hand corner of the scatter plot of stimulus dependent salience vs. empirical salience (Fig. 4.3B). Bubble B, centered on the upper lip, has a lower stimulus dependent salience, but is looked at slightly more often than bubble A, placing it in the upper left-hand corner of the scatter plot. Bubble C, showing hair, has the strongest stimulus dependent salience of all four bubbles, but is only rarely looked at, placing it in the lower right corner of the plot. Bubble D, also showing hair, has very low stimulus dependent and empirical salience, placing it in the lower left corner of the plot. In this specific example, stimulus dependent salience and empirical salience appear unrelated.

To determine the predictive power of the feature-driven model, we correlate the predicted fixation probabilities for individual bubbles with their empirical salience (both log transformed, see Methods). 4.3B shows a scatter plot of stimulus dependent salience of all bubbles in the *expression* task versus their empirical salience. It shows a weak, albeit not significant, correlation ($p>0.05$, t-test). Similarly, no significant correlation is observed for the *space* task (Fig. 4.3C). In the remaining two tasks, *gender* and *influence*, we do observe a significant correlation. This shows that the strength of the correlation of low-level features with selected fixation points varies as a function of the task for photographs of faces as well as of natural environments.

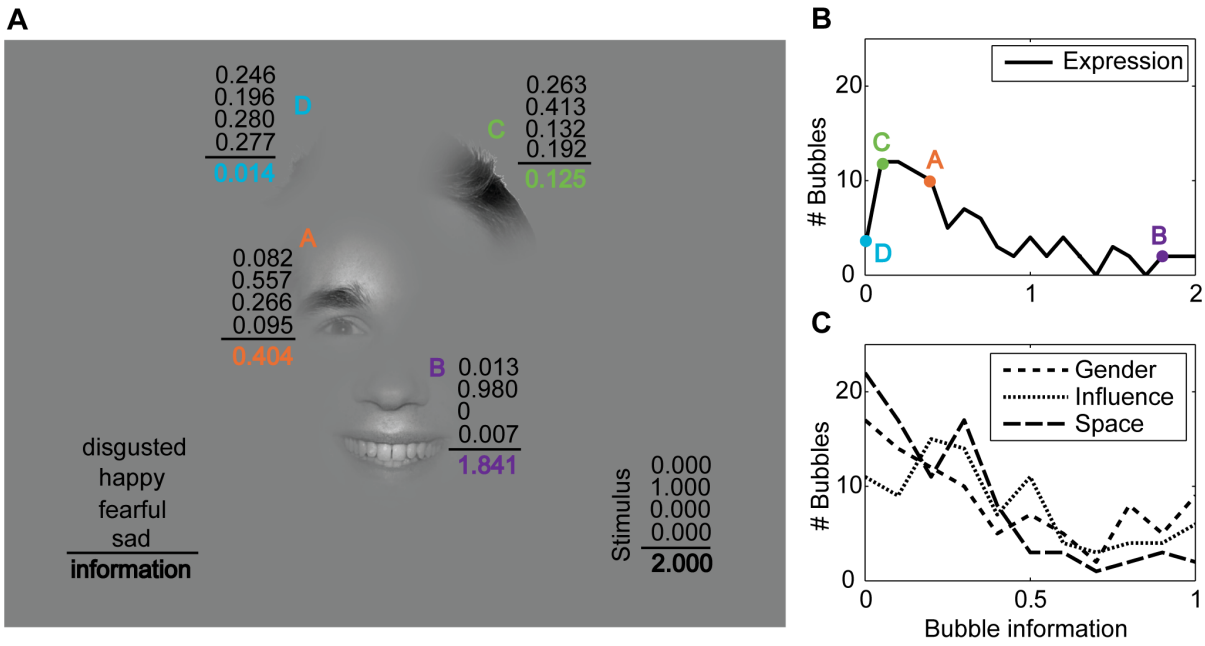


Figure 4.4 Bubble Information.
(A) Example stimulus from the *expression* task where the individual bubbles are labeled by their fitted response distributions and the corresponding bubble information. The four numbers above the black line give the response probabilities for the classes "disgusted," "happy," "fearful," and "sad." The bold number below the line gives the bubble information (in bit). For the whole stimulus, the measured response distribution and stimulus information (in bit) is given in the lower right corner. (B) The distribution of bubble information for the *expression* task. The bubble information of the four bubbles of the example stimulus is marked by colored dots. (C) The distribution of bubble information for the other three tasks *gender*, *influence* and *space*.

4.3.6 Correlation of bubble information with fixated bubbles

We take the contribution of a bubble to *stimulus information* as a surrogate for high-level information. We estimate bubble information for all bubbles that were shown in isolation or in combinations by assuming that the information of individual bubbles in a stimulus is integrated according to the p-model. Under this assumption, bubble information can be estimated in a global fit that maximizes the agreement between the subjects' responses to all stimuli and the alleged information contained in each single bubble (see Methods). This global fit estimates the information contained in each bubble, including those that were shown in isolation.

As a model of information integration we use the p-model introduced above. The results of the global fit based on the p-model may be viewed as a high-level feature specific to the context of the current task. Fig. 4.4A shows an example of a stimulus of the *expression* task

where the total measured stimulus information is 2 bit. The individual bubbles are labeled with their fitted response distributions and bubble information values. The global fit gives the information content as 0.40, 1.84, 0.13, and 0.01 bit for bubbles A, B, C, and D, respectively. In turn, estimating the stimulus information by the p-model results in 1.97 bit. This is close to the measured stimulus information with an error of 0.03 bit. Over all the bubble stimuli, the mean errors of stimulus information predicted from the fitted bubble answer distributions are 0.32, 0.20, 0.26, and 0.24 bit for the tasks *expression*, *gender*, *influence*, and *space*, respectively. For comparison, we computed the errors that would be expected if the predictions by the p-model were the true underlying response distributions of the stimuli (see supplementary text in section 4.6.2). In that case, the subjects sample their responses from the predicted response distributions and the resulting average errors serve as lower bounds for the expected errors. The resulting errors are 0.29, 0.16, 0.16, and 0.18 bit (*expression*, *gender*, *influence* and *space*). This implies that the deviation from the p-model stays within a factor of 2 of the theoretical lower limit and is consistent with the observation above that the p-model faithfully describes the dependence of stimulus information on the number of bubbles (Fig. 4.2). Hence, bubble information is reliably estimated by the global fit with the p-model. Fig. 4.4B and C show the frequencies of bubble information for the four tasks. The majority of bubbles have low bubble information values. Only a few have very high information. Bubble information varies over the whole possible range in all four tasks.

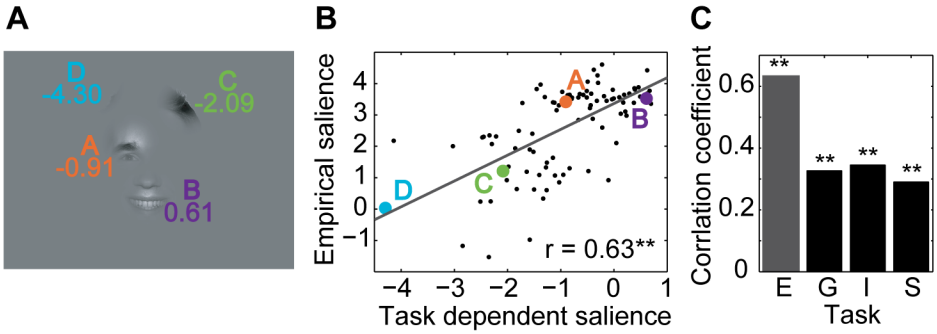


Figure 4.5 Relationship of task dependent and empirical salience.
(A) Example stimulus of the *expression* task with individual bubbles labeled by their bubble information. (B) Scatter plot of bubble information and empirical salience for the *expression* task. The positions of the example bubbles are marked by colored dots. The correlation coefficient r is given as a figure inset. (C) Correlation coefficients for all four tasks. Two stars mark highly significant correlations ($p<0.01$, t-test).

We now investigate the correlation between bubble information and empirical salience (both log transformed, see Methods). Fig. 4.5A shows the example stimulus with the individual bubbles labeled by their bubble information, and Fig. 4.5B shows a scatter plot of bubble information and empirical salience for the *expression* task. Bubble A, located on the right eye, is somewhat informative and looked at very often, placing it in the upper right corner of the plot. Bubble B, located on the smiling mouth, is much more informative than A but looked at only slightly more often, placing it in the upper right corner of the plot, to the right of bubble A. Bubble C, showing hair, has less information than A and B but is still somewhat informative. It is looked at less often than A and B. Bubble D, finally, has almost no information and is also looked at very seldom. In this specific example, bubble information and empirical salience are closely related.

Investigating the complete set of bubbles, we find that for the *expression* task the correlation of bubble information and empirical salience is highly significant ($p<0.01$, t-test). Although there is a noticeable drop in correlation for the tasks *gender*, *influence,* and *space*; all are

highly significant ($p<0.01$, t-test) as well (Fig. 4.5C). Hence there are strong correlations between bubble information and empirical salience in all four tasks.

4.3.7 Correlation of spatial arrangement with fixated bubbles

We use a generative model to predict the empirical salience of a bubble independent of its visual content, but given its location. The generative model, as defined in the Methods section, predicts gaze trajectories on a stimulus given the initial fixation spot and the spatial arrangement of bubbles. Please note that the spatial arrangement of the bubbles alone does not contain information on the frequency of fixations on different bubbles. The model takes into account the central bias of fixations and geometric constraints on the length and direction of saccades. It does not incorporate an explicit inhibition of return (see Discussion). Both the central bias of fixations and the geometric constraints on saccades are grand averages over a large number of full field stimuli from many different categories (see Methods). The model generates fixation sequences on bubble stimuli. From these sequences the average probabilities of fixating individual bubbles on a stimulus are computed. These only locally valid values are now transformed to a global scale in the same way as the relative frequencies of fixations made by actual subjects were transformed into empirical saliences (see Methods). We now consider the correlation of this spatial bias salience with empirical salience (both log transformed). Fig. 4.6A shows an example of a stimulus from the *expression* task where the individual bubbles are labeled with their spatial bias saliences, and Fig. 4.6B shows a scatter plot of spatial bias salience versus empirical salience. Bubbles A and B are looked at very often and have relatively high spatial bias saliences, which is probably due to the fact that they are close to the center of the stimulus and close to each other. Bubbles C and D, which are farther away from the center and have lower spatial bias saliences, are looked at much more rarely. In this specific example, spatial bias and empirical salience are closely related.

For all bubbles of the *expression* task, the correlation between spatial bias salience and empirical salience is highly significant. For the other three tasks, the correlation is highly significant as well (Fig. 4.6C). The correlation of empirical salience with the prediction based on spatial properties is of comparable strength in all four tasks.

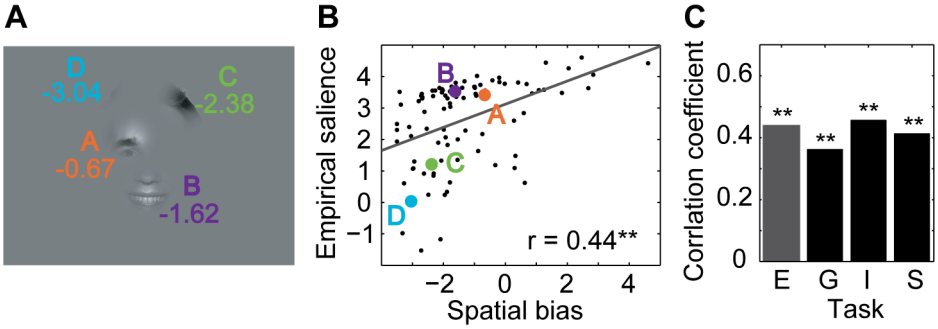


Figure 4.6 Relationship of spatial bias and empirical salience.
(A) Example stimulus of the *expression* task with individual bubbles labeled by their spatial bias salience. (B) Scatter plot of spatial bias and empirical salience for the *expression* task. The positions of the example bubbles are marked by colored dots. The correlation coefficient r is given as a figure inset. (C) Correlation coefficients for all tasks. Two stars mark highly significant correlations ($p<0.01$, t-test).

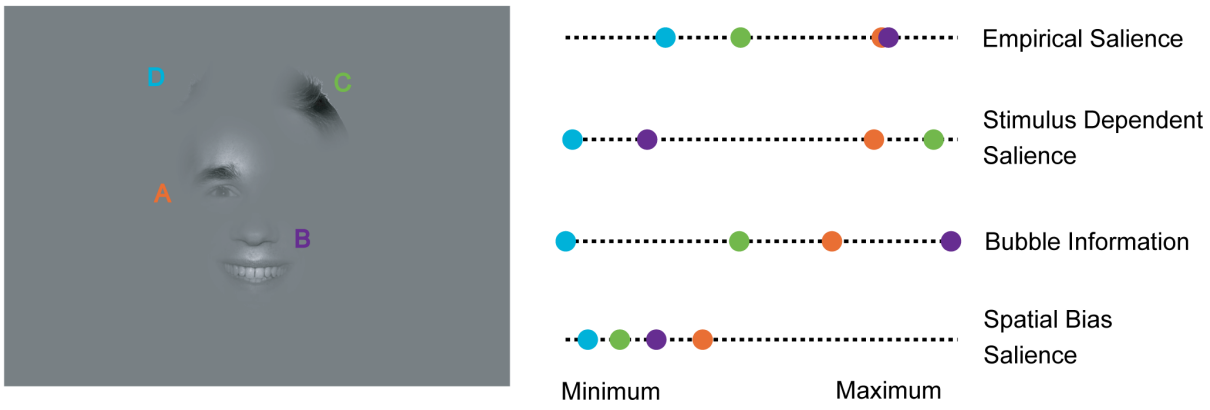


Figure 4.7 Relationship between empirical salience, stimulus dependent salience, bubble information and spatial bias salience for an example stimulus.
The example stimulus from the *expression* task is given on the left. The four values characterizing each bubble are shown on their respective scales (right panel). The range of spanned values for each variable is mapped to the same interval for comparison. The colors code for the identity of the different bubbles.

4.3.8 Partializing the information in low-level stimulus features, bubble information, and spatial arrangement

For a combined view we compare the values of all three predictor variables and empirical salience for the example stimulus (Fig. 4.7). Gathering the information from Fig. 4.3, 4.5, and 4.6 reveals bubble information as the best predictor (e.g., the order of the bubbles according to bubble information is the same as according to empirical salience), followed by the spatial bias and the stimulus dependent salience. This example is reasonably representative for the *expression* task. In other tasks the contribution of stimulus dependent salience, bubble information, and spatial bias salience is more balanced. However, the individual correlations of empirical salience with the three predictors do not address how much the effects of one of these predictor variables are already addressed by another, because of correlations between individual predictors. In the following we address this question, which is crucial for the investigation of the causal role of the individual predictors.

We employ a multivariate linear model to predict empirical salience from the joint set of all the predictors. We analyze how well a linear combination of the stimulus dependent salience, bubble information, and spatial bias salience of each bubble can explain the attention it attracts, as reflected by the empirical salience values. As in the pair-wise correlations, we use the log transform of each predictor variable and correlate with the log transform of empirical salience. The model structure is as follows:

$$\begin{aligned}\log(&Empirical\ Salience)\\ &= \beta_1\\ &\quad +\beta_2 \log(Spatial\ Bias)\\ &\quad +\beta_3 \log(Stimulus\ Dependent\ Salience)\\ &\quad +\beta_4 \log(Task\ Dependent\ Salience)\end{aligned}$$

To address correlations between individual predictor variables, we use semi-partial correlations, which correlate one predictor with empirical salience while controlling for the effect of all other predictors (see Methods).

Table 1 gives the results of this correlation analysis for the four tasks, and Fig. 4.8 summarizes these results visually. The multivariate regression coefficient (R) is highly

significant ($p<0.01$, F-test) for all four tasks, but varies considerably across tasks. For *expression*, the multivariate correlation is highly significant, with 48% of variance explained. Bubble information is the best individual predictor, the pair-wise correlation being highly significant. The individual predictive power of spatial bias salience is smaller, but the pair-wise correlation is still highly significant. Stimulus dependent salience, on the other hand, does not significantly correlate with empirical salience. These results indicate that subjects have much information about where to expect informative bubbles *a priori* and that their attention is guided by this task dependent knowledge. This is exactly what one would expect of a system that is specialized in effectively recognizing facial *expression*. It is clearly inconsistent with a purely bottom-up driven account of overt attention. For the *gender* task, the multivariate correlation coefficient is smaller than for *expression*, but still highly significant, with 27% of variance explained. Spatial bias salience and bubble information have almost the same pair-wise correlation coefficient, both correlations being highly significant. In contrast to the *expression* task, the pair-wise correlation of the stimulus dependent salience is also significant. For the *influence* task, the multivariate correlation is also highly significant, with 36% of variance explained. Again all three predictors show significant, even highly significant, pair-wise correlations. Spatial bias salience has the highest correlation coefficient, followed by bubble information and stimulus dependent salience, the latter two being almost identical. For the *space* task, the multivariate correlation coefficient is smallest, but still highly significant, with 25% of variance explained. Spatial bias salience is the best predictor, followed by bubble information. Both these pair-wise correlations are highly significant. In contrast to *influence*, the correlation coefficient of stimulus dependent salience is very small and not significant.

