Vision in context

from behavior to neurons

Wietske Zuiderbaan

Context in vision – from behavior to neurons

ISBN 978-94-6332-201-0

Printed by GVO Drukkers & Vormgeving B.V.

Foar myn mem

Vision in context

From behavior to neurons

Visuele perceptie in context

Van gedrag naar neuronen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 7 juli 2017 des middags te 2.30 uur

door

Wietske Zuiderbaan

geboren op 12 juni 1982

te Wijckel

Promotoren: Prof. dr. S.O. Dumoulin
Prof. dr. F.A.J. Verstraten

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van de Nederlandse Wetenschaps Organisatie subsidie 452-08-008 aan S.O. Dumoulin

Contents

Chapter 1

General introduction

Vision is one of the important senses that make us able to observe and interact with our environment. When we look around in the world, we see many different objects and shapes. In our interaction with the world around us, we need to be able to search, locate and recognize objects. Therefore, the visual system needs to segment and detect objects from the visual scene.

Visual processing (in primates) starts with the retina being stimulated by light. The visual system has to interpret this incoming light so we are able to make sense of our environment. Since the information coming from the retina is noisy and ambiguous, the visual system needs to make an interpretation of it. It does this based on our knowledge of the world. Using this prior knowledge, the visual system derives the most likely cause of the retinal image.

How does the visual system make this interpretation of the outside world? It does this in a hierarchical way. The retinal image that contains luminance information is being processed into ever more abstract representations. The brain processes the retinal image via different visual areas. Early visual areas process information from a small region of the visual field and represent simple sensory-driven image properties such as contrast and orientation (Hubel & Wiesel, 1962). Later visual areas extract their meaningful information from the representation of the early visual areas. The representation in later visual areas becomes more complex. Later visual areas process the information from a larger region of the visual field and represent more meaningful information about for instance the presence and shape of objects.

The interaction between the sensory-driven and knowledge-based information in visual perception is illustrated in the visual illusion in Fig. 1 (Adelson, 1995). The two squares indicated with ‘A’ and ‘B’ have the same local sensory-driven input based on the brightness of the square. When the two squares are presented in isolation, the squares are perceived as being equally gray. However, when perceiving the squares in a context in which square A is illuminated directly and square B lies in the shadow, our subjective experience is that the squares have a different brightness (Adelson, 2000).

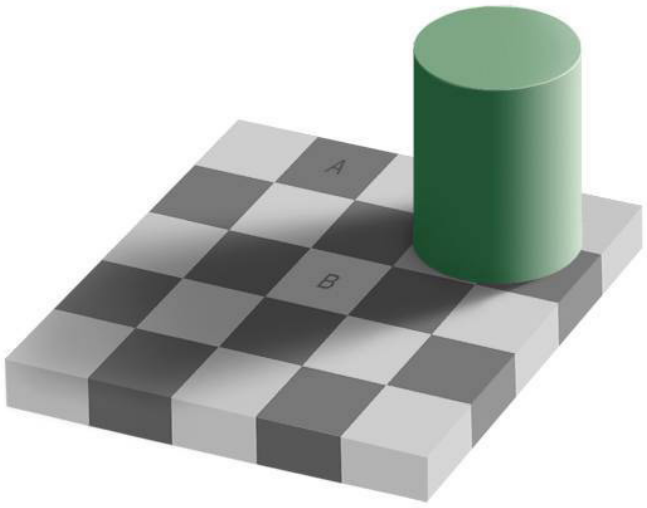


Figure 1 The checker shadow illusion from Adelson (figure adapted from Adelson, 1995). This visual illusion illustrates the interaction between the sensory-driven and knowledge-based information in visual perception. The local sensory-driven information of Square A and B is similar, but the subjective experience based on the context of the global percept is that they have a different brightness (Adelson, 2000).

How the retinal image is processed in our brain and how this influences our behavior is a big question in neuroscience. **In this thesis, we investigated how visual perception is affected by both the sensory-driven and knowledge-based information.** We will investigate the effect that these sources of information have on our behavior and on the neural signal. For this, we quantified both sensory-driven and knowledge-based information of natural images. To study the effects on our perceptual functioning we use a psychophysical method and to study how the information is represented in the neural signals in visual cortex we use computational neuroimaging techniques.

Synthetic versus natural images

In the experiments described in this thesis, we used different stimuli. We used both synthetic and natural images. The synthetic images dominated in their content of contrast-energy and mainly contained sensory-driven information, whereas the natural images contain both sensory-driven and knowledge-based information.

Synthetic images

The synthetic images were randomly generated patterns consisting of a hexagonal grid that was partially filled with a high-contrast pattern. Fig. 2 shows an example of the synthetic images used in the experiment. The synthetic images contained mainly sensory-driven information.

Figure 2 Examples of the synthetic images that we used in this thesis. The synthetic images were used to study the effects of sensory-driven information on visual perception.

Natural images

The natural images used in our experiments (Fig. 3A) were taken from the 'Berkeley Segmentation Dataset and Benchmark' database (D. Martin, Fowlkes, Tal, & Malik, 2001). Where the dominant information of the synthetic images was the sensory-driven image feature contrast-energy, we quantified this in the natural images by calculating the amount of Root-Mean-Squared (RMS) contrast for the natural images (Fig. 3B). The RMS contrast is the ratio between the local image intensity and the average image intensity.

The natural images contain besides sensory-driven information also knowledge-based information. For every image of the 'Berkeley Segmentation Dataset and Benchmark' dataset, there are manually labeled images available. The manually labeled images were made by human observers and contain a description of the subjective importance of the images (Fig. 3C). The task of the observers was to identify the most important aspects of the scene. We used the average manually labeled images of five observers as our definition of the knowledge-based image information.

The database of the manual labels was developed as a ground truth for object and boundary detection algorithms in computer vision. When parts of an image are

shown in isolation, human observers are poor at identifying the boundaries and junctions (McDermott, 2004). The sensory-driven image features that are used to identify the boundaries and junctions in natural scenes are highly ambiguous. Therefore, boundaries and junctions need to be inferred using the global context of the scene and our knowledge of the world.

For the dataset of images that we used in our experiments, Martin et al. also compared the manual labels with the performance of local boundary detector algorithms. Where the algorithms detected parts of the manual labels, a significant portion was not detected (D. R. Martin, Fowlkes, & Malik, 2004). This suggests that the manual labels were not only based on local sensory-driven image properties, but also on global image context and our knowledge of the world.