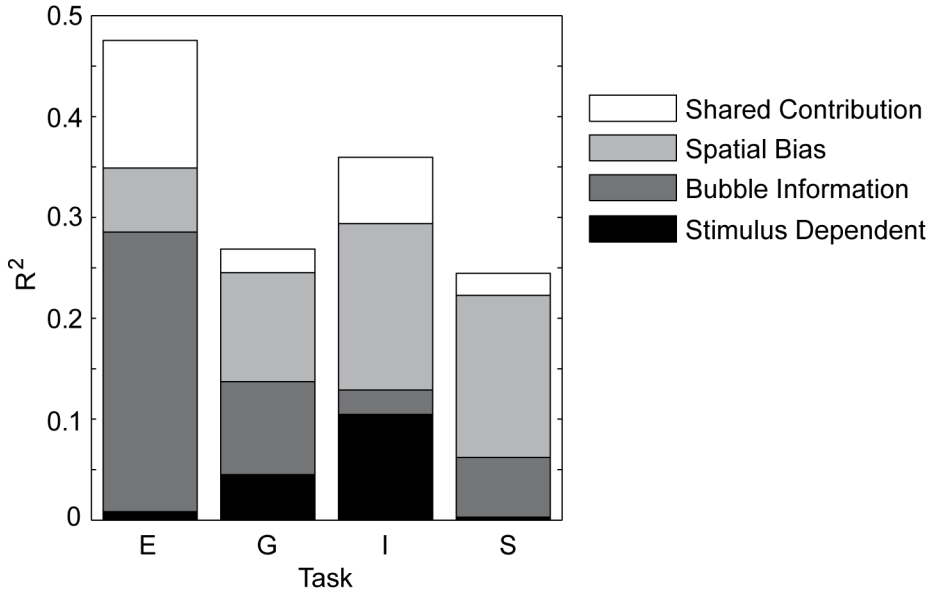


Figure 4.8 Influence of the three factors on empirical salience.
The multivariate regression results are given for all four tasks *expression* (E), *gender* (G), *influence* (I), and *space* (S). The height of each bar depicts the R^2 value; each shaded area represents the squared semi-partial correlation coefficient, which reflects the unique contribution of the respective factor. The white area in each bar represents the amount of variability of empirical salience that can explained by more than one factor.

The previous observations on the relative predictive power of individual predictors in the different tasks are also supported by the semi-partial correlation analysis. The only exceptions are a rather large decrease from the pair-wise to the semi-partial correlation for bubble information in the *influence* task, reflecting a rather small unique influence of bubble information on empirical salience, as well as a noticeable decrease of the semi-partial

correlation coefficient compared to the pair-wise correlation coefficient for the spatial bias based predictor in the *expression* task.

On the level of individual predictors, we make several observations. Spatial bias salience shows a strong and stable contribution in all four tasks. The unique contribution of bubble information is strong as well, but varies considerably over tasks. In the case of *influence,* it is not even significant. Stimulus dependent salience is the weakest predictor of the three, but shows significant correlations in *gender* and *influence*. Each single predictor shows significant normal and semi-partial correlations in at least some of the tasks. Furthermore, the relative contributions of the predictors, in terms of both uncontrolled and semi-partial correlations, vary considerably over tasks. Hence the contribution of the three different factors is dependent on the task, and none can be generally dismissed in an explanation for guidance of overt attention.

Task		Stimulus dependent salience	Bubble Information	Spatial bias	All together
Expression	Correlation coefficient *r*	**0.130**	**0.631** **	**0.437** **	
	Semi-partial correlation coefficient *sr*	0.091	0.527 **	0.252 *	
	Multivariate regression				$R^2 = 0.476$**
Gender	Correlation coefficient *r*	**0.235** *	**0.326** **	**0.362** **	
	Semi-partial correlation coefficient *sr*	0.213 *	0.304 **	0.329 **	
	Multivariate regression				$R^2 = 0.269$**
Influence	Correlation coefficient *r*	**0.324** **	**0.345** **	**0.456** **	
	Semi-partial correlation coefficient *sr*	0.324 **	0.155	0.406 **	
	Multivariate regression				$R^2 = 0.360$**
Space	Correlation coefficient *r*	**0.067**	**0.290** **	**0.412** **	
	Semi-partial correlation coefficient *sr*	0.055	0.243 *	0.401 **	
	Multivariate regression				$R^2 = 0.245$**

Table 4.1 Results of the multivariate regression. Pair-wise regression coefficients and semi-partial regression coefficients for the different predictors are given for each task. The total variance of empirical salience that is explained by all three factors is given in the last column. One star marks significant correlations ($p<0.05$); two stars mark highly significant correlations ($p<0.01$).

4.4 Discussion

4.4.1 General summary

In this study, we quantify and compare the influence of low-level stimulus features, task dependent features, and spatial biases on overt visual attention. The major achievement is a direct and quantitative comparison of the individual influences of these factors on fixation behavior in a single study. The experimental approach builds on the bubble paradigm as introduced by Gosselin and Schyns (Gosselin & Schyns, 2001). It makes use of visual stimuli composed of small image patches, called bubbles, based on face images and forest scenes.

Subjects classified stimuli according to facial expression and gender, or according to scenic openness and human influence, respectively. The subjects' eye movements show that bubbles are not scanned for information and verify our assumption that bubbles are treated as perceptual units. To each bubble, we assigned an empirical salience that adequately represents the fixation probability of the bubble. We further quantitatively assessed three factors that are thought to influence visual attention: first, stimulus dependent salience reflecting the probability of fixating a bubble given its luminance and texture contrast; second, bubble information reflecting how much information a bubble contains with respect to the classification task; and third, spatial bias salience reflecting the fixation probability given the location of the bubble. Bubble information was estimated based on the subjects' classification responses to stimuli composed of one or several bubbles using a model of information integration. We showed that this model is a reasonable approximation of the integration. Interestingly, we found that information of individual bubbles is integrated even if bubbles originate from different images of the same class and independent from their spatial arrangement in the case of forest scenes.

Having measured the three factors bubble information, stimulus dependent salience, and spatial bias salience, we then quantified how well they predict fixation behavior. We found that a substantial portion of variance of empirical salience could be explained by all three factors combined, although the share of variance explained varies across tasks. Pair-wise correlations between empirical salience and each of the factors indicate clear differences between the three factors. Empirical salience shows high correlations with spatial bias throughout all four tasks, whereas both the correlations with stimulus dependent salience and bubble information vary strongly with tasks. Stimulus dependent salience is the weakest predictor, but reaches significant levels in the *gender* and *influence* tasks. Bubble information is the best predictor in the *expression* task but for the other tasks it reaches slightly lower correlations with empirical salience than does spatial bias. Surprisingly, the semi-partial correlation coefficients, which reflect the unique contributions of each predictor controlling for the influences of the other factors, are only slightly lower than the pair-wise correlation coefficients. This indicates that all three factors act almost independently on visual attention. In summary, we find that all factors contribute, but that the absolute and relative strength of contribution depends on the task.

We now look into the potentially critical issues and shortcomings of our paradigm. These fall into two overall categories. First, we discuss the validity of our different measures. Second, we analyze how much the results obtained using our bubble paradigm generalize to more natural conditions.

4.4.2 Validity of bubble measures

Empirical salience. One basic assumption of the present approach is that the empirical saliences of different bubbles are independent from each other — i.e. the empirical salience of a bubble is not influenced by any other bubble on the same stimulus. An indicator of a violation of this assumption would be a change of the ratio of fixations falling onto two bubbles when other bubbles were presented simultaneously. We tested whether empirical salience values can predict the average number of fixations made by the subjects onto each bubble in all stimuli. The test resulted in very small errors showing that our assumption of independence between bubbles with respect to empirical salience is not violated.

Stimulus dependent salience. We characterized stimulus dependent salience as the conditional probability of fixating on an image patch given its local luminance and texture contrast for several reasons. First, these two low-level features were shown to correlate with

fixation behavior in many previous studies (Reinagel & Zador, 1999; Einhäuser & König, 2003; Parkhurst & Niebur, 2004; Açık et al., 2009). Hence, the present study can be compared directly with this previous work. Second, in an independent study, we observed that the strength of influence of different low-level features on overt visual attention is highly correlated over image categories and tasks (Betz et al., 2010). Hence, the potential benefit of additional features appears small. Third, as a control we compared stimulus dependent salience with a measure of salience obtained by a publicly available model, often used as a baseline (Itti & Koch, 2001; Walther & Koch, 2006). Indeed, the correlation of the two sets of saliences is high and in all tasks in the range of 0.4 – 0.7. Furthermore, the correlation of salience according to the model by Itti and Koch with empirical salience of bubbles is not qualitatively different from the data presented here. Fourth, previous studies showed that luminance contrast influences the response of area V1, but not the response of higher areas (Rolls & Baylis, 1986; Avidan et al., 2002). These results indicate that luminance contrast is a good measure for the relevance of stimulus dependent signals in early visual cortex and justifies the term "low-level". Fifth, another recent study claims that stimulus dependent salience is well described by luminance contrast without the need to introduce more complex kernels (Kienzle et al., 2009). Sixth, texture contrast, which is defined as second-order luminance contrast, is usually considered a low-level feature in that sense as well and triggered some debate in the literature (Parkhurst & Niebur, 2004; Açık et al., 2009). For these reasons we decided to base our characterization of low-level contributions on luminance and texture contrast.

Spatial bias salience. We characterized spatial bias salience through a generative model of fixation behavior. The model takes into account the central bias of fixations (0^{th} order) and geometric constraints on the length and direction of saccades (1^{st} order). While the location of a particular fixation has an influence on the next fixation, we do not model higher order dependencies. Specifically, we do not account for inhibition of return, which would be a 2^{nd} order relation of direction and length of saccades. Inhibition of return is characterized as a small delay of saccades that return to the location of a previous fixation. As the current investigation is not concerned with these dynamic aspects, it is not of relevance here. Furthermore, recent studies report that inhibition of return might actually not change viewing strategy for complex scenes (Hooge et al., 2005; Smith & Henderson, 2009).

Bubble information. Estimation of bubble information is based on the complete data set and involves a specific model of information integration. Both issues are considered in turn. In principle it would have been possible to estimate bubble information directly from stimuli presenting single bubbles only. This approach comes, however, with several disadvantages. First, the presentation of only single bubbles as stimuli is rather inefficient. To get reliable estimates of bubble information, each single bubble stimulus would have to be shown much more often. Since a participant cannot respond to the same single bubble stimulus twice and should not see individual bubbles too often, many more participants would be needed. Additionally, the responses on stimuli with several bubbles would be left unused, further diminishing efficiency. Given that in the present study 75 subjects were investigated, more than in any of the eye tracking studies cited above, this issue of efficiency quickly gets prohibitive. Second, using qualitatively different stimuli for computing empirical salience and bubble information potentially introduces systematic biases. For example, the difficulty of the classification task is increased considerably on single bubbles compared to stimuli with several bubbles. This might lead to performance near chance level, which in turn could cause subjects to lose motivation and concentration. Third, for the purpose of the present study our interest is focused on an estimate of bubble information in the context of the stimulus. In the event that estimates of information of isolated bubbles and bubbles in more complex context

diverge (e.g. a systematic increase or decrease), the latter would be the relevant measure as it matches the viewing conditions during the task. These reasons further grow our confidence in the validity of the applied methods.

Several models of information integration are conceivable. The mode of information integration is an important topic in its own right and a complete treatment is beyond the scope of the present paper. We assume a probabilistic integration model but also considered two other models of information integration: first a local model that captures stimulus information by the maximally informative bubble, second a global model that differs from the probabilistic model by capturing contra factual evidence for the different choice possibilities. Compared to the p-model these models both show lower performance (see supplementary text in section 4.6.4). Furthermore, under the assumption of the p-model being the true model of information integration, the estimates for bubble information resulting from the global fitting procedure are unbiased and have moderate variance (see supplementary text in section 4.6.2 and Fig. 4.14 in same section). This indicates that the predictions based on the p-model are generally good estimates of bubble information.

In conclusion, although we did not show that a probabilistic model for information integration is the true or optimal model we demonstrated that the estimates for stimulus information obtained through it are robust and consistent with the majority of the data. The influence of configural information in face stimuli has been described before and does not pose a problem in the current context. The question of which is the optimal model of information integration is left to be answered by future research.

For these reasons we decided to show stimuli with varying numbers of bubbles in a homogeneous set and to employ the information integration model and global fitting procedure. In so doing we assess the two behavioral measures, bubble information and empirical salience, from the same subjects during the same experimental trials and make optimal use of experimental data to improve the signal to noise level.

Effects of bubble position on information integration. The discrepancy between stimulus information in bubble stimuli of condition *same* versus *permuted* in task *expression* (see Fig. 4.2) could have several causes. The faces are similarly positioned in all stimuli so that the location of the bubbles hints at which bubbles contain relevant information: subjects might know a priori where informative regions, e.g. the eyes or the mouth, are located and select fixation targets accordingly. Furthermore, faces are special perceptual stimuli. Specific brain areas are devoted to the processing of face stimuli, and identification can be completely disrupted by reversing a face image (Kanwisher & Yovel, 2006; Freiwald et al., 2009). Position effects could, therefore, play a more important role for the classification of face images than for the classification of forest scenes (Calder et al., 2000). Indeed, a major effect of permutations in the *expression* task is a largely improved fit of the p-model. This indicates that, once the standardized positioning is violated, different bubbles are treated as independent pieces of information, enabling the "normal" mode of information integration. The effect of bubble position is less pronounced in the *gender* task. For the *gender* stimuli, supposedly more regions contain information and the correlation between bubble position and bubble information is weaker. In summary, our data indicate that position effects have some influence in face stimuli, but less so in the forest scenes.

4.4.3 Generalization to full scenes

Do the observed correlations between empirical salience, on the one side, and stimulus dependent salience, bubble information, and spatial bias salience, on the other side generalize

to full field images? This is a variation of the eternal question where to place the balance between complex natural conditions and well controlled laboratory stimuli. Here, the answer depends critically on whether the four measures we employ are preserved on full field stimuli. For example, it is decisive whether the empirical salience of image patches measured on full field stimuli is comparable to the empirical salience measured on bubble stimuli. In the same way, bubble information, stimulus dependent salience and spatial bias salience need to be preserved. If the four measures that characterize a bubble were preserved when the bubble is embedded in a full field stimulus then the relationship between the measures, in particular the correlations between them, would be preserved as well and our results should generalize to full scene viewing. We consider this question for each of the measures in turn.

Stimulus based salience, as we defined it, is just dependent on a local image patch. It is thus preserved for full field stimuli. Bubble information measures how much information with respect to a task is contained within a single bubble. The amount of information contained appears largely independent of bubble context and thus only depends on the image patch itself. Spatial bias salience, as we define it, is based on global fixation and saccade biases assessed from a large variety of full field stimuli. Hence, the effect of spatial bias should be largely independent of whether an image patch is embedded into a full field or bubble stimulus. The question of whether the measure of empirical salience is preserved on full field stimuli is more intricate. The observer may very well fixate image regions in the bubble stimuli that would never draw her attention given the complete image. We tested this by correlating empirical salience of bubbles with the fixation densities of the full field images containing those bubbles ($r=0.79$, $r=0.75$, $r=0.55$, $r=0.32$ for *expression*, *gender*, *influence,* and *space*, respectively; $p < 0.01$ in all cases). Since empirical salience of individual bubbles is well preserved on full field stimuli, we expect that our findings generalize to full scene viewing.

Previously, it was debated whether the informative regions uncovered by Gosselin and Schyns' bubble paradigm (Gosselin & Schyns, 2001) are valid for full scene viewing as well. Murray and Gold argue that the bubble stimuli change the information integration strategy employed by the observer (Murray & Gold, 2004). A former study showed that observers used different stimulus regions to identify faces, depending on which regions were covered by Gaussian white noise (Schwartz et al., 1998). It is conceivable that for full field images, which include redundant features, observers normally base their classification decision on only one or two of these features. The bubble stimuli force the observers to use different features on different trials, because only small fragments of the stimulus are shown on any given trial (Gosselin and Schyns argue, however, that these concerns are unfounded (Gosselin & Schyns, 2004)). These potential problems are not relevant for our study since we do not claim that certain bubbles would be used by the observers to solve the classification task on full fields, whereas other bubbles would not. Instead, we quantify the information of each single bubble, i.e., how well the task can be solved given only this bubble. By using the information integration model, we actually incorporate the observer's strategy to use different image regions, depending on which regions are shown. Hence, our measure of task dependent information is not invalidated by the use of bubble stimuli.

In summary, we consider the present experimental paradigm a most sensible compromise, balancing between the complexities of natural conditions and well controlled laboratory stimuli, and suitable for the questions addressed.

4.4.4 Relationship of low-level and high-level features to bottom-up and top-down neural signals

One of the most debated issues concerning overt visual attention is the role of bottom-up and top-down signals on a neural level. This issue is not integral part of the results of the current study. In the present study we discuss the influence of stimulus dependent salience and bubble information. Stimulus dependent salience translates directly to low-level stimulus features and to some degree, these features can be identified with bottom-up signals. It has been shown that neurons in V1 are sensitive to these features (Hubel & Wiesel, 1961; Rolls & Baylis, 1986; Avidan et al., 2002). To reach relevant motor centers and influence eye movements, these signals have to traverse the hierarchy of the visual system (Fellerman & Van Essen, 1991). This may be viewed as a bottom-up process. The second measure, bubble information, relates to high-level features of the visual stimulus interpreted in a specific context. Considering complex response properties in high-level brain areas, these are a natural place to extract such information (Tanaka, 1996). Again, in view of abundant connectivity, it is plausible that such information is sent down to lower areas of the hierarchy in a top-down manner. However, receptive field properties of neurons in V1 are complex, and non-classic surround effects are far from understood (Olshausen & Field, 2005). Furthermore, it has been proposed that essential characteristics of a salience map are already captured in the response properties of V1 neurons (Li, 2002). For that reason we are cautious using the terms *top-down* and *bottom-up signaling*, and we took care not to make unwarranted speculations about the site of the integration of the observed contributions of low-level and high-level stimulus features.

4.4.5 A Unified theory of overt visual attention

Many low-level image features were suggested to play an important role for the guidance of visual attention (Itti & Koch, 2001). When compared to random image locations, fixated regions of natural and artificial images are characterized by higher decorrelation of intensities of nearby image points (Reinagel & Zador, 1999; Parkhurst & Niebur, 2003), higher luminance contrast (Reinagel & Zador, 1999; Parkhurst & Niebur, 2003, 2004; Einhäuser et al., 2006), texture contrast (Parkhurst & Niebur, 2004; Einhäuser et al., 2006), color contrast (Frey et al., 2007; Frey et al., 2008), orientation contrast (Parkhurst et al., 2002), flicker and motion contrast (Carmi & Itti, 2006), strong statistical dependencies between frequency components of different orientation like curved lines (Saal et al., 2006) , edges (Tatler et al., 2005), occlusions or isolated spots (Krieger et al., 2000), and disparity (Jansen et al., 2009). These effects, however, appear to be relatively weak (Tatler et al., 2005), and another study reports that locations of extremes of luminance intensity, luminance contrast, high spatial frequency content, and edge density do not match with locations of fixations (Mannan et al., 1996). Yet another study puts forward contradicting evidence in favor of the role of high spatial frequency content (Baddeley & Tatler, 2006). The strength of these effects was found to vary with image type (Parkhurst et al., 2002; Açık et al., 2009). Still, the idea is that with increasing complexity of the features investigated a faithful description of human overt visual attention can be reached.

This line of research has come under attack from two sides. On the one hand, Kienzle and colleagues show that much of the observed correlation of selected fixation points in a free viewing task on gray-scale images of natural scenes can be captured by an extremely simple center surround mechanism (Kienzle et al., 2009). On the other hand, recent studies found that high-level features play an important role in overt visual attention and act more strongly on fixation behavior than low-level features when subjects engage in visual search tasks (Chen & Zelinsky, 2006; Underwood et al., 2006; Einhäuser et al., 2008). In more natural settings, task

and context have a strong impact on eye movements as well (Rothkopf et al., 2007). Also models of visual attention that employ top-down processing were successfully applied to visual search tasks (Tsotsos et al., 1995; Rao et al., 2002; Turano et al., 2003; Hamker, 2004; Navalpakkam & Itti, 2005). Recent work tries to combine low-level and high-level cues (Cristino & Baddeley, 2009; Vincent et al., 2009). The latter study specifically investigates the salience of light sources (very high luminance contrast) in natural scenes at dawn and dusk. They show that high-level features and spatial biases make the largest contribution in a mixture model, which is in line with the results reported here. However, in the work by Vincent et al. (Vincent et al., 2009) the definition of high-level features like foreground/background contains a subjective component and might correlate strongly with low-level features like disparity. Indeed, we could recently demonstrate that disparity has a strong influence on the selection of fixation points in stereoscopic presentation (Jansen et al., 2009), close regions being viewed earlier than far regions. Furthermore, about 40% of this effect survives in 2D presentation. This highlights the problem to define objectively low-level and high-level cues and to analyze their independent contribution to the guidance of gaze movements. Some experimental studies assessed the informativeness of image regions by subjective ratings (Mackworth & Morandi, 1967; Antes, 1974); or they made use of identified informative regions of face images for different tasks (Malcolm et al., 2008). In agreement with our data, these investigations show that fixation patterns vary for different tasks even if the visual input is identical — i.e., that high-level features like task dependent information have an influence on attention, and that more informative regions are fixated upon more often than less informative ones. The advantage of our approach is that it enables us to quantitatively measure task dependent information in an objective way. Another study presents an information theoretic approach to the combination of different cues (Kanan et al., 2009). They demonstrate that the model clearly outperforms models with pure bottom-up architectures. Furthermore, Ehinger and colleagues give a highly informative comparison with current contextual guidance models (Ehinger et al., 2009). Our results are in line with these studies. Averaged over all the tasks investigated, high-level features contribute more than low-level features.

One center issue of the debate about low-level and high-level features is whether, and to what degree, they have a causal role versus pure correlative effects. A study on images whose luminance contrast was locally modified shows that fixations are attracted by increases as well as decreases of luminance contrast, but that the effect within the region of normal variance of luminance contrast is small (Einhäuser & König, 2003). Furthermore, these observations cannot be explained by induced changes in texture contrast (Açık et al., 2009). This argues against a causal effect, but in favor of a pure correlative effect of luminance contrast in a free viewing task on natural stimuli. Our present results agree with the aforementioned studies inasmuch as the low-level factors exhibit, on average, weak effects on fixation behavior. However, our analysis of the correlation of empirical salience with the three predictors uncovers a surprising fact. The semi-partial correlations are only a little smaller than the full correlations. This indicates little redundancy of the three predictors — i.e. low-level features are not coincident correlations of high-level features in many tasks. This argues that none of the predictors can be neglected, but that a true integration is to be achieved. This is very much in the spirit of recent proposals, putting the problem of overt attention in a Bayesian framework (Zhang et al., 2008; Gao & Vasconcelos, 2009).

Concerning the role of spatial biases on visual attention, it was pointed out that the spatial bias towards the screen center has to be taken into account when studying the effect of image features on selection of fixation points (Mannan et al., 1996; Tatler, 2007). Furthermore, some work has been done on the statistical properties of saccade length and directions.

Human saccades can be modeled as a Levy flight with a heavy-tailed distribution (Brockmann & Geisel, 1999) and it can be shown that under certain assumptions such a distribution leads to optimal scanning behavior. Research on higher order correlations, i.e. dependencies of selected fixation points within a trajectory, is still rare (Tatler & Vincent, 2008). Given our current knowledge of spatial properties, a comparison of several models of fixation behavior revealed that the best performance is obtained from a strategy combining top-down information and spatial bias, which, however, was defined as the restriction of fixations to one side of the image (Turano et al., 2003). Our results support this view, showing a surprisingly high correlation between spatial bias and visual attention. This effect is strong and consistent in all tasks tested. This contrasts with the emphasis on low-level and high-level features in current models of visual attention. Forthcoming models should put the spatial properties of eye movements on an equal footing with other factors.

The present study contributes to focusing discussions of models of attention on quantitatively testable properties. Low-level stimulus features, task dependent information content, and spatial viewing biases jointly explain a substantial fraction of the variation of empirical salience — i.e., a unifying theory of visual attention will have large predictive power. Furthermore, each of the three factors contributes significantly. A unified theory of overt visual attention has to account for all of them.

4.5 Methods

4.5.1 Experimental setup

Ethics statement
All subjects were informed about the experimental procedure, the eye-tracking device, and their right to withdraw from the experiment at any time. However, they were initially kept naïve as to the purpose of the experiment and were debriefed after the experiment. All participants consented in writing to take part in the experiment and to allow scientific usage of the recorded data. The experimental procedure conformed to the Declaration of Helsinki and national guidelines.

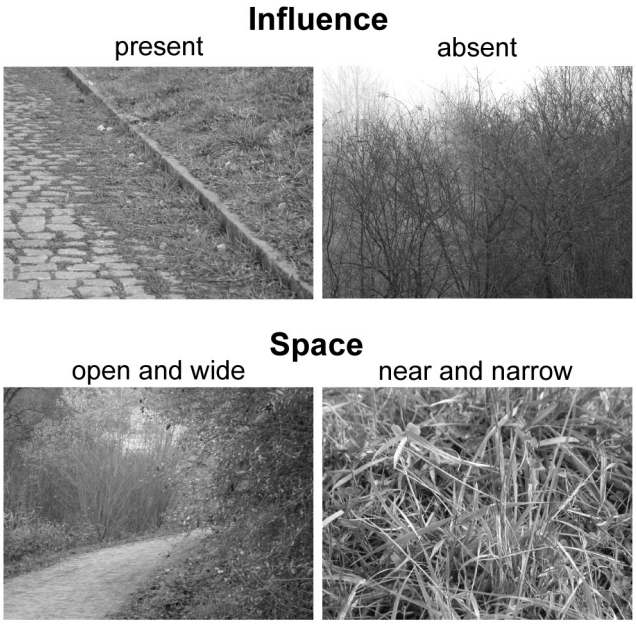


Figure 4.9 The different stimulus classes.
Subject's had to classify faces and forest scenes according to four tasks (*expression*, *gender*, *influence*, and *space*). For the forest scenes, the different response possibilities are given above the example stimuli. The stimuli are shown as full fields and are used for bubble stimuli construction. For copy right reasons, we cannot show the face stimuli here but we refer the reader to Tottenham et al. (2009). The face stimuli are taken from the "NimStim" stimulus set.

Participants
75 student volunteers participated in the experiment (39 female, 36 male). Their ages ranged from 18 to 41, with a mean of 24.2 years. All had normal or corrected-to-normal vision, which was confirmed by a vision test with Landolt rings. Participation was voluntary, and participants either were granted extra course credits or received monetary compensation for their participation.

Apparatus and recording
Participants' eye movements were recorded with the head-mounted Eyelink II eye-tracking system and the Eyelink II software package (SR Research, Ltd., Mississauga, Ontario, Canada). Monocular eye-position data were sampled with infrared-based tracking only, using a sampling rate of 250 Hz. The saccade classification of the Eyelink system is based on velocity and acceleration. A saccade starts if an initial acceleration threshold of 8000°/s^2 is exceeded and a distance of at least 0.1° is covered with a minimal velocity of 30°/s. Fixation points are then defined by the samples in between two saccades. Stimuli were presented on a 21-inch Samsung Syncmaster 1100 DF 2004 (Samsung Electronics Co. Ltd., Korea) CRT monitor at a distance of 80 cm from the subject, using a display resolution of 1024x786 pixels and a refresh rate of 120 Hz. These settings resulted in a spatial resolution of 33 pixels per degree of visual angle. No headrest was used.

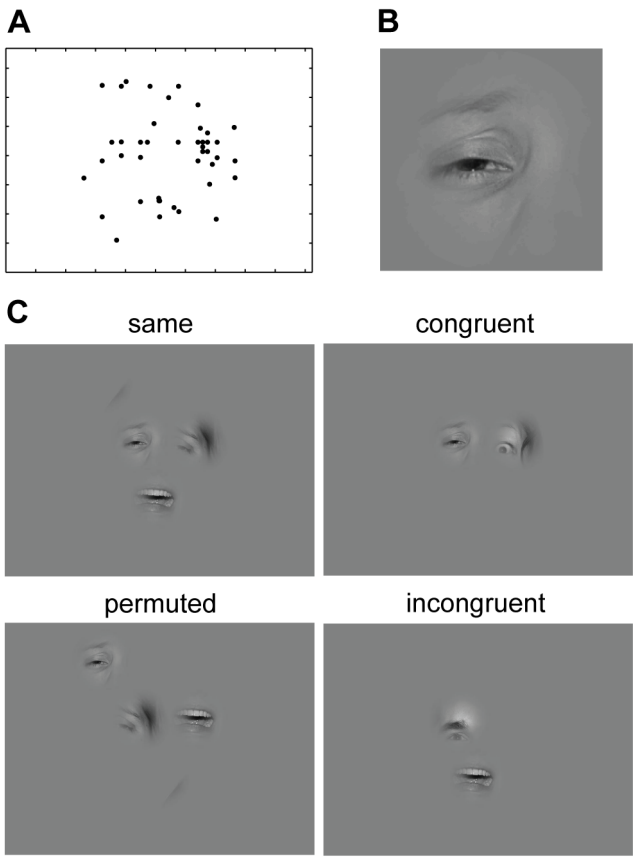


Figure 4.10 Bubble stimuli.
(A) Distribution of bubble positions for the *expression* task. (B) A single bubble based on a patch of 6 visual degrees from a full field face stimulus. The patch was filtered using an eccentricity dependent frequency filter simulating the drop of spatial acuity and a Gaussian mask to avoid edge effects. (C) Different types of bubble stimuli were generated. Stimuli of the *same* condition are built from patches of the same image. Stimuli of the *congruent* and *incongruent* condition are built from patches of different images of the same class or of different classes, respectively. *Permuted* stimuli were created for each of the three conditions by shuffling the positions of bubbles.

Stimuli
All stimuli were based on gray-scale face images (Tottenham et al., 2009) and forest scenes (the forest scene photographs were used with permission from W. Einhäuser and P. König (Einhäuser & König, 2003)). Photographs used for the construction of stimuli were selected on the basis of pre-experiments (forest scenes: Steinwender J (2005) Bachelor's thesis; faces: pre-experiment, data not shown). Face images had to be classified in different **tasks** (see

below) according to gender (***gender***) and expression (***expression***), forest images according to scenic openness (***space***) and human influence (***influence***). Only photographs that were evaluated consistently by all participants of the pre-experiments were included in the present study. These responses defined the different **classes** used below during stimulus construction. We selected a total of 24 photographs of faces and 36 photographs of forest scenes. The stimulus sets were balanced in the context of each of the four tasks. In 4% of all trials, stimuli were photographs shown in full field condition (Fig. 4.9). Although these full fields were shown during the main experiment to control for changes in classification, their main purpose was to serve as a basis for the creation of bubble stimuli. In 96% of the trials, bubble stimuli constructed from the same basic set of photographs were presented. These were created in three steps. First, 6.0° square patches were selected from the available full field photographs. Second, the image patches were space-variant filtered, imitating the retinal resolution as a function of eccentricity, and masked by a Gaussian envelope. Third, these bubbles were recombined and placed on an equiluminant gray background in different ways to create a variety of bubble stimuli. A total of 2061 gray-scale stimuli were used.

The selection of image patches from full fields was governed by the following criteria: first, we selected image patches from locations where the fixation density obtained in the pre-experiments was very low or very high. This way of selecting patch positions yields a set of patches with diverse empirical saliences. Second, since bubbles should be independent units of information, they must not overlap. Third, for each bubble on a particular full field stimulus, there should be bubbles on other full field stimuli that occupy the same position. This constraint allowed controlling for position effects when combining bubbles from different full fields. Ideally, some of these bubbles on other full fields should be close to minima and some to maxima of their respective fixation distribution. We used a randomized algorithm to generate an appropriate selection. Since the aligned geometry of the face stimuli made it impossible to fully satisfy the latter constraint, a residual set of bubbles for the face stimuli was selected by hand. The resulting distribution of bubble centers for the *expression* task is shown in 4.10A.

The selected image patches were first filtered using an eccentricity-dependent frequency filter that simulates the decline of visual acuity towards the edges of the visual field as resulting from the non-uniform distribution of photoreceptors on the human retina (Sere et al., 2000). This approach ensures that all information present in a bubble can be gained by fixating the bubble center and that scanning bubbles is inefficient. To prevent potential artifacts resulting from sharp apertures, the space-variant filtered patches were masked using an isotropic Gaussian window with a standard deviation of 1.0°. This made the bubbles blend inconspicuously into the gray background. An example is shown in Fig. 4.10B.

The final bubble stimuli were created by combining bubbles. In a small fraction, individual bubbles were shown (12%). The remaining stimuli were composed of two (42%), three (26%), four (14%) or five bubbles (2%). Combining several bubbles, depending on their full field stimulus of origin, allows different **conditions** (Fig. 4.10C). **Same** stimuli (50% of all stimuli, including single bubbles) were composed entirely of bubbles from the same full field image. **Congruent** stimuli (15%) were composed of bubbles from different full fields that were classified in the same way during the pre-experiments (they belong to the same class). **Incongruent** stimuli (15%) were composed of bubbles from full fields of different classes. **Permutations** (16%) were created by shuffling the positions of the bubbles. The final stimulus set was created using a randomized algorithm that optimized the set with respect to the constraint that each individual bubble should appear in the same number of stimuli.

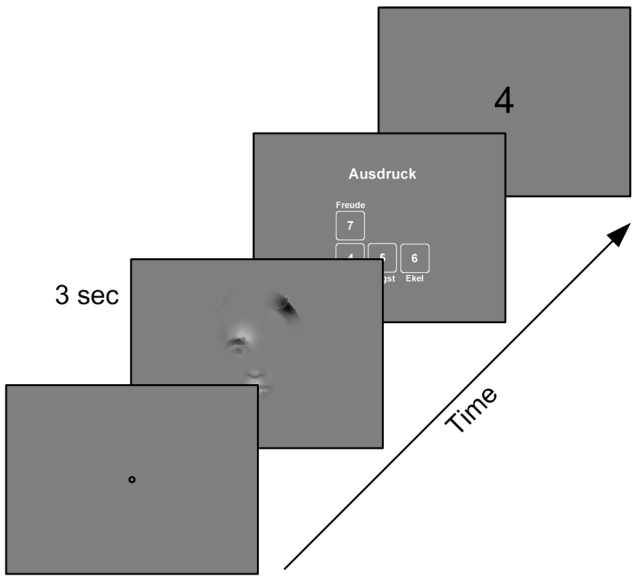


Figure 4.11 Experimental procedure.
Each trial began with the presentation of a fixation point used for drift correction. Subsequently, the stimulus was presented for 3 seconds. The response screen was displayed until the subject responded to the classification task by pressing one of the indicated keys. The subject's choice was then shown as feedback.

Classification tasks

During the experiment, participants classified visual stimuli in four different tasks. In the first task, participants tagged stimuli according to the facial *expression* of the actors into the classes "happy," "sad," "fearful," or "disgusted." Similarly they classified *gender* into "male" or "female." For the *space* task, participants were asked to choose between "close and narrow" or "wide and open." They were instructed to respond "close and narrow" if the image was a close-up or if it would not be possible to leave the scene—for example, if leaves and branches were blocking the view. They were told to respond "open and wide" if it was possible to look far ahead. For the *influence* task, we asked participants to look for indicators of human influence such as houses, roads and paths, trunks of trees, fences, and hewn stones and to classify the stimuli into either "present" or "absent." The wording of the instructions was the same for all participants.

Procedure

A complete experimental session lasted approximately one hour. It was divided into four blocks, one for each of the four classification tasks. Face stimuli and forest scene blocks were presented alternately. In the beginning of the experiment, participants were instructed about the procedure, and example bubble stimuli were shown. They were directed to classify the stimuli by pressing numbers on the keyboard's keypad and to take their best guess in cases where they were not sure about the stimulus' class.

Before the beginning of each block, the eye tracker was calibrated, and task and answer choices for that block were explained and exemplified. Each block consisted of 70 trials that were presented in constrained random order (see below). Each trial began with the presentation of a fixation cross in the middle of the screen. Whenever the fixation of the cross indicated a notable decline in tracking quality, the eye tracker was recalibrated. This ensured that the mean tracking error for at least one eye was always lower than 0.4°. If the cross was fixated properly, the conductor of the experiment triggered the stimulus presentation. We excluded the very first fixation from all subsequent analysis, as it directly reflects the preceding fixation of the fixation cross. The trial lasted for 3 seconds and was followed by the answer screen, which stayed on until participants responded by using the keyboard. There was no time limit for the decision. Before the next trial started, visual feedback of the participant's response was given to minimize classification errors due to typos (Fig. 4.11).