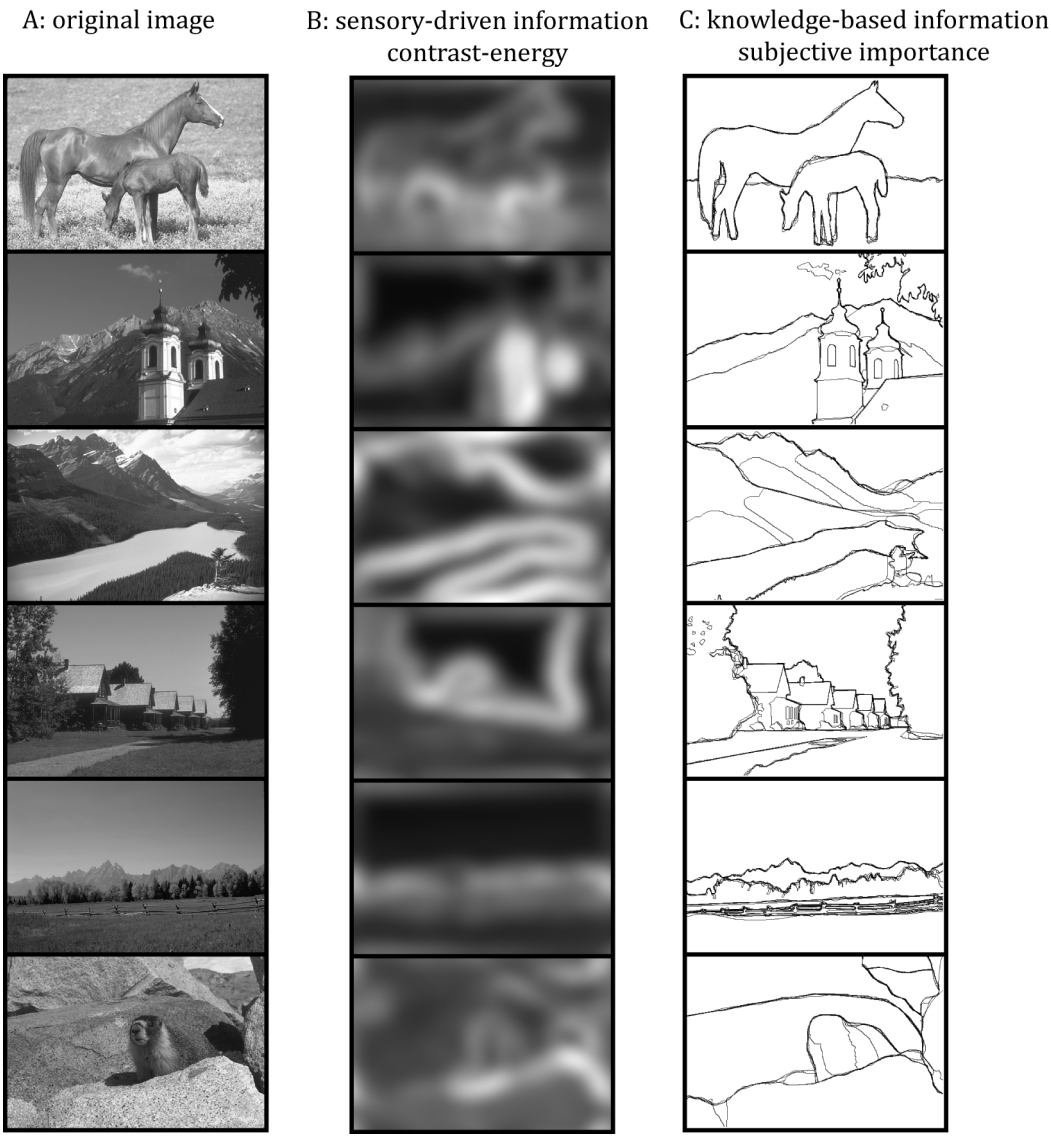


Figure 3 A: In this thesis we used natural images from the 'Berkeley Segmentation Dataset and Benchmark' database (D. Martin et al., 2001). We investigated how visual perception is affected by both the sensory-driven and knowledge-based information. B: To quantify the sensory-driven information we calculated the RMS-contrast of the images. C: For all the images of this database human observers manually labeled the important aspects of the image that represent the subjective importance of the image. We used the subjective importance to quantify the knowledge-based information of the images.

Behavior

Change blindness

When we look around in the world, we have the experience that we perceive the world around us in high detail, like a camera. However, even when we are confronted with large changes in a visual scene, we often fail to notice them. And once we notice the change, it is hard to imagine that we didn't see it. This phenomenon is called change blindness and illustrates that not every part of the visual scene is processed in such great detail as we intuitively experience.

The visual system is bombarded with a large amount of information from the light that falls on our retina, and our brain needs to make a coherent percept from this. However, the brain only has a limited amount of processing capacity, therefore, not every part of the visual scene can be processed in great detail.

In a change blindness experiment, subjects need to detect a change in two alternating images. Some changes are easier to detect than others. Current theories propose that our ability to detect changes in a visual scene is influenced by the knowledge-based image representation of the scene. Therefore, changes that occur in this knowledge-based image representation are detected faster (O'Regan, Rensink, & Clark, 1999; Rensink, ORegan, & Clark, 1997; Sampanes, Tseng, & Bridgeman, 2008; Shore & Klein, 2000; Stirk & Underwood, 2007). On the other hand, the representation in early visual cortex is dominated by sensory-driven information such as contrast-energy of the image (Boynton, Demb, Glover, & Heeger, 1999; De Valois & De Valois, 1988; Dumoulin, Dakin, & Hess, 2008; Mante & Carandini, 2005; Olman, Ugurbil, Schrater, & Kersten, 2004).

Our visual percept is made from the interactions of both the sensory-driven information and the knowledge-based information of the scene. In **chapter 2** we investigated whether these interactions are reflected in our behavior, more specifically on our ability to detect changes in a visual scene.

Neurons

Neuronal response properties

Since the visual system receives an enormous amount of information from the retina, it needs efficient ways to interpret and represent the relevant information from the retinal image. The brain interprets this information by the computations on the retinal image. A way to understand the computations of the brain is by the ability to predict how neurons will respond to stimulation.

Contrast response function

The responses of neurons in early visual cortex are known to be strongly modulated by contrast (Boynton et al., 1999; De Valois & De Valois, 1988; Dumoulin et al., 2008; Mante & Carandini, 2005; Olman et al., 2004). These responses do not increase linearly with increasing contrast, but the responses saturate at high contrast levels. How a neuron will respond to different levels of contrast is also referred to as the contrast response function (CRF) (Albrecht & Hamilton, 1982). The non-linear CRF makes neurons only sensitive to differences in a limited range of contrasts. This range is dynamic and can be modulated by different mechanisms (Carandini & Heeger, 2011). The non-linear CRF can be for instance shifted by the average contrast in the environment. This shift of the CRF explains adaptation effects; when the average contrast level changes, the visual system is able to optimally detect differences around the new average level of contrast.

Receptive field properties

Neurons in visual cortex respond to a certain region in the visual field, which is called the receptive field of this neuron (also referred to as the classical receptive field). When stimulating this part of the visual field, the neuron is activated. Neurons in early visual cortex have small receptive fields and will therefore only contain information from a small region of the visual scene. Neurons with small receptive fields will lack an amount of information, which makes their representation to be ambiguous. A way to solve this

ambiguity is by putting the information in context. A meaningful percept can be made when the information of smaller receptive fields are integrated over a larger space in the visual field.

Throughout the visual system, neurons are found that have so called center-surround receptive fields (for review see: (Allman, Miezin, & McGuinness, 1985; Carandini, 2004; Cavanaugh, Bair, & Movshon, 2002; Fitzpatrick, 2000; Hubel & Wiesel, 1968)). The response of these neurons is adjusted based on information that is outside their classical receptive field. When the information in the center is different from the information in the surround, these neurons will give the biggest response. The functional property that these neurons have is to detect changes in the visual scene; they make a comparison between the information of the center and the surround. These neurons can detect for instance changes in luminance, orientation or direction (Hubel & Wiesel, 1962, 1968). Center-surround modulations are supposed to play a role in mechanisms such as edge detection and are also widely used for this purpose in computer vision algorithms.

Receptive field properties are expected to play a functional role in visual computations. These properties become progressively more sophisticated in later visual areas. Receptive fields in later visual areas show responses to more complex stimulus properties. Studying the receptive field properties is needed to gain a better understanding of their computational function in the visual system.

Receptive fields of neurons that are nearby in visual cortex respond to regions nearby in the visual field. As a result, the visual cortex is organized into retinotopic maps. They are called retinotopic since they show a similar layout as in the retina and are also referred to as visual field maps (Wandell, Brewer, & Dougherty, 2005). These visual field maps can be measured using fMRI.

fMRI

Functional Magnetic Resonance Imaging (fMRI) is a commonly used technique to measure brain activity. The brain needs oxygen to function, and the fMRI signal is based on the differences in the local concentration of blood oxygen. The activation measured with fMRI is an indirect measurement of the activation of a population of neurons (Moseley & Glover, 1995). FMRI measures the activation of volumes of the cortex that are typically of a size of 2x2x2mm. These volumes are called voxels (volume-element) and contain about a million of neurons. Since this is a non-invasive technique it can be used to measure neural activity in human subjects. This technique is widely used to localize brain areas that are involved for specific tasks. However, fMRI can also be used to study more complex neuronal properties.

Computational neuroimaging

Conventional fMRI studies focus on whether a certain task or stimulus leads to an increase of neural activity, and in which brain areas this activation is found. The drawback of this approach is that it ignores the differences in neuronal properties. For example, different cortical locations respond to different parts of the visual field. An approach that takes these differences into account is computational neuroimaging. In computational neuroimaging, fMRI is not only used for investigating which parts of the brain respond to a certain stimulus, it also investigates how this part of the brain responds to the stimulus. In this approach, fMRI is used as a tool to test hypotheses about the neural computations (Wandell, 1999). There are various methods of computational neuroimaging: visual field mapping, biologically inspired models such as population receptive field models and classifiers that decode information from the fMRI signal.

Visual field mapping

The methodology of fMRI can be used to distinguish the different visual field maps in cortex. For this, the traveling-wave method (Engel, Glover, & Wandell, 1997) (also known as the phase-encoded paradigm (Sereno et al., 1995)) was designed. This technique infers

the preferred spatial location in the visual field for every voxel. For this, different parts of the visual field are stimulated while measuring the fMRI response simultaneously. While the subjects fixate at the center of the visual field, expanding and contracting rings are presented. The expanding and contracting rings stimulate parts of the cortex that respond to different peripheral locations of the visual field. These stimuli are used to define the borders of the visual field maps based on eccentricity, i.e. the distance to fixation. Rotating wedges stimulate parts of the cortex that respond to locations in the visual field that are different in preferred polar-angle. These stimuli are used to distinguish the borders between the visual field maps. By reconstructing the positions of the visual field on the cortical surface, the visual field maps can be drawn based on the borders of polar-angle and eccentricity (Fig. 4).