The stimuli shown to each participant and their order were selected by a randomizing algorithm that respected the following constraints: for each participant, each stimulus was shown at most once; each bubble was presented at most four times; and stimuli with the same bubble were not shown in direct succession. Furthermore, on average, each stimulus should be shown to 8 participants, and the variation in the number of participants that have seen a particular stimulus should be as small as possible.

4.5.2 Data analysis

In the following, we first define a measure for the empirical salience of bubbles as quantified by fixation probability. Then we derive measures for the spatial bias, and the stimulus dependent and task dependent effects. These three measures will be used to investigate the relative contributions to the empirical salience of stimuli. All three measures put the bubbles in a global order.

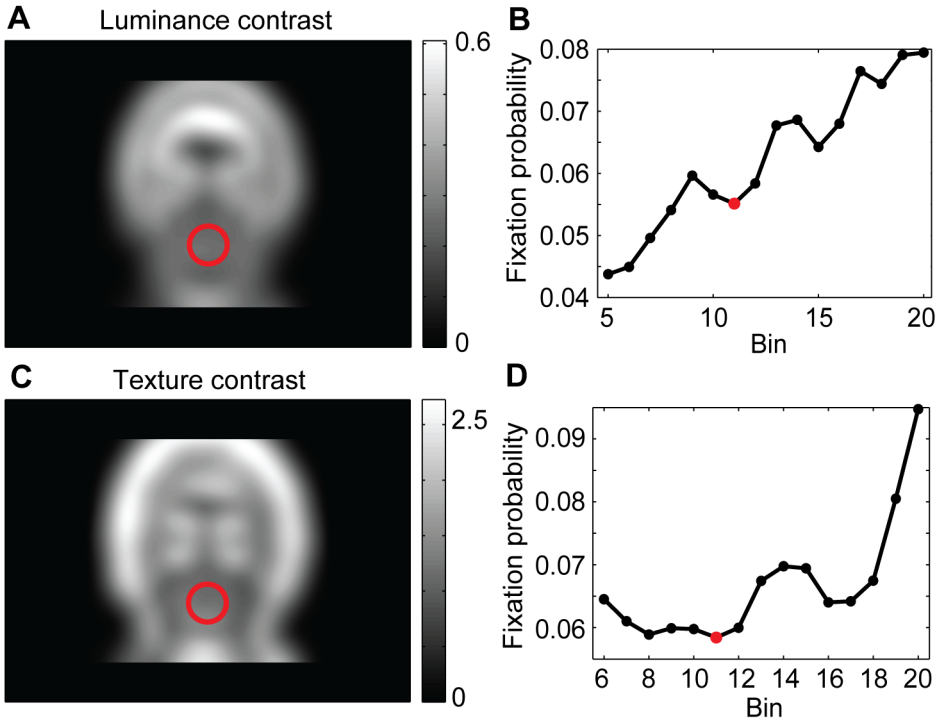


Figure 4.12 Computation of stimulus dependent salience.
For each bubble, stimulus dependent salience was computed by considering the luminance and texture contrast map of the embedding full field (A and C). Luminance and texture contrast at the location of the bubble (marked by red circles for one example bubble) are then mapped to fixation probabilities (red dots). These mappings (B and D) map luminance and texture contrast bins (see text) to fixation probabilities and were obtained in a baseline study using a large number of stimuli from different categories. The resulting fixation probabilities based on luminance and texture contrast were multiplied yielding the stimulus dependent salience.

Empirical salience
To obtain a global quantification of empirical salience, we assume that on any stimulus *S*, the ratio between the number of fixations at a bubble *A*, *Fs(A)*, and the number of fixations at another bubble *B*, *Fs(B)*, is independent of the context in which both are presented. This implies

$$\frac{F_S(A)}{F_S(B)} = \frac{E_A}{E_B} \quad (1)$$

for any stimulus *S* with bubbles *A* and *B*, where E_A and E_B are global measures of *empirical salience,* which are independent of stimulus context. From this, it follows that the equation

$$\frac{F_S(A)}{\sum_{L \in S} F_S(L)} = \frac{E_A}{\sum_{L \in S} E_L} \tag{2}$$

holds for any stimulus *S* and any bubble *A*. Because every stimulus was presented to several subjects, we have, in fact, several left-hand sides of this equation. We average them for each stimulus and bubble. Next, the resulting equations are grouped into a linear system, and we compute the *empirical salience* as the best approximate solution. We eliminate one degree of freedom by imposing a scale, demanding that all empirical saliences sum to one.

Stimulus dependent salience

To characterize the bottom-up contribution to fixation behavior, we use a feature-based salience model. It models the conditional probability of fixating a location of an image, given a set of local low-level image features. Here we consider luminance contrast and texture contrast as features.

Luminance contrast is defined as the standard deviation of the luminance intensity in an image patch, normalized by the mean intensity of the entire image (Einhäuser et al., 2006, Reinagel & Zador, 1999). We calculate it using circular patches weighted by a Gaussian window, *G*, in close analogy to the computation of a bubble. Formally, the luminance contrast of a pixel, *LC(x)*, is given by

$$LC(x) = \frac{1}{I_{task}} \sqrt{\left(I(x-\Delta) - \bar{I}(x)\right)^2 \cdot G(\Delta)} \tag{3}$$

where *I(x)* is the map of luminance intensity at each pixel, Δ is the displacement relative to the center of the bubble, and $\bar{I} = I \cdot G$ is the smooth luminance map obtained by a convolution with a Gaussian of the same size as the Gaussian used in bubble construction. Please note that the normalization deviates from the definition given by Reinagel and Zador (Reinagel & Zador, 1999) and Einhäuser et al. (Einhäuser et al., 2006). In these previous studies luminance contrast was normalized in each individual image. Here, however, the bubble stimuli show only a limited aperture of the respective full field stimulus. Hence varying normalization of bubble stimuli, due to not visible differences in the respective full field stimuli would make contrast values incomparable. Furthermore, in conditions *congruent* and *incongruent,* several different full field stimuli contribute. There is no obvious generalization of an image-specific normalization procedure to these conditions. For these reasons we follow the suggestion of Zhang et al. (Zhang et al., 2008) and normalize luminance contrast by the mean luminance contrast over all the images of one task (I_{task}). This is based on the assumption that the influence of a bubble's contrast on the viewing behavior depends on the whole range of contrast values appearing in the images of one category. Fig. 4.12A shows a luminance contrast map of one of our full field stimuli.

Texture contrast is defined as the standard deviation of the luminance contrast values in an image patch, normalized by the mean luminance contrast of the entire image (Einhäuser et al., 2006). Formally,

$$TC(x) = \frac{1}{LC_{task}} \sqrt{\left(LC(x-\Delta) - \overline{LC}(x)\right)^2 \cdot G(\Delta)} \quad (4)$$

where $\overline{LC} = LC \cdot G$ is the map of the Gaussian weighted mean luminance contrasts. Analogous to luminance contrast, we normalize by the mean luminance contrast over all images in one task. The luminance contrast map, *LC*, used for the computation of the texture contrast, is calculated with a Gaussian window of a quarter of the size of a bubble. For the subsequent computation of texture contrasts, the same Gaussian window, *G*, as for bubble creation is used. The luminance contrast and texture contrast of a single bubble are defined as the contrast values at the center of the bubble.

Based on the feature contrasts of each bubble, we now derive a scalar describing the stimulus dependent contribution to fixation probability (Fig. 4.12). In a previous study we investigated the relation of luminance contrast and texture contrast with fixation probability in natural stimuli (Schumann et al., 2007; Açık et al., 2009). From the observed distribution of selected fixation points and the image statistics, we used Bayes' rule to determine the conditional probability to fixate a given location. Importantly, the data were well described by a model assuming independent contributions of luminance contrast and texture contrast. Here we use this mapping, which originates from an independently obtained data set, to predict fixation probability based on the luminance contrast and texture contrast of the bubble stimuli.

For computational efficiency and optimal usage of data we bin the luminance contrast and texture contrast values of each image. We chose 20 bins with boundaries so that the number of available image locations falling into each bin is constant. Next, the probability of a feature value (luminance contrast or texture contrast) occurring at a fixated location was calculated. Then priors on the image features and fixation locations are computed. The priors on the image features are constant due to the equilibration of the distribution. The priors for the fixation locations were estimated for each image category. Both the feature and fixation location priors were corrected for the spatial viewing biases to obtain a measure based purely on low-level image features. The probability of fixating a location, given its local features, was then estimated using Bayes' rule. Finally, the stimulus dependent salience value of each bubble was calculated as the product of the fixation probabilities based on luminance and texture contrast.

Spatial bias salience

As a next step we investigated to what degree the fixation of bubbles can be predicted by a spatial bias towards the screen center (Tatler, 2007) and the statistics of saccade length and orientation (Bahill et al., 1975). Fig. 4.13 shows the structure of a generative model based on bubble positions and on the parameters of the Gaussian window used for bubble construction (bubble masks), global fixation statistics (central bias), and saccade statistics. Using the specific bubble locations as input to the model is necessary to account for the strong fixation preference towards bubbles found in the experimental data, the very purpose of using bubbles. The fixation and saccade bias maps are derived from empirical data recorded in a previous study of our laboratory (Walter A (2006) Bachelor's thesis. Walter showed images of urban scenes/man-made objects, natural images, fractals, and pink noise images under a free viewing condition to 27 participants. We pooled over all her data from all of these categories.). The fixation bias map contains the distribution of fixations in absolute (screen) coordinates; the coordinates of fixations relative to their preceding fixations form the saccade bias map. For each trial, both maps are computed and convolved with a Gaussian kernel, with a standard deviation of 0.5° and then normalized to integral of one. Finally, we average across trials weighting each trial equally independent of the number of fixations made.

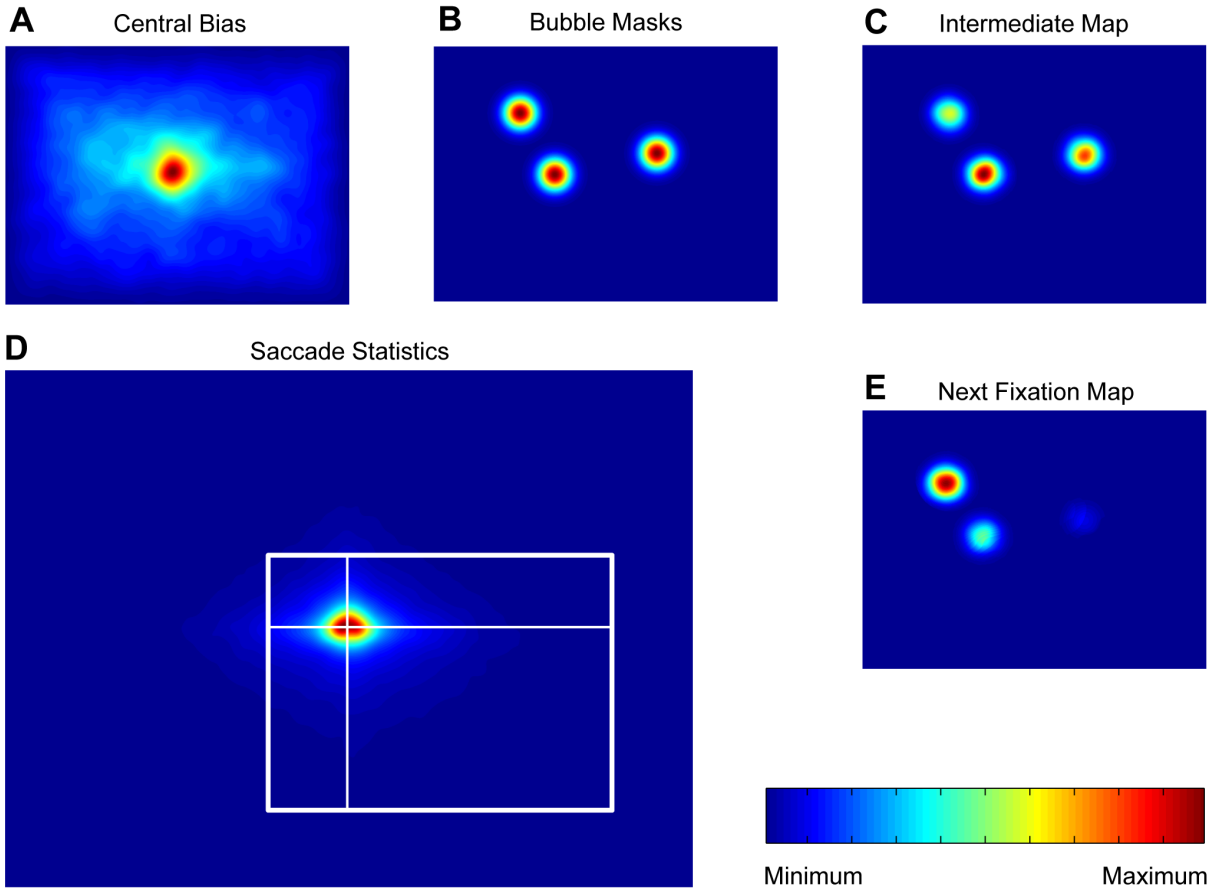


Figure 4.13 Simulation of fixation trajectories based on spatial biases.
Spatial bias salience was computed from simulated fixation trajectories based on the central bias of fixations, saccade statistics, and bubble positions. Given the current fixation location, the next fixation is generated by, first, multiplying the central bias map (A) with the bubble position map (B). Second, the resulting intermediate map (C) is multiplied with the probability distribution over saccade vectors (D) centered at the current fixation. The next fixation is then sampled from the resulting map (E). For example, assuming a fixation of the upper left bubble in panel C, the multiplication (indicated by the white coordinate frame) of the intermediate map (C) and saccade statistics (D) results in the depicted next fixation map (E). Repeating this sampling procedure resulted in the simulated fixation trajectory.

Based on the three maps, we simulate gaze trajectories of 75 virtual participants in 280 trials each in close analogy to the actual experiments. For each simulated trial, the global stimulus independent fixation map and the stimulus specific bubble mask are combined by point-wise multiplication. This combination results in an intermediate map of the spatial bias specific for the position of the bubbles in the stimulus considered (Fig. 4.13C). Next, the saccade bias map is combined with the intermediate map by first aligning the center of the saccade map with the last fixation location (or the screen center for the first fixation within a trial), then multiplying both maps point-wise and normalizing the result to integral one (Fig. 4.13E). The next simulated fixation is then randomly drawn from that probability distribution. This procedure is repeated until as many simulated fixations are drawn for the simulated trial as were made in the corresponding original trial.

From the simulated data we obtain a scalar measure for the fixation probability of each bubble, independent of the task and the spatial structure of the respective full field images. Instead, this spatial bias salience is based solely on the spatial position of the bubbles and the global properties of fixation points and saccades.

Bubble information

To characterize task dependent influences on fixation behavior, we derive a scalar measure for the information a bubble contains with respect to a classification task. First, we assume that each individual bubble is associated with a probability distribution that captures how likely the subjects are to decide for each stimulus class (***response distribution***). If this distribution is flat, the bubble contains no information relevant for classification, and performance of subjects viewing only this bubble would be at chance level. If one of its components is one and all others are zero, then the bubble contains maximal information. This is captured by the entropy of a bubble's response distribution. If $I(B)$ denotes the information content and $P_R(B)$ denotes the response distribution of bubble B, with entropy $E(P_R(B))$, then

$$I(B) = E_{max} - E\big(P_R(B)\big) \tag{5}$$

where E_{max} denotes the maximal entropy that can occur for probability distributions like $P_R(B)$ and depends only on the number of degrees of freedom of $P_R(B)$. For tasks *expression*, *gender*, *influence,* and *space,* E_{max} is 2, 1, 1, and 1, respectively.

Second, along the same lines we assume that our participants' responses to a stimulus S are independent and identically distributed according to the response distribution of the stimulus. In the case of a stimulus S composed of a single bubble B the distribution of observed answers is an estimate of $P_R(B)$. To estimate the empirical saliences of single bubbles from measured classification responses to stimuli composed of several bubbles, we need to make an assumption on how the response distributions for single bubbles are related to the joint response distribution of a stimulus containing those bubbles. Here, we assume optimal probabilistic integration of the independent response distributions of single bubbles (**p-model**). We describe the response distribution $P_R(S)$ of a stimulus $S = \{B_1, \ldots, B_n\}$ by the function Z operating on the individual response distributions $P_R(B_1), \ldots, P_R(B_n)$.

$$Z\big(P_R(B_1), \ldots, P_R(B_n)\big) = P_R(S) \tag{6}$$

We call Z the information integration function. It integrates the response distributions of single bubbles independent of the bubbles' absolute position or their relative arrangement. Furthermore, it does not relate to the visual content of the bubbles. It is defined as

$$Z\big(P_R(B_1), \ldots, P_R(B_n)\big)[c] = \frac{1}{\omega} \prod_{i=1,\ldots,n} P_R(B_i)[c] \tag{7}$$

where $\omega = \sum_d \prod_{i=1,\ldots,n} P_R(B_i)[d]$, with the summation over different stimulus classes d, is the appropriate normalization. Z is formally derived by writing the probability for a stimulus $S = \{B_1, \ldots, B_n\}$ to be of class c, $P_R(S)[c]$, in terms of the corresponding probabilities for the individual bubbles in S to be of class c, $P_R(B_i)[c]$ under the assumption that the individual bubbles are independent.

For each stimulus, we can formulate an equation like (6). Hence for each task, we can formulate as many equations like (6) as there are stimuli in that task. These equations operate on response distributions. Each equation can, however, be transformed into a set of scalar equations by considering the different components of the response distributions (probabilities for the different classes) separately. This yields 1791, 600, 588, and 585 equations for the tasks *expression*, *gender*, *influence* and *space*, respectively (*expression* has four instead of two response possibilities, yielding more scalar equations). This contrasts with 282, 94, 88, and 89 free parameters in the four tasks, equaling the number of bubbles used for stimulus

construction in these tasks, multiplied by the number of possible responses minus 1. We solve this over a determined system of non-linear equations by a maximum likelihood method. Details of this fitting procedure are given in the supplementary text in section 4.6.1. Finally, we determine estimated bubble information from the estimated response distributions of single bubbles according to equation (5).

Correlation analysis
We employ pair-wise correlation analyses (Pearson's correlations) to address the net effect of individual predictor variables. To address how well a linear combination of the stimulus dependent salience, the bubble information, and the spatial bias salience of each bubble can explain the attention it attracts, as reflected by the empirical salience values, we employ a multivariate model. Finally, to correlate one predictor with empirical salience while controlling for the effect of all other predictors, we use semi-partial correlations. For example, when we are interested in the correlation of bubble information and empirical salience controlled for the influence of stimulus-based salience and spatial bias salience, we consider the residuals of a multivariate correlation (with intersection) of stimulus-based salience and spatial bias salience with bubble information. These residuals are the differences between the prediction of the multivariate model and the actual bubble information values. We now correlate these residuals with empirical salience. The result is called the semi-partial correlation coefficient of bubble information and empirical salience.

For all, the simple pair-wise correlation analysis, the multivariate correlation and the semi-partial correlation analysis, we used the log transform of the predictor variables and the log transform of empirical salience. This standard practice (Tabachnick & Fidell, 2007) has the main effect of making the distributions of the individual variables more normal.

4.6 Supplementary material

Description of the computation of bubble information and of other models of information integration.

4.6.1 Fitting single bubble answer distributions

Let S be a stimulus made up of a number of bubbles $B_1^S, \dots, B_m^S$. If it was presented to k participants, k classification responses are available. By assumption (compare Methods), these k responses are drawn from the same underlying stimulus response distribution. We denote that underlying response distribution by s. The k responses yield an estimate of s. This estimate is a random variable following a multinomial distribution. The parameters of that distribution are given by the underlying stimulus response distribution s. By assumption, we know that s is related to the response distributions of the individual bubbles that make up the stimulus by an information integration model I. Let $b_1^S, \dots, b_m^S$ denote the underlying response distributions of the individual bubbles of stimulus S. It holds: $s = I(b_1^S, \dots, b_m^S)$. We want to estimate $b_1^S, \dots, b_m^S$ based on the measured responses of our participants. Estimating $b_1^S, \dots, b_m^S$ is harder than estimating the underlying stimulus response distribution s, since we cannot directly observe realizations of the corresponding random variables. However, we can estimate $b_1^S, \dots, b_m^S$ by maximizing the likelihood of the observed classification responses with respect to the response distributions of individual bubbles. In our case, the maximum likelihood estimates of $b_1^S, \dots, b_m^S$ are those that minimize the sum-of-squares error between the classification responses predicted from them and the actual classification responses. We perform this maximum likelihood fit for every task separately, always fitting all bubbles of

one task simultaneously. We used a genetic algorithm to minimize the sum-of-squares (MatLab's *ga.m* function, Mathworks, Natick, MA, USA). We used the genetic algorithm instead of gradient based methods, since we initially experimented with a multitude of information integration models, some of them possessing non differentiable structure. The fitting procedure was repeated 15 times (each time with different random starting values for the genetic algorithm), and the best fit was chosen.

4.6.2 Comparing the predicted stimulus response distributions with the measured stimulus response distributions

We used the p-model to fit the response distributions for single bubbles. The bases for these fits are the stimulus response distributions measured throughout the experiment on stimuli with 1 to 5 bubbles. As an additional test for the validity of the p-model we predicted the stimulus response distributions from the fitted response distributions for single bubbles using the p-model. We give the absolute difference of stimulus information of the predicted response distribution and stimulus information of the measured response distribution averaged over all stimuli within one task to indicate the fitting error. As a lower bound, we compute the error that would be expected if the predicted stimulus response distributions were the true response distributions. Under this assumption, the measured response distributions are built up from samples drawn from the predicted response distributions. Hence, we computed the average absolute difference between stimulus information of the "resampled" response distribution and the assumed response distribution.

4.6.3 Validating the fit

Because of the random nature of the fitting algorithm and the measured stimulus answer distributions, we investigated the quality of the estimation process. We investigated the bias and variance of the estimated underlying bubble distributions using simulations of the experimental process and the fitting procedure, based on random initial values for the underlying bubble distributions (these values were, however, chosen to be close to the estimated underlying bubble answer distributions to achieve maximal comparability). The simulation proceeded as follows. First, we choose underlying response distributions for single bubbles. Second, we use the integration model to compute the response distributions for whole stimuli. Third, we resample those distributions, taking into account the number of subjects that actually saw each of the stimuli in the real experiment. Fourth, we perform the above described fit to obtain estimates of the underlying response distributions for single bubbles based on the resampled responses to whole stimuli.

These four steps are repeated 30 times, yielding pairs of underlying bubble response distributions and their estimates for every repetition. We are interested in the bias and variance of the entropy of the estimated response distributions, because the entropy is what we correlate with empirical salience and our other measures. Hence, for each pair of the underlying "true" bubble response distribution and its estimate, we compare the entropies of the distributions. Figure 4.14 shows how the entropies of the estimated bubble response distributions relate to the entropy of the true response distributions for different values of the true entropy for the *expression* task. The estimates appear unbiased, and the variance is moderate. The situation for the other tasks is qualitatively the same.

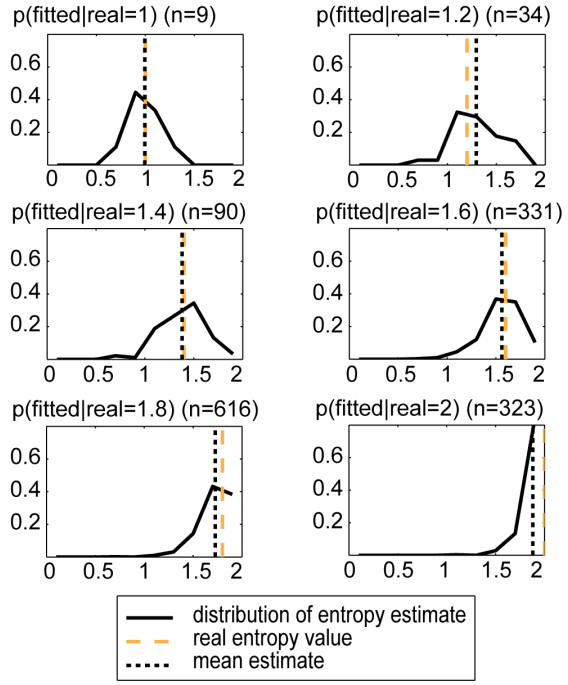


Figure 4.14 Distributions of bubble entropy estimates for different initial bubble entropies (see text) for the expression task. Each plot accumulates the data for all initial bubble entropies in a small interval around the displayed value. The distributions were obtained from simulations (see text).

4.6.4 Other models of information integration

We assume a probabilistic integration model but also considered two other models of information integration. First, a local model captures stimulus information by the maximally informative bubble (max-model). This model always selects the bubble with the highest information. Formally:

$$Z_{Max}(P_R(B_1), \ldots, P_R(B_n)) = P_R(\underset{B_i, i=1..n}{\operatorname{argmax}} I(B_i))$$

Where P_R denotes the response distributions, B_i denotes an individual bubble and $I(B_i)$ denotes its information content. Z_{Max} is a function on the response distributions of the individual bubbles (just like in the case of the p-model) and its value is the response distribution with maximal information.

Second, we considered a global model that differs from the probabilistic model by capturing contra factual evidence for the different choice possibilities (ce-model). Formally:

$$Z_c(P_R(B_1), \ldots, P_R(B_n))[c] = \frac{1}{\omega}(1 - \prod_{i=1,\ldots,n}(1 - P_R(B_i)[c]))$$

We evaluate these models using the method employed to generate Figure 4.2 (see Results). We give the squared error between the same condition and the model prediction (divided by the maximal information of the task) averaged over bubble numbers and tasks. The average squared relative error for the max-model is 0.044276. For the ce-model the average error is 0.23654. The error for the p-model is 0.02518.

4.7 References

Açık A, Onat S, Schumann F, Einhäuser W, König P (2009) Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories. Vision Res 49: 1541–1553.

Antes JR (1974) The time course of picture viewing. J Exp Psychol 103: 62–70.

Avidan G, Harel M, Hendler T, Ben-Bashat D, Zohary E, Malach R (2002) Contrast sensitivity in human visual areas and its relationship to object recognition. J Neurophysiol 87: 3102–3116.
Baddeley RJ, Tatler BW (2006) High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. Vision Res 46: 2824–2833.
Bahill AT, Adler D, Stark L (1975) Most naturally occurring human saccades have magnitudes of 15 degrees or less. Invest Ophthalmol 14: 468–469.
Betz T, Kietzmann TC, Wilming N, König, P (2010). Investigating task-dependent top-down effects on overt visual attention. J Vis 10: 3 1-14.
Brockmann D, Geisel T (1999) Are human scanpaths Levy flights? Artificial Neural Networks 1: 263–268.
Buswell GT (1935) How people look at pictures. A study of the psychology of perception in art. Chicago: University of Chicago Press.
Calder AJ, Young AW, Keane J, Dean M (2000) Configural information in facial expression perception. J Exp Psychol Hum Percept Perform 26: 527–551.
Carmi R, Itti L (2006) Visual causes versus correlates of attentional selection in dynamic scenes. Vision Res 46: 4333–4345.
Chen X, Zelinsky GJ (2006) Real-world visual search is dominated by top-down guidance. Vision Res 46: 4118–4133.
Cristino F, Baddeley R (2009) The nature of the visual representations involved in eye movements when walking down the street. Vis Cogn 17: 880–903.
Ehinger KA, Hidalgo-Sotelo B, Torralba A, Oliva A (2009) Modeling Search for People in 900 Scenes: A combined source model of eye guidance. Vis Cogn 17: 945–978.
Einhäuser W, König P (2003) Does luminance-contrast contribute to a saliency map for overt visual attention? Eur J Neurosci 17: 1089–1097.
Einhäuser W, Kruse W, Hoffmann KP, König P (2006) Differences of monkey and human overt attention under natural conditions. Vision Res 46: 1194–1209.
Einhäuser W, Rutishauser U, Frady EP, Nadler S, König P, Koch C (2006) The relation of phase noise and luminance contrast to overt attention in complex visual stimuli. J Vis 6: 1148–1158.
Einhäuser W, Spain M, Perona P (2008) Objects predict fixations better than early saliency. J Vis 8: 18 11–26.
Einhäuser W, Rutishauser U, Koch C (2008) Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. J Vis 8: 2 1–19.
Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1: 1–47.
Freiwald W, Tsao D, Livingstone M (2009) A face feature space in the macaque temporal lobe. Nat Neurosci 12: 1187 - 1196.
Frey HP, König P, Einhäuser W (2007) The role of first- and second-order stimulus features for human overt attention. Percept Psychophys 69: 153–161.
Frey HP, Honey C, König P (2008) What's color got to do with it? The influence of color on visual attention in different categories. J Vis 8: 6 1–17.
Gao D, Vasconcelos N (2009) Decision-theoretic saliency: computational principles, biological plausibility, and implications for neurophysiology and psychophysics. Neural Comput 21: 239–271.
Gosselin F, Schyns PG (2001) Bubbles: a technique to reveal the use of information in recognition tasks. Vision Res 41: 2261–2271.
Gosselin F, Schyns PG (2002) RAP: a new framework for visual categorization. Trends Cogn Sci 6: 70–77.
Gosselin F, Schyns PG (2004) No troubles with bubbles: a reply to Murray and Gold. Vision Res 44: 471–477; discussion 479–482.

Hamker FH (2004) A dynamic model of how feature cues guide spatial attention. Vision Res 44: 501–521.
Hooge I, Over E, van Wezel R, Frens M (2005) Inhibition of return is not a foraging facilitator in saccadic search and free viewing. Vision Res 45: 1901–1908.
Hubel DH, Wiesel TN (1961) Integrative action in the cat's lateral geniculate body. J Physiol 155: 385–398.
Itti L, Koch C (2001) Computational modelling of visual attention. Nat Rev Neurosci 2: 194–203.
Itti L, Baldi P (2009) Bayesian surprise attracts human attention. Vision Res 49: 1295–1306.
Jansen L, Onat S, König P (2009) Influence of disparity on fixation and saccades in free viewing of natural scenes. J Vis 9: 29 21–19.
Kanan C, Tong MH, Zhang L, Cottrell GW (2009) SUN: Top-down saliency using natural statistics. Vis Cogn 17: 979–1003.
Kanan C, Cottrell G (2010). Robust classification of objects, faces, and flowers using natural image statistics. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2010: 2472-2479.
Kanwisher N, Yovel G (2006) The fusiform face area: a cortical region specialized for the perception of faces. Philos Trans R Soc Lond B Biol Sci 361: 2109–2128.
Kienzle W, Franz MO, Schölkopf B, Wichmann FA (2009) Center-surround patterns emerge as optimal predictors for human saccade targets. J Vis 9: 1–15.
Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol 4: 219–227.
Krieger G, Rentschler I, Hauske G, Schill K, Zetzsche C (2000) Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. Spat Vis 13: 201–214.
Land M, Mennie N, Rusted J (1999) The roles of vision and eye movements in the control of activities of daily living. Perception 28: 1311–1328.
Land MF, Hayhoe M (2001) In what ways do eye movements contribute to everyday activities? Vision Res 41: 3559–3565.
Li Z (2002) A saliency map in primary visual cortex. Trends Cogn Sci 6: 9–16.
Mackworth NH, Morandi AJ (1967) The gaze selects information details within pictures. Perception 11: 547–551.
Malcolm GL, Lanyon LJ, Fugard AJ, Barton JJ (2008) Scan patterns during the processing of facial expression versus identity: an exploration of task-driven and stimulus-driven effects. J Vis 8: 8 1–9.
Mannan SK, Ruddock KH, Wooding DS (1996) The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. Spat Vis 10: 165–188.
Murray RF, Gold JM (2004) Troubles with bubbles. Vision Res 44: 461–470.
Navalpakkam V, Itti L (2005) Modeling the influence of task on attention. Vision Res 45: 205–231.
Navalpakkam V, Itti L (2007) Search goal tunes visual features optimally. Neuron 53: 605–617.
Olshausen BA, Field DJ (2005) How close are we to understanding v1? Neural Comput 17: 1665–1699.
Parkhurst D, Law K, Niebur E (2002) Modeling the role of salience in the allocation of overt visual attention. Vision Res 42: 107–123.
Parkhurst DJ, Niebur E (2003) Scene content selected by active vision. Spat Vis 16: 125–154.
Parkhurst DJ, Niebur E (2004) Texture contrast attracts overt visual attention in natural scenes. Eur J Neurosci 19: 783–789.

Peters RJ, Iyer A, Itti L, Koch C (2005) Components of bottom-up gaze allocation in natural images. Vision Res 45: 2397–2416.
Privitera CM, Stark LW (2000) Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations. IEEE Trans Pattern Anal Mach Intell 22: 970–982.
Rao RP, Zelinsky GJ, Hayhoe MM, Ballard DH (2002) Eye movements in iconic visual search. Vision Res 42: 1447–1463.
Reinagel P, Zador AM (1999) Natural scene statistics at the centre of gaze. Network 10: 341–350.
Rizzolatti G, Riggio L, Dascola I, Umilta C (1987) Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. Neuropsychologia 25: 31–40.
Rolls ET, Baylis GC (1986) Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. Exp Brain Res 65: 38–48.
Rothkopf CA, Ballard DH, Hayhoe MM (2007) Task and context determine where you look. J Vis 7: 16 11–20.
Saal H, Nortmann N, Krüger N and König P (2006) Salient image regions as a guide for useful visual features. In Proceedings of the IEEE Systems, Man and Cybernetics Society. Conference on Advances in Cybernetic Systems, AICS 2006.
Sere B, Marendaz C, Herault J (2000) Nonhomogeneous resolution of images of natural scenes. Perception 29: 1403–1412.
Schumann F, Açık A, Onat S, König P (2007) Integration of different features in guiding eye-movements [Abstract]. Proceedings of the 7th Meeting of the German Neuroscience Society/31th Göttingen Neurobiology Conference.
Schumann F, Einhäuser-Treyer W, Vockeroth J, Bartl K, Schneider E, König P (2008) Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. J Vis 8: 12 11–17.
Schwartz O, Bayer HM, Pelli D (1998). Features, frequencies, and facial expressions [Abstract]. Investigative Ophthalmology and Visual Science 39: 173.
Schyns PG, Bonnar L, Gosselin F (2002) Show me the features! Understanding recognition from the use of visual information. Psychol Sci 13: 402–409.
Smith T, Henderson J (2009) Facilitation of return during scene viewing. Vis Cogn 17: 1083–1108.
Steinwender J (2005) Context dependency of overt attention in natural scenes. Bachelor's thesis, University of Osnabrück.
Tabachnick BG, Fidell LS (2007) Using Multivariate Statistics Pearson Education, Inc.
Tanaka K (1996) Inferotemporal cortex and object vision. Annu Rev Neurosci 19: 109–139.
Tatler BW, Baddeley RJ, Gilchrist ID (2005) Visual correlates of fixation selection: effects of scale and time. Vision Res 45: 643–659.
Tatler BW, Baddeley RJ, Vincent BT (2006) The long and the short of it: spatial statistics at fixation vary with saccade amplitude and task. Vision Res 46: 1857–1862.
Tatler BW (2007) The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. J Vis 7: 1–17.
Tatler B, Vincent B (2008) Systematic tendencies in scene viewing. Journal of Eye Movement Research 2: 5.
Tatler B, Vincent B (2009) The prominence of behavioural biases in eye guidance. Vis Cogn 99999: 1–26.
Tootell RB, Silverman MS, Switkes E, De Valois RL (1982) Deoxyglucose analysis of retinotopic organization in primate striate cortex. Science 218: 902–904.
Torralba A (2003) Modeling global scene factors in attention. J Opt Soc Am A Opt Image Sci Vis 20: 1407–1418.

Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, Marcus DJ, Westerlund A, Casey BJ, Nelson C (2009) The NimStim set of facial expressions: judgments from untrained research participants. Psychiatry Res 168: 242–249.