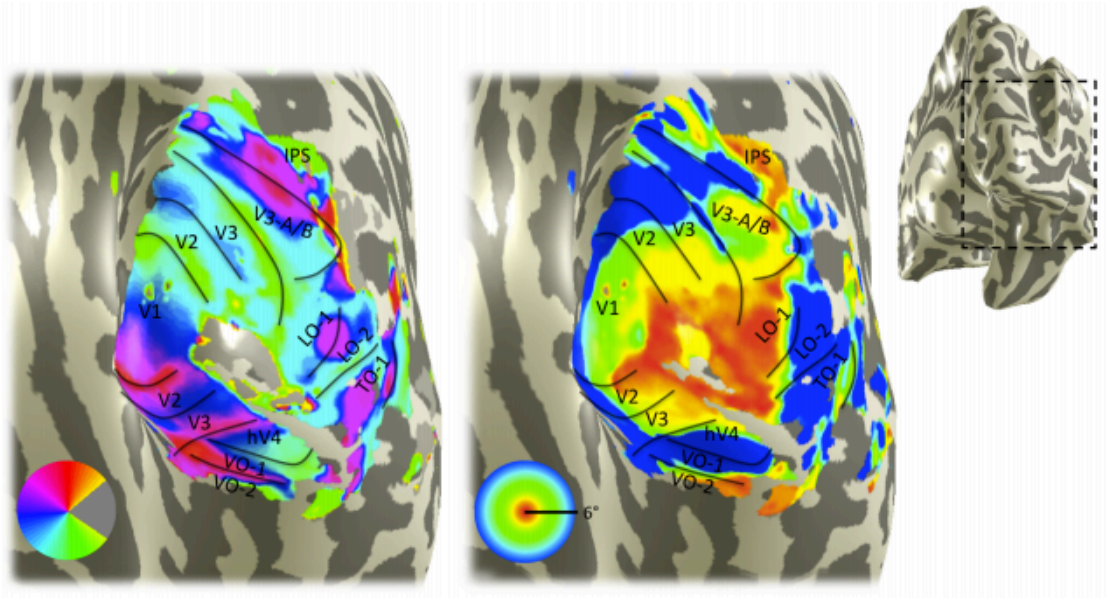


Figure 4 Visual field maps in occipital cortex can be identified from polar angle (left) and eccentricity (right). Figure from (Wandell, Winawer, & Kay, 2015).

Population receptive field mapping

The traveling wave method estimates the preferred location of a voxel as a point in visual space. However, neurons in visual cortex do not only respond to a point, but to a region in the visual space, called the receptive field. fMRI measures from a neuronal population

instead of a single neuron, therefore, the aggregate receptive field measured with fMRI is referred to as the population receptive field (pRF) (Dumoulin & Wandell, 2008; Victor, Purpura, Katz, & Mao, 1994). The properties of pRFs in visual cortex can be measured using the pRF-method (Dumoulin & Wandell, 2008). The stimuli used for the estimation of the pRF-model are bar apertures that move trough the visual field in 8 different directions. The original pRF-model describes the pRF with a single circular Gaussian in the visual field. In short, the parameters of the pRF are estimated by combining the measured fMRI time series with the presented stimulus over time. A prediction of the time series is made based on the overlap of the pRF and the stimulus. The model estimates the parameters of the pRF that give the best prediction to the measured time series (Fig. 5). This method estimates for every cortical location not only the parameters for its preferred location in the visual field, but also for the size of this region.

The conventional pRF-model uses only a single circular Gaussian to describe the pRF, and it therefore doesn't capture any influences of possible surround effects of the pRF. In **Chapter 3** we extend the conventional pRF-model by adding a suppressive surround to the pRF.

The strength of the pRF-model is that it allows the selection of voxels that respond to certain parts of the image. Therefore, it is possible to select responses elicited by different contrast-energy levels in natural images. In **Chapter 5** we use the pRF-model to derive contrast response functions from the inherent variation in contrast of natural images.

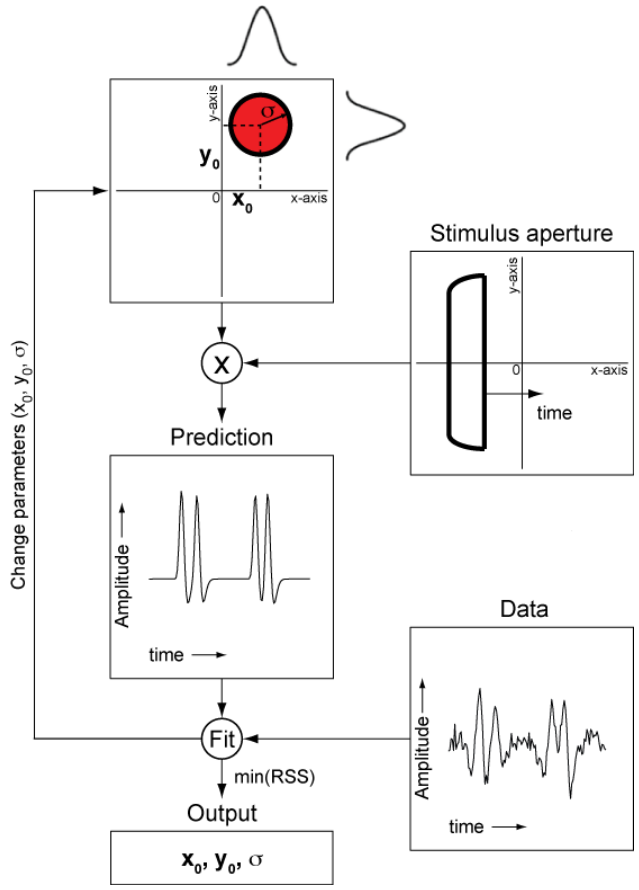


Figure 5 Model fit routine of the original (One Gaussian) pRF-model. For every cortical location a pRF is estimated. The pRF is modeled as a single circular Gaussian in the visual field with parameters for its location (x_0,y_0) and for the size of the pRF (σ). The pRF-model makes a prediction of the measured fMRI signal based on the overlap of the pRF and the stimulus over time. The parameters of the pRF are those that give the best prediction to the measured fMRI signal.

Brain reading

Another challenge of computational neuroimaging is to predict the mental state from the measured brain activity. Different studies showed that it is possible to determine from the measured fMRI signal what the subject was seeing, hearing, remembering or even dreaming (for review see: (Tong & Pratte, 2012). Most of these studies make use of the statistical differences between activation patterns that are elicited by different stimuli. For example, in the study of Haxby et al. they trained a classifier to determine whether the

subject saw an image with a house or a face (Haxby et al., 2001). The drawback of these approaches is that it is not possible to make a prediction when the stimulus comes from a different category, for instance an image of a car.

A different approach is to use biologically inspired models that are able to make predictions about the presented stimulus. The advantage of this approach is that it is possible to make a prediction about any stimulus containing properties that are defined by the model. The pRF-model is such a biological model that estimates for every cortical location the region in the visual field that it responds to.

Using the pRF-model we can make a prediction of the brain activity for any visual stimulus. In **Chapter 4** we describe an fMRI study in which we used the pRF-model to identify the visually presented image. We identified both synthetic and natural images based on the sensory-driven contrast-energy information.

Theories of neural mechanisms of inference

Where visual processing starts with the extraction of simple sensory-driven image properties, the computations become more complex in later stages of visual processing. Our visual percept is made from the interactions of both the sensory-driven information and the knowledge-based information of the scene. It is unclear how this interaction is reflected in the neural signal.

Different computational and statistical theories propose a different effect on the neural signal for the interaction between the sensory-driven and knowledge-based information. In these theories, later visual areas form a perceptual hypothesis about the external source, which they feedback to early visual areas. The feedback modulates the sensory-driven neural representation in early visual areas according to this perceptual hypothesis. The difference between the theories lies in the effect of the feedback; it can be either boosting or suppressing the neural representation of the early visual areas.

Fig. 6 illustrates two different theories that predict opposing effects of the feedback on the internal representation of early visual areas. On the one hand, predictive coding states that the feedback from later areas subtracts the perceptual hypothesis from the

neural representation of early visual areas (Friston, 2002; Mumford, 1992; Rao & Ballard, 1999). By subtracting the perceptual hypothesis from the internal representation of early visual areas, the unpredicted information remains represented in the signals of early visual areas. On the other hand, the efficient coding theory proposes that the internal representation of early visual areas is enhanced according to the perceptual hypothesis (Barlow, 2001; Friston, 2002; Rao, 2005; Series, Lorenceau, & Fregnac, 2003; Simoncelli & Olshausen, 2001). This causes the internal representation to sharpen the sensory-driven representation according to the perceptual hypothesis.