Tsotsos JK, Culhane SM, Wai WYK, Lai Y, Davis N, Nuflo F (1995) Modeling visual attention via selective tuning. Artificial Intelligence 78: 507–545.

Turano KA, Geruschat DR, Baker FH (2003) Oculomotor strategies for the direction of gaze tested with a real-world activity. Vision Res 43: 333–346.

Underwood G, Foulsham T, van Loon E, Humphreys L, Bloyce J (2006) Eye movements during scene inspection: A test of the saliency map hypothesis. Eur J Cogn Psychol 18: 321–342.

Vincent BT, Baddeley R, Correani A, Troscianko T, Leonards U (2009) Do we look at lights? Using mixture modelling to distinguish between low- and high- level factors in natural image viewing. Vis Cogn 17: 856–879.

Vinette G, Gosselin F, Schyns PG (2004) Spatio-temporal dynamics of face recognition in a flash: it's in the eyes. Cogn Sci 28: 289–301.

Walter A (2006) Baseline Study on Overt Visual Attention. Bachelor's thesis, University of Osnabrück.

Walther D, Koch C (2006) Modeling attention to salient proto-objects. Neural Networks 19: 1395–1407.

Wolfe JM, Horowitz TS (2004) What attributes guide the deployment of visual attention and how do they do it? Nat Rev Neurosci 5: 495–501.

Yarbus AL (1967) Eye movements and vision. New York: Plenum.

Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW (2008) SUN: A Bayesian framework for saliency using natural statistics. J Vis 8: 32 31–20.

5. General discussion

5.1 Discussion

Summary of results
This thesis investigates influences of stimulus-dependent factors and contextual factors on visual processing. To gain a better understanding of influences of temporal context on early visual processing we investigated neuronal processing in V1 using the method of optical imaging with voltage sensitive dyes, which allowed us to measure neuronal population activity across several millimeters of cortical area with high temporal resolution.

To characterize the cortical layout in V1 concerning the mapping of orientation, previous to further investigation, we used smoothly moving grating stimuli. Optical responses to moving gratings were measured for a long time (VSDI measurements: Shoham et al., 1999), but the obtained responses held ambiguous stimulus information, as they did not show signatures of stimulus motion. Thus, prior to the investigation of influences of contextual factors, we investigated systematically to what extend properties of the current stimulus are reflected in V1 population responses, under stimulation with moving gratings (Study 1, Section 2). In addition to the strong and well-known response to the orientation of the grating, we were able to show simultaneous encoding of stimulus motion, across the population of V1 neurons. While the orientation of a moving grating led to a stationary pattern on the cortex, the motion of the grating's individual stripes led to propagating waves. The retinotopic waves traversed the orientation map homogeneously, which indicated that orientation and location information are processed independently. The signal magnitude of the propagating waves was two orders of magnitude smaller than that of the stationary orientation response. The critical step that allowed us to visualize these waves in the recorded data was the use of singular-value-decomposition, as an alternative method of analysis. Still, we were only able to resolve these signatures in our highest signal-to-noise ratio recordings. The results provide evidence that, at the first cortical stage, the two stimulus features, location and orientation, are processed independently across the neuronal population (multiplexing).

To investigate the influence of the context of previous stimulation on the processing of the current stimulus (Study 2, Section 3), we measured responses to abrupt stimulus changes in cat V1. We were able to find evidence for two different encoding schemes: We found ongoing encoding of stimulus features when stimuli were presented in fast succession (at 33 Hz), consistent with earlier studies (Benucci et al., 2009). However, slower stimulus sequences (10 Hz) revealed an alternative encoding scheme. Population tuning in V1 no longer represented the complete stimulus content, but those orientations that were newly added or removed. Thus, we found two characteristically different encoding schemes: ongoing encoding after short presentation (30 ms) of the preceding stimulus, and encoding of difference after longer presentation (100 ms). We found this neuronal behavior for artificial grating stimuli, as well as for filtered natural stimuli that contained a range of different spatial frequencies. We wanted to test whether our results are contingent on the chosen recording method of VSDI, which is highly sensitive to subthreshold activity. Therefore, we conducted additional experiments in which we recorded spiking activity of individual neurons, using electrodes. There, we were able to confirm a representation of difference in superthreshold responses of individual neurons. Further, to test whether the results are contingent on presentation of stimulus sequences, we also used isolated presentation of stimulus transitions between only two stimuli. There, we confirmed our results. Yet further, to test whether change coding is only found for a presentation duration of 100 ms, we also tested a duration of 500 ms in an additional experiment. After such longer presentation of the preceding stimulus, we found

representation of difference also. Thus, we showed that representation of difference is not exclusive to a frequency of 10 Hz, and the results suggest that it can generally be found for lower temporal frequencies. To explore the underlying mechanisms of change coding, we assessed possible contributions of adaptation and stimulus off-responses, whose signatures we also measured using stimulus sequences. We then proposed that encoding of difference involves the interplay of these two mechanisms, both depending on stimulus timing also. Overall, we found that the previous stimulus strongly influences the processing of current stimulus content when it is presented for relatively long durations of 100 ms and 500 ms, resulting in an encoding of the difference between past and present sensory input.

Finally, to better understand how contextual factors interact with stimulus-driven factors in visual sampling behavior, we conducted a study to quantify the influence of the task context in relation to the influence of spatial biases and low-level image features on human eye movement control (Study 3, Section 4). We tested visual stimuli composed of image patches in a variety of tasks. Obtaining eye-tracking and behavioral data (button presses in forced choice classification tasks) from 75 participants and, subsequently, fitting them to an information theoretic model, allowed us to take a quantitative approach to task-dependent, contextual effects. We found that all three factors made individual contributions towards the selection of attended locations, with little redundancy among them. On average, spatial viewing biases, which only depend on the location of a patch, made the strongest contribution. Second, on average, were factors related to the task context, which were based on how informative the content of an image patch is in a given task. Third were low-level image features, which, on average, contributed least to eye movements, yet they also showed a significant individual contribution. All three factors made independent and significant contributions to eye movements and together explained a substantial fraction of the variance in how often different image patches were selected for fixation. The results emphasize the need to integrate all three, task context dependent factors, spatial biases, and stimulus-driven factors in a model of human eye movements.

The following sections discuss connections of the presented work to related concepts and research, address possible consequences, and provide outlooks on possible future work. The results of the individual studies were already discussed separately (in sections 2.5, 3.5, and 4.4).

The choice of stimuli

Whether results obtained using artificial stimuli generalize to natural conditions is an interesting topic. When investigating the V1 representation of stimulus motion (Study 1) we used artificial gratings to obtain a high signal to noise ratio. The detected signals carrying information about stimulus motion were two orders of magnitude smaller than the orientation responses and only detectable in our highest quality recordings (were the limiting factor is likely of technical origin, as discussed in 2.5.1). It remains to be investigated whether the independent simultaneous encoding of the two stimulus features generalizes to stimuli that are more natural. Natural stimuli, however, usually provide smaller response amplitudes than gratings (see Study 2, and e.g. Onat et al., 2011). A methodological advancement could provide a higher signal to noise ratio than is currently achievable. This might enable the investigation of parallel encoding of orientation and local contrast using stimuli that are more natural.

When investigating the influence of a previous stimulus onto the processing of the next (Study 2), we used different types of stimuli, ranging from artificial gratings to widely filtered

natural stimuli. We found 10-Hz-change-coding independent of the stimulus type. Although we modified the natural stimuli to obtain dominant vertical and horizontal frequencies, necessary for parametric control in our experiments, these stimuli share important properties with natural stimuli, that the grating stimuli do not: The phase spectrum and the frequency spectrum along the spatial frequency dimension is left untouched in the filtering process. In the broadly filtered stimuli, additionally, we presented context from other orientations together with the dominant orientations. This reduced the amplitude of the effect but not the qualitative result. Our results show that change coding is not contingent on the presentation of abstract simple stimuli, but indeed generalizes to scenarios in which complex stimuli are used, that are comprised of contours from natural images. These results further emphasize the need to consider cortical encoding of difference in theoretical frameworks that aim to predict responses to natural stimuli.

Furthermore, an interesting topic is the possible interaction of the two different encoding schemes for stimuli with orientation changes at multiple timescales. In our presented data, we found either ongoing encoding or encoding of difference, in dependence of the presentation duration of the stimuli. In the natural input, different changes could take place on different timescales in an interleaved fashion. Within the same visual stream of input, one orientation can change after a short time, while another orientation can change after a longer time. This makes it reasonable to hypothesize that both encoding schemes can be found in the cortex simultaneously under such kind of stimulation. It could be useful to investigate empirically such a possible co-occurrence of different coding schemes and their interaction, to specify response predictions to complex natural movies.

The main discussion about the use of natural stimuli versus artificial stimuli concerns the use of these stimulus types in the investigation of early visual areas (Felsen & Dan, 2005; Rust & Movshon, 2005). In the investigation of functions involving higher areas, such as eye movement control, natural images are often used. Nevertheless, also in our investigation of the influence of the task context on human eye movements, we used images that are modified. There, we extracted image patches from photographs of forest scenes and faces that reveal only a limited controlled subset of the image content. We argued (in detail in the study in section 4.4.3) for a generalizability of our results to full field scenes. By using local image patches, we created individual perceptual units (individual pieces of visual information), which were required, in the developed methodology, to objectively quantify task-dependent information content of particular image regions (using the degree of agreement between participants with respect to a classification task).

However, the controlled use of pieces of information in the form of image patches is a versatile and powerful approach to investigate questions involving the use of task-related information. In particular, the method and stimuli developed in the presented study (Study 3), as well as the measured data, provided a foundation for further investigations. Sinke (2008) adapted the method and stimuli to also investigate the influence of the gist of a scene (provided by low-spatial-frequency global image content, presented together with the local image patches) on the selection of fixation locations. Zennig (2009) and Herkenhoff (2012) investigated cultural differences in the use of information in scene viewing and classification, between German and Japanese participants, with a special focus on faces (based on the presented data from German participants and additionally obtained data from Japanese participants).

Generally, in each empirical study the question where to place the balance between complex natural conditions and well-controlled laboratory stimuli has to be addressed anew. In Study 1, we used gratings, a stimulus type, which is known to be highly effective in driving

responses in V1, to detect responses to an additional stimulus property. In Study 2, we used classical grating stimuli, as well as filtered natural stimuli. The orientation-sensitive filtering was necessary to investigate changes in orientation content, in line with previous related research (e.g. Ringach et al., 1997; Benucci et al., 2009). In study 3, we used image patches extracted from natural images to evaluate task-dependent information content. In the presented experimental paradigms, we chose stimuli suitable for the respective questions addressed, striking compromises between the ecological validity of natural stimuli and the controllability of artificial stimuli.

Perceptual experience

Neuronal activity in V1 can sometimes directly be linked to perceptual phenomena. In an earlier study, Jancke et al. (2004) found cortical correlates of perceptual illusion in V1. In particular, a stimulus without real motion, that is known to create a perceptual motion illusion, created cortical activity indistinguishable from that to real stimulus motion (Jancke et al., 2004). It is thus reasonable to hypothesize that the two different coding schemes observed in Study 2 (ongoing encoding and encoding of difference) also have perceptual correlates.

There is a possible connection of the two different coding schemes to the flicker fusion frequency, which is the frequency above which individual images are not perceived separately anymore, but instead, as a smoothly changing stream of input. The flicker fusion frequency for cats is in a range of 45-70 Hz (Taravella & Clark, 1963; Schwartz & Cheney, 1966). Although the experimental conditions in our setup are different to the ones in which these thresholds were measured (the most important difference being that the cats were anesthetized in the experiments reported here), the values point to possible perceptual correlates of the qualitatively different coding schemes reported (ongoing encoding at 33 Hz and encoding of difference at 10 Hz), namely perception of smooth motion and flicker. This connection would need to be investigated, possibly by specification of the relevant timescales of ongoing encoding versus encoding of difference in awake animals, also including other areas (e.g. cat area 17).

Predictive coding

Encoding of differences in V1 is compatible with theories of predictive coding (Friston, 2005; Clark, 2013). Clark (2013) concludes that the predictive coding approach is the "best clue yet to the shape of a unified science of mind and action." Predictive coding theories inherently combine action and perception, since in their theoretic framework predictions about future input are propagated down to the earliest levels of the perceptual hierarchy (e.g. Rao & Ballard, 1999). The organism is assumed to possess knowledge about the effect of its own actions on possible input, which influences its predictions about it, in dependence of its actions.

Further, predictive coding theories propose that deviations from the cortically generated predictions about the input are propagated up the visual hierarchy as error signals. In this framework, assuming that the default prediction is that no change occurs, representations of change can be interpreted as instances of such error signals. It is unknown how interactions with high-level mechanisms affect representation of differences in awake animals. Considering the temporal statistics of natural input, as discussed in the introduction, the prediction that no change occurs in orientation on a timescale of 100 ms and longer is ecologically worthwhile, and in accordance with natural input statistics (Kayser et al.,

2004).[18] Thus, even in lack of voluntary or attentional top-down influences, the first cortical processing stage may already incorporate predictive coding mechanisms, dependent on input timing.

Cortical encoding of difference in connection with eye movements
Given that under natural conditions visual input is actively selected by eye movements, it is an interesting question whether cortical encoding of difference could be observed in combination with eye movements. We found ongoing encoding for fast changes (33 Hz), and encoding of difference for slower changes (10 Hz). We proposed that these two processing schemes, observed for the different presentation durations, complement two different modes of sensory sampling.

Specifically, the first case concerns the processing across different fixations. The change coding that we observed in our experimental setup, in anaesthetized cats, can be described as a difference operation over time between two visual inputs that originate from the same location in the environment. However, the situation might be different in awake animals. The spatial mapping in V1 is in reference to the retina and not in reference to the visual field. Thus, when eye movements are made, a given cortical location will receive information from different locations in the environment. Thus, local operations in the cortex are, in principle, sufficient to compare information originating from different locations in the environment. Accordingly, change coding could not only be used to process changes of visual information in time, but also to process differences in visual information across space (Gawne & Woods, 2003). In the study (section 3.5.2), we propose that the characteristically high correlation of activity in the retina in the beginning of a fixation (as opposed to later during fixation; Desbordes & Rucci, 2007), i.e. at low temporal frequencies, might be compensated at the cortical level by the observed difference representation, effectively reducing redundancies in the input. In general, encoding of difference across successive fixations could play a role in the perception of spatial patterns and in the detection of salient structures, by reducing responses to similar input across fixations, while enhancing responses to dissimilar input across fixations.

The second case concerns the processing during fixation. During fixation, microscopic eye movements, which operate at higher temporal frequencies, decorrelate retinal activity (Kuang et al., 2012). Thereby, small spatial details in visual structures are emphasized (Rucci et al., 2007; Desbordes & Rucci, 2007; Rucci, 2008; Kuang et al., 2012). Here, the faithful cortical representation of current stimulus content, which is observed at higher frequencies (33 Hz), may result in a direct transmission of the acquired information, which already has reduced correlations due to the microscopic eye movements, further downstream. Thus, the two different processing schemes observed for the different presentation durations (ongoing encoding and encoding of difference) might complement the two different modes of sensory sampling by eye movements (during fixation and immediately after a saccade).

In addition, under natural conditions and in awake animals, additional factors are expected to play a role. In the investigation of contextual factors in eye movement control (Study 3), we showed that task context individually influences the selection of fixation locations, in addition to low-level features and spatial biases. Task context was also shown to affect responses in V1 via top-down signals (Heanny & Schiller PH, 1988; Watanabe et al., 1998; see Posner & Gilbert, 1999 for review of attentional modulation in V1). Task-dependent effects in V1

[18] This applies to the amplitudes of collinear orientations. Phases of orientations, on the other hand, correlate on shorter timescales (Betsch et al., 2004).

include changes in neuronal response properties that allow neurons to carry more information about a stimulus attribute, when it is relevant in the given task (Li et al., 2004). It was also found that practicing a task could lead to long-term changes in V1 neuronal response properties (Schoups et al., 2001). The other way around, activity in V1 can sometimes even be used to predict task performance (Ress et al., 2000). Thus, task context can influence V1 activity not only via the selection of visual input by eye movements, but also via top-down signals. Future research needs to elaborate, how such top down influences interact with the observed encoding of difference across time, in addition to eye movements.

Simultaneous and independent processing of different stimulus features

Of the different stimulus features that are processed in the primary visual cortex, in Study 2, we concentrated on stimulus orientation. We found encoding of difference in slow (10 Hz) stimulus sequences. From the response we measured at a given time, it is not possible to infer what image was presented at the screen at the corresponding time. Instead, the change that occurred from one image to the next can be inferred. Does V1 possess incomplete information about the current stimulus? In Study 1, we found that the two stimulus dimensions, orientation and local contrast, are represented independently. Although the retinotopic response, encoding local contrast, was two orders of magnitude smaller than the orientation response, the results show that different properties of the stimulus are processed in parallel at the same cortical locations. It is, thus, possible that in parallel to the representation of differences observed for orientation, a faithful representation of current stimulus content exists in the retinotopic representation, which Study 2 was not designed to investigate. Thus, in light of the multiplexing capabilities of V1 (observed in Study 1), it is possible that the orientation response, investigated in Study 2, does not represent all information available in V1. Alternatively, other stimulus properties, such as local contrast, could be processed in parallel, and provide additional stimulus information.

Concluding remarks

It is clear that a concept of perception, that sets the construction of faithful representations as its main function, is simplified and that perceptual processes are influenced by more than the current stimulus alone. To understand the role of context in perceptual processing it can help to conceptualize perception in a broader framework. That is, understanding the function of perception as to aid the generation of beneficial behavior.

In real-world scenarios, sensory input is not arbitrary. If it were, nothing could be gained by integrating information across time and space or across different sources. However, there are dependencies in the world. It was mentioned that natural visual input possesses characteristic statistical properties, such as correlations across time and space (see Simoncelli & Olshausen, 2001; Geisler, 2008). Additionally, there are lawful relationships between action and sensory input (sensorimotor contingencies, O'Regan & Noë, 2001).[19] And, of course, there are also higher-level, complex lawful relationships in the world. Such dependencies determine the entities that provide relevant context in the processing of current stimulus content.

An adaptation of the nervous system to such dependencies allows the use of relevant context. This has advantages for the organism, among which are the following. It allows for predictions about future sensory input (e.g. Rao & Ballard, 1999; Friston, 2005; Clark, 2013). It allows for increasing efficiency in perceptual processing, by using knowledge of lawful relationships (or learned probabilities) to reduce redundancies (Barlow, 1961). Related to the

[19] For example, when the eyes are moved to the left, visual input shifts to the right.

last, it allows selecting (by active sensing) or enhancing (by neuronal processing) those signals that are most informative. On a systemic level, it allows the organism to take guided actions to increase the chances of beneficial outcome. Organisms that learn or have adapted to these dependencies can exploit them and have an evolutionary advantage.

These ideas stress that an influence of context on perceptual processes is not to be seen as an exception or an artifact, interfering with unbiased stimulus processing. Instead, it is vital for the organism that it uses lawful relationships in the world and between the world and its own body, which exist under natural conditions, in the processing of current stimulus content. To be able to observe how perceptual processes have adapted to such lawful relationships, they have to be investigated in conditions in which ecologically relevant relationships are applicable.

In this framework, the immediate past and the task of the animal can provide relevant context. These contextual factors are investigated in this thesis. We have shown a strong effect of temporal context in the primary visual cortex, and found significant independent contribution of the task context on eye movements, in addition to spatial and stimulus-driven factors. The empirical results presented here provide foundations for an improved theoretical understanding of the role of context in perceptual processes.

5.2 References

Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: Rosenblith W (editor) Sensory communication. Cambridge (MA): MIT Press: 217–234.

Benucci A, Ringach DL, Carandini M (2009) Coding of stimulus sequences by population responses in visual cortex. Nat Neurosci 12: 1317-1324.

Betsch B, Einhäuser W, Körding KP, König P (2004) The world from a cat's perspective – statistics of natural videos. Biol Cybern 90: 41–50.

Clark A (2013) Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. Behav and Brain Sci 36: 181-204.

Desbordes G, Rucci M (2007) A model of the dynamics of retinal activity during natural visual fixation. Vis Neurosci 24: 217–230.

Felsen G and Dan Y (2005) A natural approach to studying vision. Nat Neurosci 8: 1643–1646.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360: 815–836.

Gawne TJ, Woods JM (2003) The responses of visual cortical neurons encode differences across saccades. NeuroReport 14: 105–109.

Geisler WS (2008) Visual Perception and the Statistical Properties of Natural Scenes. Annu Rev Psychol 59: 167–92.

Heanny PE, Schiller PH (1988) State dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks. Exp Brain Res 69: 225–44.

Herkenhoff S (2012) Cultural differences of fixation behavior in emotion recognition on faces. Bachelor thesis. University of Osnabrück.

Jancke D, Chavane F, Naaman S, Grinvald A (2004) Imaging cortical correlates of illusion in early visual cortex. Nature 428: 423–426.

Kayser C, Einhäuser W, König P (2004) Temporal correlations of orientations in natural scenes. Neurocomputing 52-54: 117–123.

Kuang X, Poletti M, Victor JD, Rucci M (2012) Temporal encoding of spatial information during active visual fixation. Current Biol 22: 510–514.

Li W, Piëch V, Gilbert CD (2004) Perceptual learning and top-down influences in primary

visual cortex. Nat Neurosci 7: 651–657.
Onat S, König P, Jancke D (2011). Natural scene evoked population dynamics across cat primary visual cortex captured with voltage-sensitive dye imaging. Cereb Cortex 21: 2542-2554.
O'Regan JK, Noë A (2001) The sensorimotor account of vision and visual consciousness. Behav Brain Sci 24: 939–1031.
Posner MI, Gilbert CD (1999) Attention and primary visual cortex. Proc Natl Acad Sci USA 96: 2585–2587.
Schoups A, Vogels R, Qian N, Orban G (2001) Practising orientation identification improves orientation coding in V1 neurons. Nature 412: 549–553.
Schwartz AS, Cheney C (1966) Neural mechanisms involved in the critical flicker frequency of the cat. Brain Res 1: 369-380.
Shoham D, Glaser DE, Arieli A, Kenet T, Wijnbergen C, Toledo Y, Hildesheim R, Grinvald A (1999) Imaging Cortical Dynamics at High Spatial and Temporal Resolution with Novel Blue Voltage-Sensitive Dyes. Neuron 24: 791–802.
Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24: 1193–1216.
Sinke C (2008) Effects of the gist on overt attention. Master thesis. University of Osnabrück.
Taravella CL, Clark G (1963) Discrimination of intermittent photic stimuation in normal and brain-damaged cats. Exp Neurol 7:282-293.
Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. Nat Neurosci 2: 79–87.
Ress D, Backus BT, Heeger DJ (2000) Activity in primary visual cortex predicts performance in a visual detection task. Nat Neurosci 3: 940–945.
Ringach DL, Hawken MJ, Shapley R (1997) Dynamics of orientation tuning in macaque primary visual cortex. Nature 387: 281–284.
Rucci M, Iovin R, Poletti M, Santini F (2007) Miniature eye movements enhance fine spatial detail. Nature 447: 852–855.
Rucci M (2008) Fixational eye movements, natural image statistics, and fine spatial vision. Network 19: 253–285.
Rust NC, Movshon JA (2005) In praise of artifice. Nat Neurosci 8: 1647–1650.
Watanabe T, Harner AM, Miyauchi S, Sasaki Y, Nielsen M, Palmo D, Mukai I (1998) Task-dependent influences of attention on the activation of human primary visual cortex. Proc Natl Acad Sci USA 95: 11489–11492.
Zennig C (2009) Cross-cultural comparison of visual attention and emotion recognition. Bachelor thesis. University of Osnabrück.

6. Appendix

6.1 Information on the method of voltage sensitive dye imaging

The method of voltage sensitive dye imaging used in the studies in section 2 and 3 enables real-time, in vivo optical imaging of the brain activity. It provides high spatial and temporal resolution at the same time. This section provides specific information voltage sensitive dye imaging (VSDI), and compares it with other methods used to investigate neuronal processing.

6.1.1 Mechanism

Voltage sensitive dye imaging was first used by Tasaki et al. (1968) in the squid giant axon to translate membrane potential into optical signals. The amplitude of the voltage sensitive dye signal is linearly correlated with changes in membrane potential (Cohen et al., 1974). The voltage sensitive dye molecules bind to neuronal membranes and their fluorescence properties correlate linearly to changes in membrane potential. They act as signal transducers exchanging electrical activity into optical signals within the studied tissue. This is possible because of the dye's electrochemical properties. Each dye molecule has a large dipole moment. It has a hydrophobic tail on one side and a hydrophilic part on the opposing side with a fixed charge. The dye binds to the membrane of cells with its hydrophobic tail. The fixed charge on the hydrophilic part of molecule prevents the molecule from crossing the membrane. The opposing charges within the dye molecule make it sensitive to changes of membrane potential and lead to changes in its fluorescent properties. See also Grinvald et al. (2001).

6.1.2 Temporal and spatial resolution

When the stained cortex is thus excited with light, local changes in membrane potential lead to changes in optical signals. These signals are captured with a fast CCD camera. We used a recording frame rate from 100 Hz to 200 Hz. Grinvald et al. (2001) determined an upper boundary for spatial resolution (around 50 x 50 micrometers per pixel), the limiting factors being optics and light scattering. In the results reported in this thesis, we use a resolution of 107 x 107 micrometers per pixel. The cortical area covered is a few square centimeters in size. According to Grinvald et al. (2001), a pixel viewing an area of 50 x 50 micrometers would measure a composition of signals from about 250-500 neurons. Consequently, in one of our pixels we would have contributions from 1145-2290 neurons. Since contributions from these neurons are mixed in each pixel, it is important for paradigms using this method, to make use of functional cortical organization. Here, topographic organization of cortical areas is a great advantage. In a topographic organization, neighboring cells are functionally more similar to each other than to cells that are further away.

6.1.3 Comparison to other methods

The dye signal is a sum over responses from dendrites, axons and cell bodies. However, the greatest share of the signal is generated by post-synaptic potentials, as the receiving dendritic trees cover a much larger cortical area than the cell somata (~1000 fold) (Grinvald et al., 2001). VSDI measurements in combination with other methods have confirmed that the dye signal reports changes in membrane potential (Petersen et al, 2003), in contrast to classical electrophysiological recording, which records spikes only. Being able to detect changes in sub-threshold activity can be very useful. Membrane potential much better reflects the synaptic input to neurons and the integrative processes within them than their spiking output.

Thus, visualizing subthreshold activity allows the investigation of input integration in a large cortical area. Other methods that can measure subthreshold activity are intrinsic electrophysiology, which allows measurement from only one neuron at a time, and external recording of local field potentials (LFPs), which measures responses at single cortical locations. Both methods are spatially restricted, while their temporal resolution is very good (in the sub-millisecond range). The method that is most similar to VSDI is LFP recording with multi-electrode arrays. Both have the advantage that responses can be recorded from different cortical locations simultaneously. VSDI imaging, however, reflects spatial *averages* at each pixel location, whereas multi-electrode recordings *sample* at individual locations. Signals measured with intrinsical optical imaging (without dye), like in fMRI, depend on the level of oxygen in the blood. The measured response is called hemodynamic response, or also blood-oxygen-level dependent (BOLD) signal. Recoding techniques that depend on the BOLD signal have a low temporal resolution. Intrinsical imaging is useful when focusing on spatial properties of cortical activation (e.g. MacEvoy et al. (2009) use frames of 500 ms), while VDSI allows investigation of both spatial and temporal properties, resolving temporal changes on smaller timescales (we use frames of 10 ms).

6.1.4 Conclusion

Using VSDI imaging, we can record simultaneously from large neuronal populations with high spatial and temporal resolution (Grinvald & Hildesheim, 2004). Since both of the studies in section 2 and 3 aim to investigate dynamical aspects of cortical processing in large populations of neurons tuned to different stimulus features, VSDI imaging was the appropriate choice.

6.1.5 References

Cohen LB, Salzberg BM, Davila HV, Ross WN, Landowne D, Waggoner AS, Wang CH (1974) Changes in axon fluorescence during activity: molecular probes of membrane potential. J Membrane Biol 19: 1–36.

Grinvald A, Shoham D, Shmuel A, Glaser D, Vanzetta I, Shtoyermann E, Slovin H, Sterkin A, Wijnbergen C, Hildesheim R, Arieli A (2001) In-vivo optical imaging of cortical architecture and dynamics. Technical Report GC-AG/99-6. The Weizmann Institute of Science. The Grodetsky Center for Research of Higher Brain Functions (revised version published in Modern Techniques in Neuroscience Research. Windhorst U, Johansson H (Editors) Springer Verlag).

Grinvald A, Hildesheim R (2004) VSDI: a new era in functional imaging of cortical dynamics. Nat Rev Neurosci 5: 874–885.

MacEvoy SP, Tucker TR, Fitzpatrick D (2009) A precise form of divisive suppression supports population coding in the primary visual cortex. Nat Neurosci 12: 637–645.

Tasaki I, Watanabe A, Sandlin R, Carnay L (1968) Changes in fluorescence, turbidity and birefringence associated with nerve excitation. Proc Natl Acad Sci USA 61: 883–888.

6.2 Rat retinotopy

6.2.1 Introduction

To prepare for experiments investigating cat primary visual cortex (presented in section 2 and 3), the author conducted experiments investigating the primary visual cortex of rats. This was done in the context of two projects, the first investigating retinotopic organization in rat primary visual cortex (Sascha Rekauzke, 2006, Diploma Thesis) and the second investigating the extension, propagation, and overlap of local activity in rat primary visual cortex under local stimulation (Jan Habijan, 2009, Diploma Thesis). Details about these projects are documented in the respective diploma theses. This section provides a brief overview of these experiments, of data analysis, and of the data.

Retinotopy is a specific kind of mapping from the visual field to the cortex, one that is topographic, meaning that neighbor-relationships are maintained. Thus, in a retinotopic mapping neighboring locations in the visual field are mapped onto neighboring locations on the cortex. It has been shown previously that the contralateral retina of the rat projects to the primary visual cortex (Rumberger et al., 2001) and ipsilateral input only reaches the lateral binocular area (Adams & Forrester, 1968). Most previous studies, investigating retinotopy in primary visual cortex, report that the superior-inferior axis of the visual field is represented on a posterior-anterior axis in the primary visual cortex (area 17), and the temporal-nasal axis is represented on a medial-lateral axis in the cortex (Adams & Forrester, 1968; Espinoza & Thomas, 1983; Gias et al., 2005), in another study however, a representation that was rotated about 45° was reported (Montero, 1973). Montero et al. found that both the dorsal and ventral temporal visual field was represented medially in the cortex, and the whole nasal visual field was represented laterally in the cortex. In the two projects mentioned above, voltage-sensitive dye imaging (VSDI) of responses to local stimuli was used to investigate retinotopic organization in rat primary visual cortex (area 17).

6.2.2 Methods

Animal preparation
The experiments described here were conducted to supplement the experiments documented in Rekauzke (2006), and Habijan (2009), as well as to provide the author with practical experience in the method of voltage sensitive dye imaging. Animal preparation and data acquisition was similar to that described in the context of the studies in section 2 and 3 (see methods sections). However, there were some differences as rats are different from cats. The following section contains a brief report of the methods with emphasis on the deviations from the methods described earlier in this thesis. For more details see Habijan (2009).

All animal experiments were approved by the German Animal Care and Use Committee and carried out in accordance with the Deutsches Tierschutzgesetz (§ 8 Abs. 1 Tierschutzgesetz). The experiments were conducted with rats (Dark Agouti, Janvier). Data was recorded from 10 male animals weighing between 218 g and 354 g. Initial anesthesia was done with chloral hydrate (400 mg per kg i.p.). Animals were artificially respirated, continuously anesthetized with 0.8% isoflurane in a mixture of O_2/N_2O (1:2 during surgery and 1:1 during measurement), and fed intravenously. Atropine sulphate was administered (0.05 ml/kg i.m.) every four hours. Before receptive field mapping, paralysis was induced with 0,09 mg/kg*h Alloferin® i.v. to prevent eye drift during measurement. Heart rate, exhaled CO_2 (3.8%), and body-temperature were constantly monitored and maintained. The skull was opened, using the bregma as a point of reference, the dura was removed, a chamber was mounted, the cortex was stained for 2-3 h with voltage-sensitive dye (RH1838), and unbound dye was washed out.

Data acquisition
The chamber was filled with Lactated Ringer's solution and sealed to maintain pressure. Before recording an image of the cortical vessel pattern was made using green light to enhance contrast. Then the camera was positioned. The focal plane was set to about 500 micrometers below the cortical surface. The cortex was illuminated with light of wavelength 630 ± 30 nm, and emitted light at wavelengths above 665 nm was collected. Stimulus presentation was locked to heart-beat for later artifact correction. Breathing was not locked to recording, because interfering with the rhythm of the pneumatic pump tended to have a negative influence physiological stability, which had a higher priority. (This is much more critical in a small animal like the rat than in cat).

Visual stimuli and presentation
The stimuli used where high-contrast phase inverting step-wave gratings. They alternated with a frequency of 2,5 or 4,5 Hz, meaning that there was a black-white-switch either 5 or 9 times per second, respectively (a full cycle is one switch and back). Luminance was 1-97 cd/m², with mean 49 cd/m². One local stimulus was 17° in size, with a spatial frequency of 0.06 c/°, so one full cycle (one black and one white bar) was visible (see Figure 6.1 for illustration of a local stimulus). In between conditions, there was a delay of at least 15 seconds. Recording was done for 1 second, including a 200 ms pre-stimulus interval. The stimulus was shown more than 800 ms, thus exceeding the recoded interval. Additionally, "blank" trials (grey screen) were recorded for later artifact correction. The grating was presented in horizontal and vertical orientation locally at 12 different positions on the monitor (100 Hz, Sony Triniton GDM-FW900, Japan), which was positioned 30 cm in front of the contra-lateral eye. The exact position of the monitor was chosen based on gaze direction, which was assessed by measuring the position of the blind spot, which was then back-projected onto the visual field (and marked on a glass panel) as well as by electrophysiology. Receptive fields of individual cells were mapped manually beforehand using a flashlight capable of projecting oriented light bars of varying length and width onto a surface. Potentials were amplified and played over audio speakers. Based on this feedback, receptive field borders were mapped onto the surface and manually noted on a glass plate. In some experiments, extra-cellular recordings were also made in response to local stimulation. Also full field stimuli (covering the whole monitor) were shown, previous to the local gratings, to check for visual responses with a strong stimulus first, before running the local paradigm.

Analysis
The recorded signal was normalized by the average (across 200 ms) pre-stimulus response-level and an average blank trial was subtracted to correct for heart-beat artifacts. The signal from the local presentation paradigm was then averaged across repetitions of each condition and across stimulus orientation, resulting in one signal for each of the twelve local stimulus positions. The activity not specific to position was removed by subtracting the average across all stimulus positions from each condition. To eliminate high-frequency artifacts and low-frequency gradients, the signal was spatially band-passed from 0.01 to 3 c/mm.

6.2.3 Results

Figure 6.1 and 6.2 illustrate retinotopic organization in rat primary visual cortex for one animal. Figure 6.1 shows average responses to local stimuli presented at twelve different positions. Each local response is an average across 80 repetitions (40 each orientation).