There is no consensus on how the interaction of the sensory-driven information and the knowledge-based perceptual hypothesis takes place in early visual areas. Where the interaction effect has only been studied using synthetic images, in **chapter 5** we investigated the effect that these interactions have on the internal representation of early visual areas using natural images.

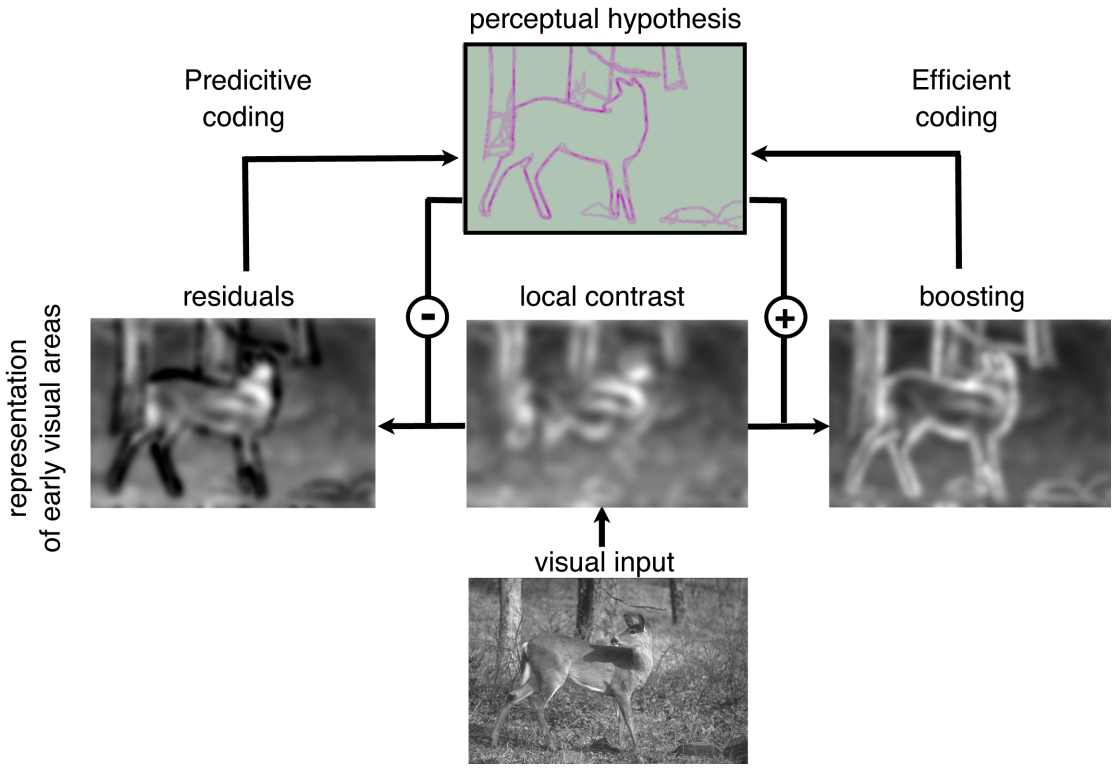


Figure 6 Two theories of neural mechanisms of inference. Opposing effects are proposed for the effect of the knowledge-based perceptual hypothesis on the internal representation of early visual areas. Where the predictive coding predicts that the perceptual hypothesis is subtracted from the representation of early visual areas, the effective coding predicts that the representation of early visual areas is enhanced according to the perceptual hypothesis.

Chapter 2

Change blindness is influenced by both low-level image properties and high-level image representation

Wietske Zuiderbaan, Jonathan van Leeuwen, Serge O. Dumoulin (under review).

Acknowledgements of author contributions: WZ and SOD designed the experiment. WZ and JvL collected the data, WZ Performed data analysis under supervision of SOD. WZ wrote the manuscript and the other authors provided critical comments.

Abstract

Our visual system receives an enormous amount of information but not all information is retained. This is exemplified by the fact that subjects fail to detect large changes in a visual scene, i.e. change-blindness. Current theories propose that our ability to detect these changes is influenced by the high-level image interpretation of an image. On the other hand, low-level image features such as contrast-energy dominate the representation in early visual cortex. Here we investigated the interaction of the low-level feature contrast-energy and the higher-level image interpretation on our ability to detect changes in a visual scene. We measured reaction times while manipulating the low and high-level image properties using the change blindness paradigm. Our results suggest that our ability to detect changes in a visual scene, as measured with a change blindness paradigm, is not only influenced by the high-level image interpretation, but also by low-level image statistics such as contrast-energy. Furthermore, we find that the low-level and high-level image properties interact. We speculate that low-level and high-level image features are not independently represented in the visual system.

Introduction

Given the overload of information that our eyes receive, the visual system needs an efficient mechanism to extract the behaviorally relevant features and ignore other features. The change blindness paradigm reveals our inability to see changes in two sequentially presented visual scenes when separated by a disruption like a saccade or a flicker (for review see: Rensink, 2002; Simons and Ambinder, 2005). Without the disruption, the change is often easy to detect. Where our intuitive idea is that our visual representation of the outside world is highly detailed, the change blindness paradigm highlights limitations of this visual representation (Rensink, 2000; Martin et al., 2001). The failures to detect changes in the visual scene highlight the limits of our ability to retain and compare information from one visual scene to the other (Simons and Rensink, 2005).

Some changes in the visual scene are easier to detect than others. Many change blindness studies focus on the notion that changes made in parts of a scene that represent the general interpretation of the image are detected faster (Rensink et al., 1997; O'Regan et al., 1999; Shore and Klein, 2000; Stirk and Underwood, 2007; Sampanes et al., 2008). What causes these changes to be detected faster? One interpretation is that these parts of a scene receive more attention, and are therefore more likely to be encoded and compared (Simons and Rensink, 2005). Some studies refer to this high-level image representation as regions of high interest (Rensink et al., 1997; O'Regan et al., 1999; Shore and Klein, 2000; Verma and McOwan, 2010), where others refer to a semantic summary of the scene (gist) that is related to our knowledge of the world (Stirk and Underwood, 2007; Sampanes et al., 2008). In all scenarios, changes in high-level image representation are detected easier.

On the other hand, we know that the early visual system, i.e. primary visual cortex, encodes low-level image features such as contrast-energy (De Valois and De Valois, 1988; Boynton et al., 1999; Olman et al., 2004; Mante and Carandini, 2005; Dumoulin et al., 2008). Previous change blindness studies did not always account for differences in low-level image features (Verma and McOwan, 2010) and low-level features might contribute to

change detection, such as saliency or size of the change (Landman et al., 2003; Verma and McOwan, 2010).

Here, we ask whether low-level image features, in particular contrast-energy, contribute to our ability to detect changes in a visual scene. Furthermore, we compared change detection for these low-level image features with the high-level image representation. Last, we investigated whether our high-level image interpretation interacts with the low-level image feature contrast-energy in our ability to detect changes to a visual scene.

To this aim, we measured reaction times (RTs) in a flicker-task using the change-blindness paradigm (Rensink et al., 1997), subjects indicated when and where they detected the change between the images. We used images from the Berkeley Segmentation Dataset and Benchmark database (Martin et al., 2001). In this dataset, Martin and colleagues asked human observers to identify the most important aspects of the image (Fig. 1). We used these manually labeled aspects of the scene to define and quantify our measure for the high-level image representation. Using the combination of the natural images together with their manually labeled images, we were able to measure the amount of change that a manipulation to an image brings both to the low-level and to the high-level image representation. We manipulated the amount of change in the high-level image representation (manually labeled aspects of the scene) and in the change of the low-level image features (contrast-energy).

We found both significantly shorter RTs for manipulations dominated by a change in the high-level image representation, as well as those dominated by changes in contrast-energy. Furthermore, these two properties interacted: shortest RTs were found when the changes both contained a large change in contrast-energy and in the high-level image representation. These results show that our ability to detect changes in a visual scene is not only influenced by the high-level image interpretation of the image, but is also influenced by low-level image statistics such as contrast-energy. Finally, our results suggest that the low-level and high-level image features are not independently processed in the visual system, but interact with each other.

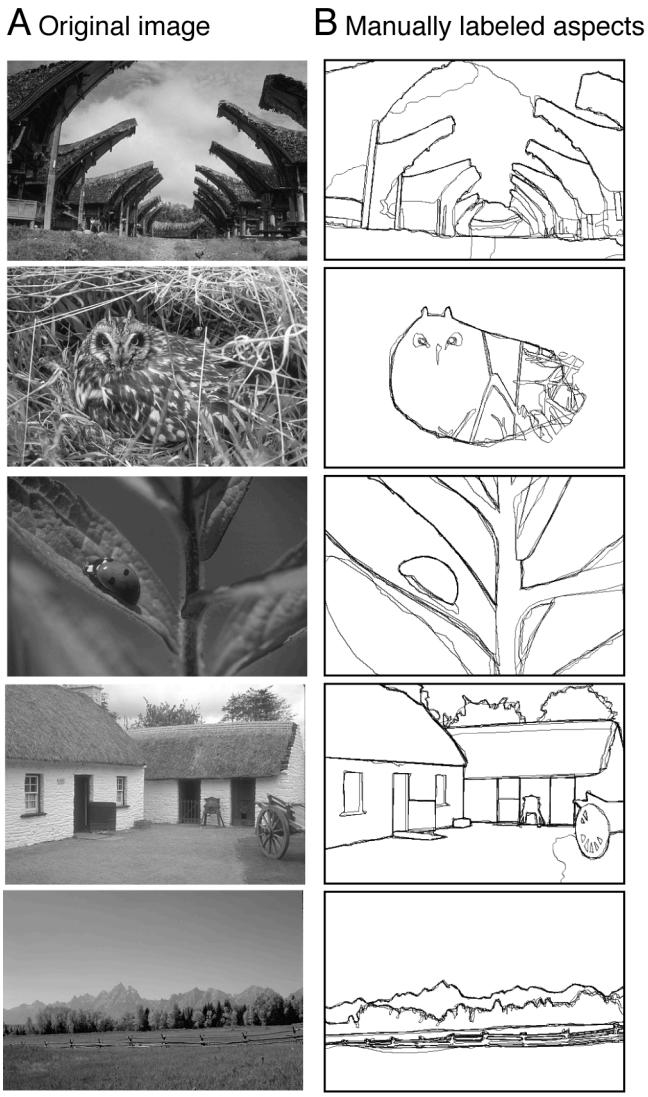


Figure 1 Five example images from the 'Berkeley Segmentation Dataset and Benchmark' database (Martin et al., 2001). For all the images of this database human observers manually identified the important aspects of the image. Different observers drew the labels, and their task was to draw lines on the image to highlight the parts of the image they considered to be important for the representation of the scene. We took the average manually labeled images of five observers (B) as our definition of the higher-level image interpretation. The pixels of the manually labeled images have values between 0 (not labeled) and 1 (pixel labeled by all 5 observers).

Methods

Participants

In total 60 subjects participated in the experiment (30 female, age range = [18 39], mean age = 23.9, SD = 4.0). The total number of 60 subjects was based on a power analysis informed by previous literature (Verma and McOwan, 2010). All subjects had normal or corrected-to-normal visual acuity. The study was approved by the local Ethics Committee of the Utrecht University and the experiments were carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All experiments were performed with the informed written consent of the subjects.

Apparatus

The experiment was programmed in MatLab (MathWorks, USA), using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Stimuli were presented on a 21-inch CRT monitor type LaCie-C22BW711 (60 Hz, 1024x768) using a Mac Pro 4.1 computer. The monitor was calibrated with a light meter type Gossen Mavo-Monitor USB. The viewing distance was 57 centimeters, which was maintained using a chin and forehead rest.

Stimuli

The images were taken from the 'Berkeley Segmentation Dataset and Benchmark' database (Martin et al., 2001). We used a selection of the grayscale images and we only used the images that were in landscape orientation. Also, we did not include any images in which a human was present in the image. Using Adobe Photoshop CS6 (Version 13.0, USA), we manipulated the image content. In total 108 different changes to the images were made, giving with 108 independent image pairs. Changes to the images were for instance deletions of (part of) objects, or changes in textures. Fig. 2 shows four example image pairs. The arrows indicate where the manipulation to the image was made.

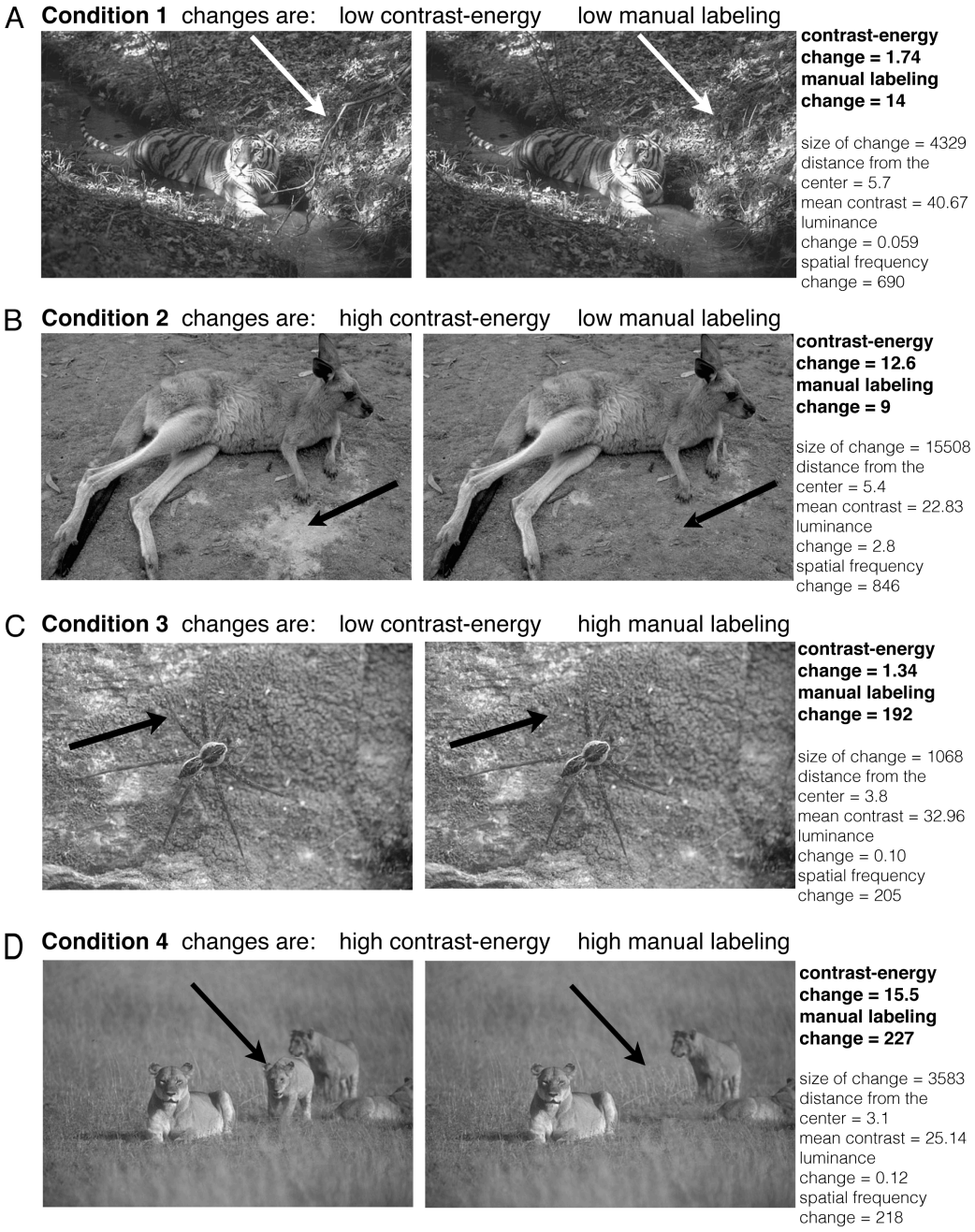


Figure 2 Examples of image pairs that were used in the experiment. The original images were taken from the 'Berkeley Segmentation Dataset and Benchmark' database (Martin et al., 2001). The arrows indicate where the manipulation in the image was made (the arrows were not present in the actual experiment). For every condition we show one example image pair. The conditions were based on the amount of contrast-energy change and the amount manual labeling change. The conditions were balanced for changes in size, distance from the center, contrast, luminance and spatial frequency.

All stimuli were presented on a gray background. The size of the images was 481x321 pixels, extending 18.6 x12.4 degrees of visual angle.

Measure of change in contrast-energy and manual labeling

The aim of the study was to investigate the role of the low-level feature contrast-energy and the high-level image interpretation on our ability to detect changes in a visual scene. We simultaneously measured the influence of both; for every image manipulation we calculated how big the change in the low-level feature contrast-energy was and its corresponding change in the high-level image representation.