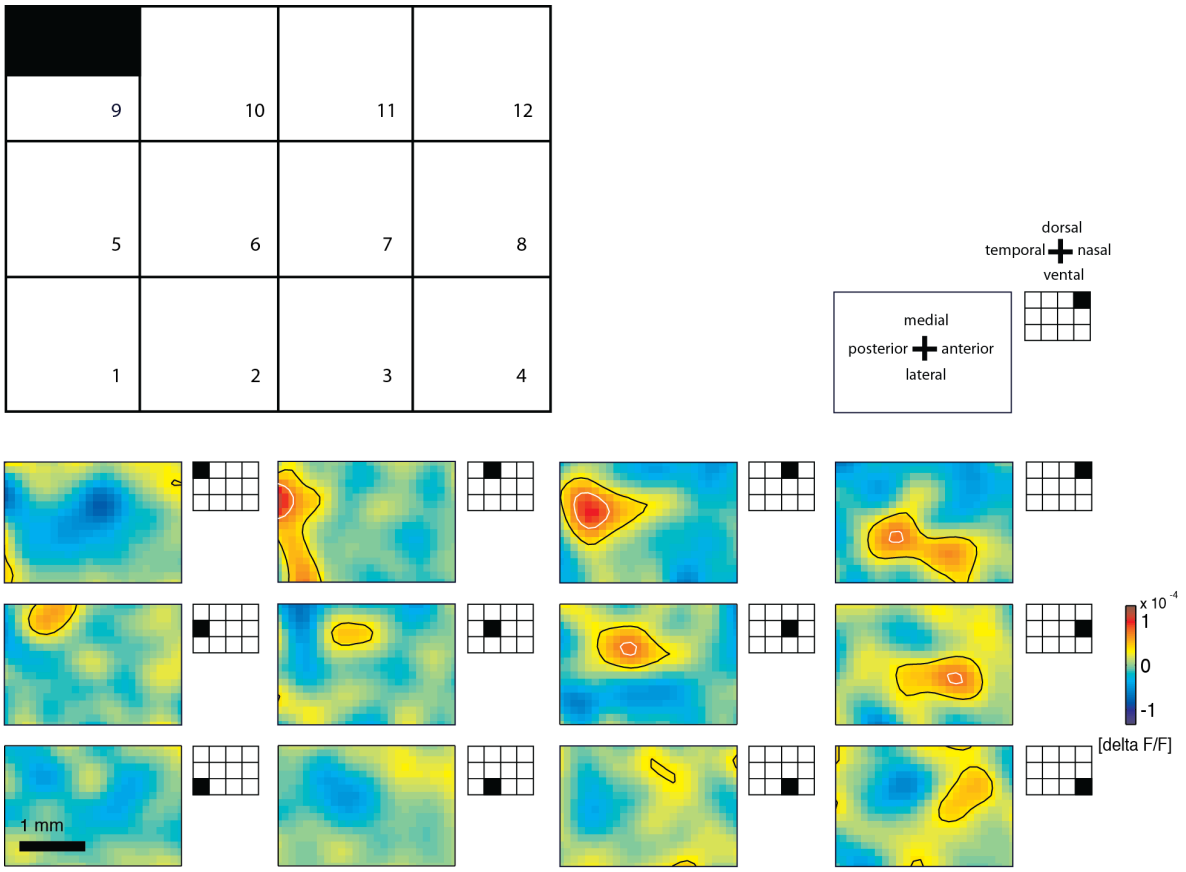


Figure 6.1: Top: Stimulus positions on the presentation screen. One of two phase-inverting gratings, horizontal orientation or vertical orientation, was shown at one of twelve possible positions on the screen. Bottom: Average responses for each of the twelve positions. Contour lines are plotted at 60% (black) and 80% (white) of maximal average activity.

Representation of stimulus location

Temporal averages were computed from 100 ms after stimulus onset to the end of stimulus presentation at 700 ms later to avoid artifacts from the initial processing delay. Contour lines were computed at 60% and 80% of the maximal activity across space and conditions. Responses to stimulation at the twelve monitor positions with contour lines overlaid are shown in Figure 6.1. In Figure 6.2 only the 60% contours from responses to local stimuli are shown and overlaid on top of the cortical vessel pattern to aid comparison of cortical position. There is a systematic relationship between the position of the local stimulus and the position of its response on the cortex: Dorsally presented stimuli (9-12) lead to activation in posterior regions of V1. Stimuli presented more ventrally (5-8, and also 1-4) lead to activation of more and more anterior regions. Along the other axis, temporal stimuli were represented more medially, while nasally presented stimuli were represented more laterally. There was no visible activation for stimulus 1,2, and 9 inside the region of interest (ROI). Considering the layout of the mapping of the other local stimuli, these stimuli are likely to be projected to regions outside the ROI, which could not be detected in this experiment.

Size of local activity

Visual responses were obtained in a region from 4 to 6.5 mm lateral and from 2 to 4 mm posterior to bregma (Figure 6.2). All local stimuli were of the same size in the visual field, but the contoured activated regions are of different sizes for different stimuli. Larger regions are activated laterally compared to medially (from nasal vs. temporal stimulation), and slightly larger regions are activated posteriorly compared to anteriorly (from dorsal vs. ventral

stimulation). This excludes the cases of stimulus number 1,2, and 9, in which no local activity was measured.

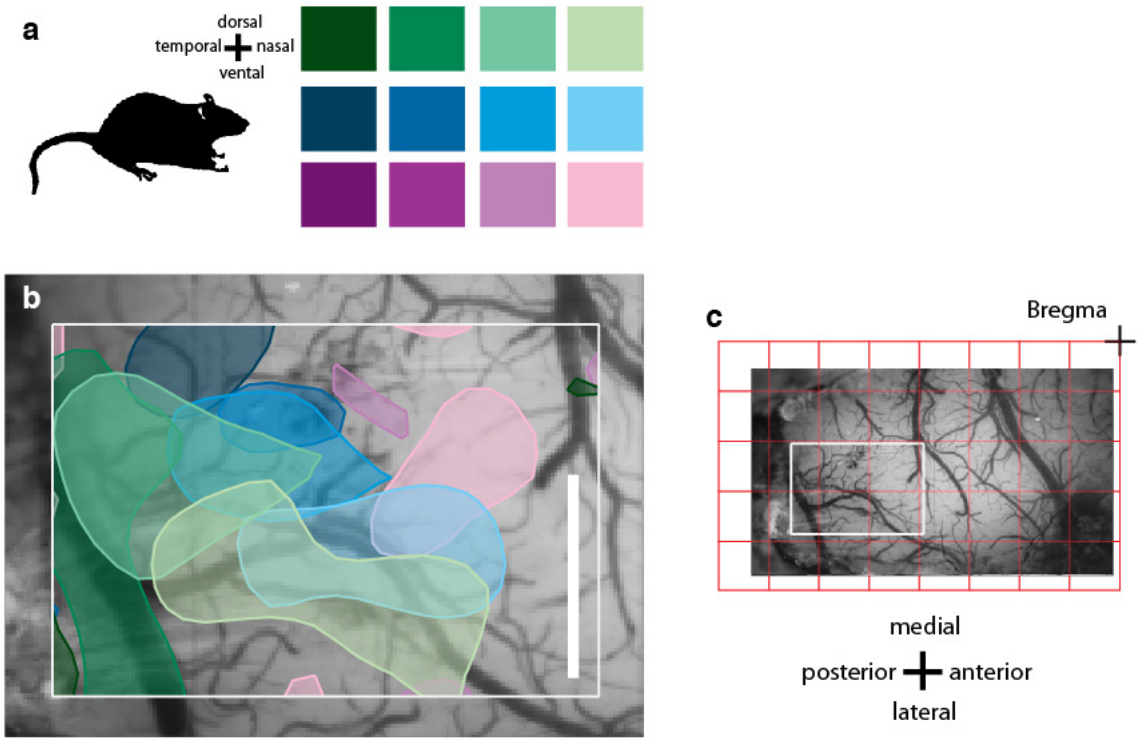


Figure 6.2: **a)** Positions on the screen are color-coded. Stimuli were presented on the left side of the animal. **b)** Contour lines of the 60% activation level for each possible position on the screen (also seen in Figure 6.1, black) are overlaid on top of each other and on top of the cortical vessel pattern inside the ROI shown in c. Contour lines and regions within contour lines were colored according to the color code shown in a). The white bar indicates one millimeter. **c)** Cortical vessel pattern underneath opening and region of interest (ROI) as white frame. The red grid is spaced in one-millimeter distances and is anchored at the position of the bregma.

6.2.4 Discussion

Representation of stimulus location

Using VSDI and localized stimuli it was possible to document the retinotopic organization in rat primary visual cortex (area 17). There was a systematic relationship between the position of the local stimuli on the monitor (or in the visual field since eyes did not move) and the cortical position of their evoked activity (Fig 6.1 and 6.2). The retinotopic representation shown here is incongruent with Montero et al.'s (1973) result that the whole (dorsal and ventral) nasal visual field is represented in the lateral visual cortex (area 17). Instead, the layout of the retinotopy is consistent with studies that found that the dorsal-nasal visual field (Figure 6.2, light green) is represented laterally, and the ventral-nasal field (Figure 6.2, light purple) is represented medially (Adams & Forrester, 1968; Espinoza & Thomas, 1983, Gias et al., 2005).

Size of local activity

Gias et al. (2005) found a relatively larger representation of the dorsal-nasal area compared to other regions of the visual field. They argue that this might partly be due to this being the binocular region, which might use a larger neural circuitry. The data presented here is consistent with this observation of cortical magnification (Figure 6.2, light and medium-light green, and light blue lead to a larger area of activity).

V1 location

The data show visual responses in a region from 4 to 6.5 mm lateral and from 2 to 4 mm posterior to bregma (Figure 6.2). It is important to realize that not the whole primary visual cortex was stimulated, since the monitor did not cover the entire visual field. Thus, areas in which visual responses were not recorded could still be part of primary visual cortex. In addition, a region could be stimulated but still not lead to a recorded response, when signal quality would be to low, as would be the case for responses outside the ROI (expected response locations for stimulus position 1,2 and 9). This means that also areas outside the ROI could be part of the primary visual cortex. So the values above are rather minimal boundaries than maximal boundaries. Interpreted this way, the observations here are consistent with results reported in the two above mentioned diploma theses: Habijan localized the primary visual cortex 4,9 to 8,2 mm posteriorly and 2,4 to 4,3 mm laterally to bregma (Habijan, 2009, p.26) and Rekauzke found it 4 to 7.5 mm posteriorly and 1.7 to 4.5 mm laterally to bregma (Rekauzke, 2006, p.23-24).

6.2.5 References

Adams AD, Forrester JM (1968) The projection of the rat's visual field on the cerebral cortex. Exp Physiol 53: 327-336.

Espinoza SG, Thomas HC (1983) Retinotopic organization of striate and extrastriate visual cortex in the hooded rat. Brain Res 272: 137-144.

Gias C, Hewson-Stoate N, Jones M, Johnston D, Mayhew JE, Coffey PJ (2005) Retinotopy within rat primary visual cortex using optical imaging. Neuroimage 24: 200- 206.

Habijan JE (2009) Repräsentation von Luminanzänderungen visueller Reize im Kortex der Ratte. Diplomarbeit. Ruhr-University Bochum.

Montero VM (1973) Evoked responses in the rat's visual cortex to contralateral, ipsilateral and restricted photic stimulation. Brain Res 53: 192-196.

Rekauzke S (2006) Populationsrepräsentationen des Gesichtsfeldes im visuellen Kortex der Ratte gemessen mit spannungsabhängigen fluoreszenten Farbstoffen. Dimplomarbeit. Ruhr-University Bochum.

Rumberger A, Tyler CJ, Lund JS (2001) Intra- and inter-areal connections between the primary visual cortex V1 and the area immediately surrounding V1 in the rat. Neurosci 102: 35-52.

7. Abbreviations

BOLD	blood-oxygen-level dependent
i.m.	intramuscular (into a muscle)
i.p.	intraperitoneal (into the peritoneum "body cavity")
i.v.	intravenous (into a vein)
LFP	local field potential
LGN	lateral geniculate nucleus
RF	receptive field
ROI	region of interest
SVD	singular-value decomposition
V1	primary visual cortex
VSDI	voltage-sensitive dye imaging

8. Acknowledgments

I would like to express deep appreciation for my two supervisors Peter König and Dirk Jancke. Their mentorship helped me not only to solve tasks at hand, but also to learn for life. I thank Sascha Rekauzke for teaching me many practical experimental skills, for his invaluable help during the experiments, for always thinking one step ahead, and for generally being the most reliable person I have come to know. I learned a lot from working with Selim Onat. I would like to thank him, in particular, for so readily sharing his valuable experiences and knowledge and for introducing me to optical imaging data analysis. It was a great experience to work together with Sepp Kollmorgen and Sylvia Schröder in the eye-tracking project. I am grateful for that experience. I thank the members of the Optical Imaging Group in Bochum Anna Palagina, Benedict Ng, Agnieszka Grabska-Barwinska, and Jan Habijan, and in Osnabrück Alper Açık and Saskia Nagel and the members of the Neurobiopsychology Group for professional feedback at various stages of my work, and for providing an excellent research environment. I thank my parents for their everlasting encouragement to always seek challenge, and Robert Märtin for proofreading and professional feedback on this thesis.

9. Declaration

All experiments reported in this thesis conformed to National and Institutional Guidelines. Experiments involving human subjects conformed to the Declaration of Helsinki.

Parts of this thesis were published in scientific journals. To achieve a consistent style throughout this thesis, the text might slightly differ from published versions. Contents in chapter 2 (study 1) were published before in the journal NeuroImage and the copyright was transferred to Elsevier. The contents are reproduced here with permission from Elsevier. Contents in chapter 3 (study 2) were published previously in the journal Cerebral Cortex under the Creative Commons Non-Commercial (CC-BY-NC) license according to Oxford Journals policy. Contents in chapter 4 (study 3) were published before in the journal PLoS Computational Biology under the Creative Commons Attribution (CC-BY) license according to PLOS policy. The references to the previously published versions are given at the beginning of each chapter and in the section below.

10. Contributions of the author

The work presented in this thesis was performed in collaboration with several scientists. Therefore, the plural form is used throughout this thesis. The author hereby ensures that she significantly contributed to all materials used in this thesis. Below, the author's individual contributions in the presented projects are specified.

Study 1: Independent encoding of grating motion across stationary feature maps in primary visual cortex visualized with voltage-sensitive dye imaging.

The author contributed to the experimental design conceptually and implemented parts of it. She did preliminary data analysis and contributed to the development of the analysis and related decisions, discussed the content of the paper and commented on the manuscript. The author measured data together with Sascha Rekauzke and Dirk Jancke. This data, significantly affected the whole study including analysis and interpretation of previously measured data.

The work was published in the journal NeuroImage.

Journal Publication

> Onat S, Nortmann N, Rekauzke S, König P, Jancke D (2011) Independent encoding of grating motion across stationary feature maps in primary visual cortex visualized with voltage-sensitive dye imaging. Neuroimage 55: 1763-1770.

Study 2: Primary visual cortex represents the difference between past and present.

The author was the main contributor to all aspects of the presented work. She implemented the experimental paradigm and did all data analysis. She conceived the experimental paradigm in close collaboration with her co-authors, and wrote the manuscript in close collaboration with Dirk Jancke. She did the data acquisition with assistance from Sascha Rekauzke.

The work was published in the journal Cerebral Cortex. In addition, it has led to three conference contributions as posters, which were prepared and presented by the author.

Related contribution: She assisted Sascha Rekauzke in data acquisition for his PhD project, which is not presented here. That project will be documented, respectively, in the PhD thesis by Sascha Rekauzke. A manuscript with her co-author contribution is in preparation.

Journal Publication

Nortmann N, Rekauzke S, Onat S, König P*, Jancke D* (2013) Primary visual cortex represents the difference between past and present. Cereb Cortex doi: 10.1093/cercor/bht318.
**These authors shared senior authorship.*

Conference Contributions

Nortmann N, Rekauzke S, Onat S, König P, Jancke (2012) Change detection in primary visual cortex. Opening Symposium of the BCAN and ICMB, Neural Computation: From Perception to Cognitive Function, Berlin, 29-31 Oct 2012 (Poster).

Nortmann N, Rekauzke S, Onat S, König P, Jancke D (2012) Primary visual cortex represents the difference in orientation content within rapid stimulus sequences. OCCAM 2012, Osnabrück, 4-6 June 2012 (Poster).

Nortmann N, Rekauzke S, Onat S, König P, Jancke D (2011) Representation of difference: Primary visual cortex encodes the relative change in orientation content within rapid stimulus sequences. SFN Annual Meeting 2011, Washington, DC, USA, 12-16 Nov 2011. Poster 799.11 (Poster).

Related contribution (not presented in this thesis)

Rekauzke S, Nortmann N, Jancke D (in preparation) Propagating waves in response to opposing luminance changes of two adjacent stimuli.

Study 3: Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention.

The author performed the experiments and data analysis in collaboration with Sepp Kollmorgen and Sylvia Schröder with an own focus on the analysis of the eye tracking data, and deriving the empirical and spatial bias salience measures. All analysis was jointly developed and discussed. The experiment was conceived and designed, and the manuscript written in close collaboration by all authors of the manuscript.

Parts of the eye tracking data were acquired and preliminarily analyzed in a pilot study and in the context of the author's Master thesis. The work presented here is based on further extensive data measurement, method development, and analysis done later, during her PhD studies.

The work was published in the journal PLoS Computational Biology. The author presented the work in an early stage in an invited talk at the Center for Nonlinear Studies, Los Alamos National Laboratory, NM, USA. In addition, two conference contributions were made: A poster that was presented by the author together with Sepp Kollmorgen and Sylvia Schröder, and a talk that was jointly prepared and given by Sepp Kollmorgen.

Journal Publication

> Kollmorgen S*, Nortmann N*, Schröder S*, König P (2010) Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. PLoS Comput Biol 6: 1-20.
>
> **These authors contributed equally to this work.*

Conference Contributions

> Kollmorgen S, Nortmann N, Schröder S, König P (2007) Integration of task dependent information of different image patches during scanning eye movements. Journal of Eye Movement Research, Special Issue: Abstracts of the ECEM2007, A2-06 (Talk).
>
> Schröder S, Nortmann N, Kollmorgen S, König P (2007) Influence of task dependent information and saliency on overt attention. Journal of Eye Movement Research, Special Issue: Abstracts of the ECEM2007, PA-08 (Poster).

Invited Talk

> Nortmann N (2007) Viewing Behavior: Influence and integration of task dependent information. Center for Nonlinear Studies, Los Alamos National Laboratory, NM, USA.

Appendix 6.2: Rat retinotopy

The author performed the presented experiments, developed the scripts for data analysis, analyzed the data, and produced the illustrations and the documentation.