Calculation of change in the low-level image property contrast

To compute the amount of contrast-energy change that the manipulation to an image brings, the first step was to calculate the local contrast-energy values both of the original image and of the manipulated image. For every pixel of the image we calculated the local contrast-energy inside a spatial window of neighboring pixels. This spatial window is defined by a Gaussian weighting function:

$$w_i = \exp - \left(\frac{(x_i - x_c)^2 + (y_i - y_c)^2}{2(\sigma)^2} \right) \quad (1)$$

where (x_i,y_i) is the location of the *i*th pixel and (x_c,y_c) is the location of the pixel at the center of the image patch. σ is the standard deviation of the Gaussian window and defines the size of the spatial window. For the size of σ, we used 19.4 pixels, which corresponds to 0.75° of visual angle.

The local contrast-energy value is based on the Root-Mean-Squared (RMS)-contrast (Pelli, 1997; Bex and Makous, 2002) which is defined as the standard deviation of the luminance intensities relative to the mean. The RMS-contrast is weighted by the Gaussian function to obtain the local contrast-energy value per pixel:

$$local\_contrast\_energy = \sqrt{\frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \frac{(L_i - L)^2}{L^2}} \quad (2)$$

Where w_i is the Gaussian weighting function. N Is the number of pixels in the spatial window. L is the mean luminance from the pixels inside the spatial window, and L_i is the luminance of the *i*th pixel.

Using the local contrast-energy values per pixel, we computed for every image pair the change in contrast-energy between the original image and the manipulated image. Fig. 3 gives an illustration of this procedure. We used the local contrast-energy values of the pixels that were changed in the image; the red line represents this region. The difference between the mean local contrast-energy values from the original image and the mean local contrast-energy values from the manipulated image defined our measure of contrast-energy change.

Calculation of change in subjective important aspects of the scene

For every image from the 'Berkeley Segmentation Dataset and Benchmark' database (Martin et al., 2001), there are manually labeled images available. In each manually labeled image, a human observer drew lines on the image to highlight the parts of the image they considered to be important for the representation of the scene. We used the average of 5 manually labeled images as our definition of the higher-level image interpretation. In Fig. 1 five examples of natural images (A) are shown together with the averaged manually labeled images (B).

We used the averaged manually labeled images to calculate the amount of change in the high-level image representation that a manipulation brings to an image. The pixels of the manually labeled images have values between 0 (not labeled) and 1 (pixel labeled by all 5 observers). The amount of change in the high-level image representation was calculated by summing the values of the pixels in the averaged manually labeled images that were manipulated. The computation is illustrated in Fig. 3. The red line in Fig. 3 represents the changed region. From this region we computed how big the change in the manually

labeled aspects of the image was and this gives us our measure of change in the higher-level image representation.

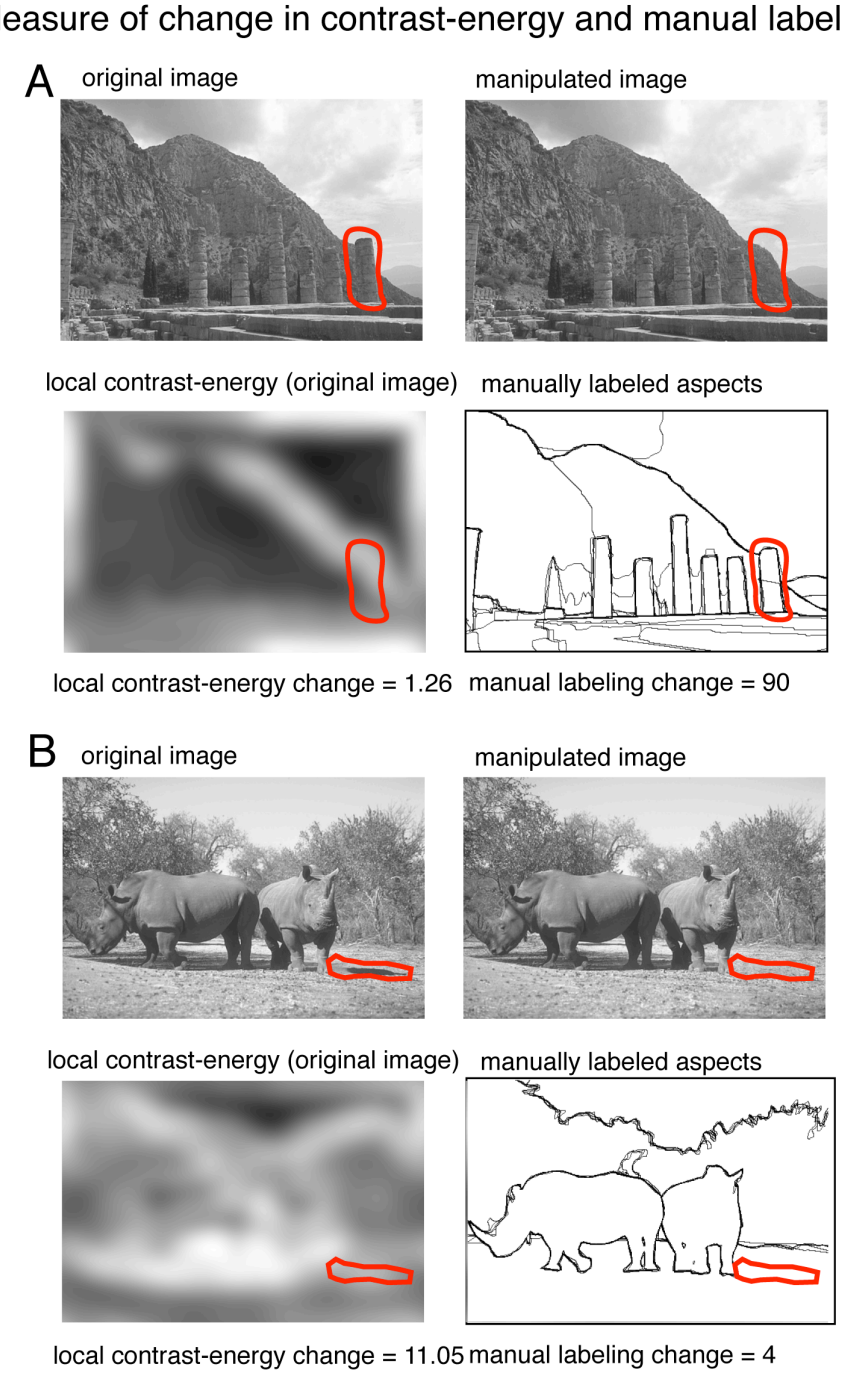


Figure 3 An illustration of the procedure of calculating the change in contrast-energy and manual labeling. Two example images (taken from the 'Berkeley Segmentation Dataset and Benchmark' database (Martin et al., 2001)) are shown with changes predominant in manually labeling (A) and contrast-energy (B). The red line represents the region manipulated to alter the image. From this region we calculated the difference in local contrast-energy as well as the amount of manual labeling change.

From this region we also computed the amount of change in contrast-energy. Two example images are shown: one in which the change is dominated by high-level aspects (Fig. 3A) and one in which the change is dominated by low-level aspects (Fig. 3B). The region was controlled for changes in size, distance from the center, mean luminance and mean contrast of the changed area and spatial frequency (see below).

Conditions

Based on the changes in contrast-energy and manually labeled aspects we defined four conditions based on the image-pairs that are 'low' and 'high' in their differences for contrast and manual labeling. We made a 2(contrast: high vs. low) x 2(manual labeling: high vs. low) within subjects design. The definition of 'low' and 'high' was based on the 50th percentile of the values for contrast-energy change and manual labeling change (median split). We took the first 50th percentile of the changes for the condition 'low' and the second 50th percentile for the condition 'high', both for contrast-energy as for the manually labeled aspects. Fig. 4 shows the histograms of the changes in contrast-energy (A) and manually labeled aspects (B); the vertical black striped line represents the 50th percentile. All the images below this boundary-line were placed in the 'low' condition. The images above the 50th percentile were placed in the 'high' condition. Fig. 2 shows for every condition an example of an image pair.

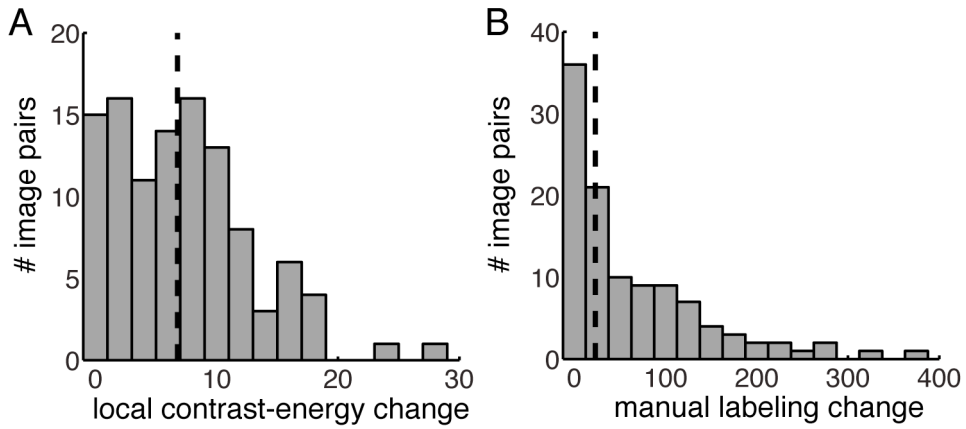


Figure 4 The distribution of local contrast-energy change (A) and change in manual labeling (B) for all the image pairs. The vertical striped black line indicates the 50th percentile used to define our four different conditions. Image pairs left of the 50th percentile were used in the 'low' condition, both for contrast-energy change and manual labeling change. Image pairs right to the 50th percentile were used in the 'high' conditions.

The conditions were balanced for changes in size, distance from the center, mean contrast, mean luminance change and spatial frequency change of the manipulated area. The size of the change was calculated as the number of pixels that were changed in the original image. The distance from the center was calculated as the eccentricity where the center of mass of the change was. The mean contrast was calculated as the average contrast of the manipulated area in the original image and the manipulated image. The luminance change was the difference of the mean luminance of the original image and the manipulated image in the area that was changed. The change in spatial frequency was the difference of the median spatial frequency of the changed area in the original image and the manipulated image. In Fig. 2 the corresponding values for the calculated changes are reported next to the image pairs. No significant changes were found using an ANOVA to test for group differences (changes in size: $F(3,104)=0.66$, $p=0.58$, distance from the center: $F(3,104)=1.75$, $p=0.16$, mean contrast: $F(3,104)=1.17$, $p=0.33$, maximum contrast of both images $F(3,104)=1.07$, $p=0.37$, minimum contrast of both images $F(3,104)=0.60$, $p=0.61$, mean luminance difference: $F(3,104)=1.03$, $p=0.38$, spatial frequency difference $F(3,104) = 0.31$, $p=0.81$), or the Kruskal-Wallis test to test for group differences (changes in size: $H(3,104)=2.76$, $p=0.43$, distance from the center: $H(3,104)=5.01$,$p=0.17$, mean

contrast: H(3,104)=4.18, p=0.24, maximum contrast of both images H(3,104)=4.51, p=0.21, minimum contrast of both images H(3,104)=1.84, p=0.61, mean luminance difference: H(3,104)=4.19, p=0.24, spatial frequency difference H(3,104) = 1.5, p=0.68).

We made 108 different image pairs (i.e. 108 different manipulations) and every condition contains 27 of these image pairs. Some of the images were used more than once, so different manipulations to the images could be made. To make sure that every subject did not see multiple manipulations to the same image, we made 6 different subsets, in which every original image was only inserted once. The subsets contained 36 different image pairs and the subsets were balanced so that every condition contained 9 individual image pairs. Every subset of images was presented to 10 subjects. Apart from these image pairs that were used for the experiment, we also used 4 trial image pairs. Every trial image pair is an example from one of the 4 conditions. These trial image pairs were used before the actual experiment began.

Design

For the change blindness experiment we used a flicker task (Rensink et al., 1997). In the flicker task, an original image was repeatedly alternated with a manipulated image, until the observer notices the change. The original and the manipulated image were presented for 600ms and were separated by a gray screen that was presented for 100ms. The separation with a gray screen of approximately 80ms prevents the observer from seeing the change due to movement transients caused by the change between the two images (Rensink et al., 1997). The maximum duration of the experiment was 240 seconds. Subjects were instructed to indicate when they saw the change by pressing the space bar. After this response, the last presented image from the image pair was presented again. At this image, the subjects indicated with the mouse where they had noticed the change in the image. After every trial the subject received feedback whether the change was correctly detected. The answer was classified correctly when the subject pointed within a small region of the change. The 36 different image pairs were presented in a randomized order for every individual experiment. The experiment started with four trial image pairs to

make the participants acquainted with the experiment, these trials were excluded from the analysis.

Analysis

We measured the time from the onset of the stimulus to the time at which the subject pressed the spacebar as the reaction time (RT). Trials in which the subject did not report to see a change (misses) or did not report the change at the accurate location (false alarms) were analyzed using repeated measures ANOVA. There were no significant differences in the proportion of false alarms ($F(3,236)=1.39$, $p=0.25$) and misses ($F(3,236)=1.62$, $p=0.19$) in the different conditions, so these trials were excluded from the analysis (61 out of 2160 trials).

The effect that the different conditions have on the RTs was analyzed using a general linear model (GLM, Matlab, Mathworks, USA). Since the distribution of RTs was skewed, we used an inverse Gaussian as a linking function in the GLM. Fig. 6B shows the inverse Gaussian fits to the data. Using the GLM approach we investigated the relation of the measured RTs with contrast-energy change, manual labeling change and the interaction between the two.

We further examined whether the interaction effect can be explained by statistical facilitation alone, i.e. RTs can be faster only because more than one feature is present. We compared the cumulative distribution functions (CDF) of the high-contrast/low-manual labeling (RT_{hl}) and low-contrast/high-manual labeling (RT_{lh}) conditions with the CDF of the high-contrast/high-manual labeling (RT_{hh}) condition (Fig. 6C). The upper bound of statistical facilitation was calculated using the race model inequality (Raab, 1962; Miller, 1982; 1986; Ulrich et al., 2007):

$$P(RT_{hh} < t) <= P(RT_{hl} < t) + P(RT_{lh} < t) \quad (3)$$

The race model inequality reflects the probability that an RT of the condition high-contrast/high-manual labeling is equal or less to the probability of the combined RT

distributions of the low-contrast/high-manual labeling and high-contrast/low-manual labeling. Shorter RTs for the condition high-contrast/high-manual labeling compared to the race-model reflect a violation of the race-model and can thus not be explained by statistical facilitation alone. Note that the race-model only provides an upper bound of statistical facilitation, since the CDF of the race-model sums to 2.

The CDFs were compared for different quantiles of the RTs (10, 20 and up to 90 percentile). We bootstrapped the CDFs over our subjects to obtain the 95% confidence interval and the p-values to test for a statistical violation of the race model for the different quantiles (corrected for multiple comparisons using the Bonferroni method).

Results

We measured detection rate and reaction times for the four different categories based on changes that are either low or high in contrast-energy (CE) and low or high in manual labeling of important aspects of the scene (ML).

The average detection rate for all changes in all image categories was 97%, i.e. all the changes to these images were detected within the 240 seconds that the images were presented. Fig. 5 shows the proportion of detections (hits), detections in which subjects failed to indicate the correct location (false alarms) and failures to detect the change (misses). Since there were no significant differences in the proportion of false alarms and misses in the different conditions we excluded these trials from the analysis. Thus only the trials in which the subjects correctly detected the changes to the images were used for further analysis.

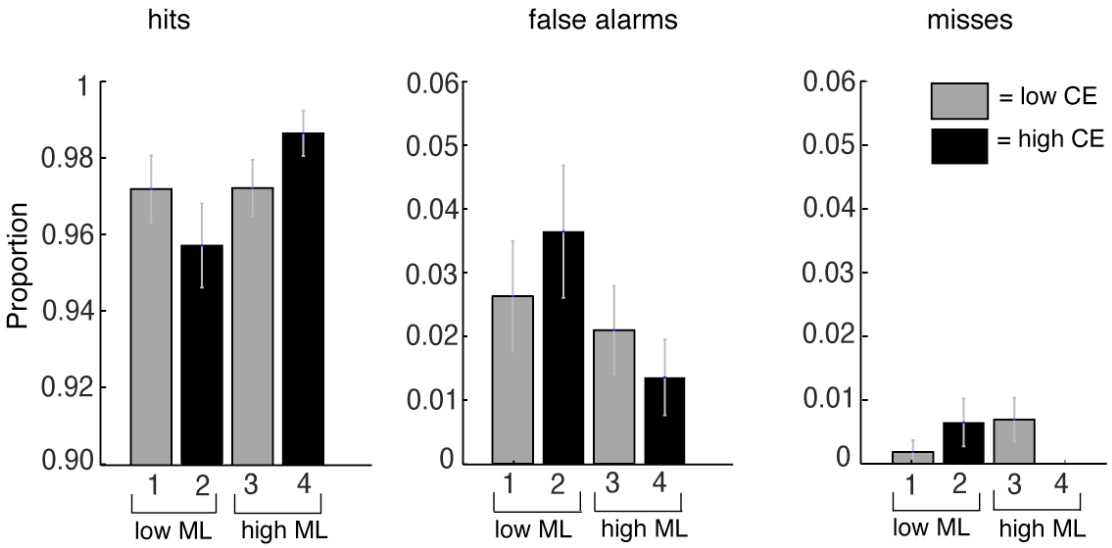


Figure 5 Proportion of detections (hits), detections in which subjects failed to indicate the correct location (false alarms) and the failure to detect the change (misses). The averaged data are the mean from all subjects, and the error bars reflect 1 standard error of the mean.

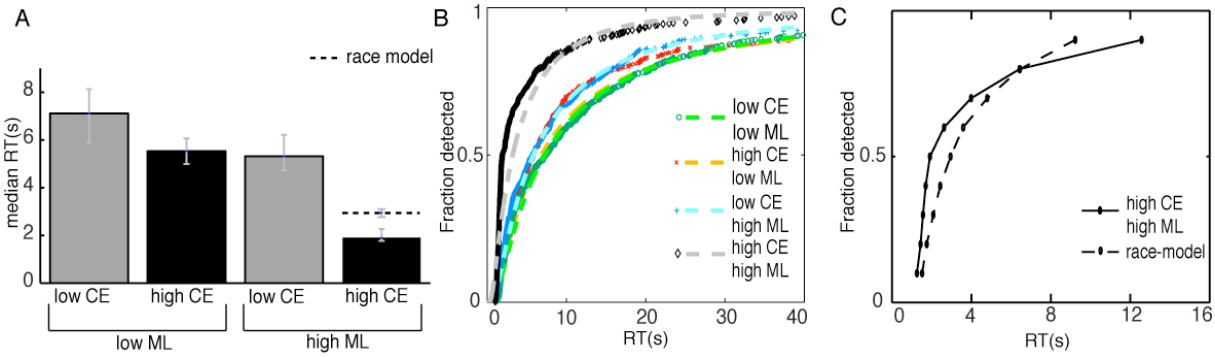


Figure 6 A: The median reaction times of all the correct detections for all subjects for the different image categories. The striped line represents the median RT of the race-model. The error bars reflect the bootstrapped 95% confidence interval. B: the cumulative distributions of the correct responses for the different conditions. We analyzed our results using a GLM approach, with the inverse Gaussian as a linking function. The striped lines are the fits to the data with an inverse Gaussian function. We found significantly shorter reaction times for both changes in contrast-energy (CE) and manual labeling (ML). Furthermore, we found a significant interaction effect, i.e. reaction times were shorter when changes affected both contrast-energy and manual labeling. C: The comparison of the CDF of the condition high-contrast/high-manual labeling with the CDF of the race-model. The CDF of the race-model is the summed CDFs of the conditions high-contrast/low-manual labeling and low-contrast/high-manual labeling. We found statistical significant shorter RTs for the condition high-contrast/high-manual labeling compared to the race-model for the 10^{th} to the 70^{th} percentile. This indicates that the interaction effect cannot be explained by statistical facilitation alone.

We measured the time from the beginning of the stimulus, until the time the subjects noticed the change and pressed the spacebar as the RT. Fig. 6A shows the median RTs per condition, the error bars reflect the bootstrapped 95% confidence intervals. The condition in which the manipulations to the image were the most difficult to detect was when the manipulation of the image was both low in manual labeling change and low in contrast-energy change, i.e. in this condition the RT was longest (median RT = 7.1 s). Comparing this condition to the condition in which the manipulations were also low in manual labeling change but were high in contrast-energy change, we see that these were

detected faster (median RT = 5.5 s, median decrease RT= 1.6 s). Comparing the first condition to the condition in which manipulations were low in the change of contrast-energy, but the changes in manual labeling were high, we also found a faster detection of the manipulations (median RT = 5.3 s, median decrease RT= 1.8 s). Last, when the manipulations were both high in the change of contrast-energy and high in the change of manual labeling, we found that the manipulations were detected fastest (median RT = 1.9 s, median decrease RT= 5.2 s). These results indicate an effect both for the amount of manual labeling change and the amount of contrast-energy change on the RT. A manipulation to an image with a larger change in either contrast-energy or manual labeling to the image led to a faster detection of the manipulation. The results also suggest an interaction effect for the amount of contrast-energy change and the amount of manually labeled change. A manipulation that was both high in its change for contrast-energy and manual labeling resulted in the fastest detection of the manipulation.

To quantify these observations, we performed a GLM-analysis to compare the effects of contrast-energy change, manual labeling change and the interaction of the two on the RTs. Since the distribution of RTs was skewed, we did not use the canonical linking function that tests for differences in the conditions based on their mean. Instead of the canonical linking function, we used the inverse Gaussian linking function to take into account the non-normal or non-Gaussian distribution of the data. The differences of the distributions are now based on the fit of the inverse Gaussian distribution. Fig. 6B shows the fits of the inverse Gaussian function to the distribution of the data for the different conditions. All fits had a variance explained $R^2 > 0.97$. Using the GLM analysis, we found a significant effect for the level of contrast-energy change: $t(2095)=9.92$, $p<<0.001$. Also, we found a significant effect for the level of manual labeling change: $t(2095)=6.54$, $p<<0.001$. Furthermore, we found a significant effect for the interaction of the amount of contrast-energy change and the amount of manual labeling change: $t(2095)=7.90$, $p<<0.001$. Thus, our ability to detect changes in a visual scene is influenced by both the higher-level image representation and the low-level image properties, and they interact.

To further examine this interaction effect, we investigated whether it can be explained by statistical facilitation alone. We compared the CDF of the high-contrast/high-manual labeling condition to the CDF of the race model (Fig. 6C). We compared the bootstrapped CDFs at different quantiles, from the 10^{th} to the 90^{th} percentile. For the 10^{th} to the 70^{th} percentile we found a statistically significant violation of the race model ($p < 0.001$ for all significant percentiles, Bonferroni corrected for multiple comparisons), with a maximum of 1 second difference in RT that was unexplained by statistical facilitation.

Discussion

In our study we used manually labeled images to investigate the simultaneous effect of low-level and high-level image features on our ability to detect changes in a visual scene. We measured RTs within a change blindness task and compared RTs over four different categories. The categories were based on the amount of change in the low-level image feature contrast-energy and the amount of change in the high-level image representation.

Shorter RTs for high change in the manually labeled aspects of the image

Analyzing the reaction times for the different categories, first, we found a significant effect for manual labeling. RTs were shorter for manipulations that have a larger change in their high-level image representation. This measure of change was based on the manually labeled aspects of the images, in which people indicated which parts represented the image most (Martin et al., 2001). Our results are in line with previous change blindness studies, which showed that changes made in parts of a scene that were considered to be important for their high-level image representation were detected earlier (Rensink et al., 1997; O'Regan et al., 1999; Shore and Klein, 2000; Stirk and Underwood, 2007; Sampanes et al., 2008; Verma and McOwan, 2010). However, different notions of the high-level image representation have been used in these studies. Some refer to it as the gist; the general interpretation of a scene (Stirk and Underwood, 2007; Sampanes et al., 2008), where others refer to it as regions of high interest (Rensink et al., 1997; O'Regan et al., 1999;

Shore and Klein, 2000; Verma and McOwan, 2010). Where our study resembles most with the studies that use the notion of the high-level image representation as regions of high interest, our definition of the high-level image representation can also partly be seen as capturing properties of a figure-ground representation of the image. Mazza et al. found that changes in images that belong to the figure are easier to detect than changes belonging to the ground of that image (Mazza et al., 2005). Where this study used non-natural experimental stimuli, we show that it holds for natural images as well.

Arguably, the manual labeled images represent figure-ground representations. Figure-ground may be thought of as more of a mid-level image representation than a high-level one, but we believe this is not the only property that is captured in the manual labeling. The manually labeled images were made subjectively by human observers with the intention to develop a ground-truth database to test automatic segmentation algorithms. Different cues for the manual labeling can be used, from low-level and mid-level to high-level knowledge based cues. So even if the manually labeled aspects of the image partly capture a figure-ground representation, we believe it also captures the high-level properties of an image.

Shorter RTs for high contrast-energy change

We also found a significant effect for the amount of contrast-energy change on the RTs. Manipulations with a larger change in contrast-energy were detected earlier. Neurophysiological data shows that the early visual cortex responds strongly to changes in contrast-energy (De Valois and De Valois, 1988; Boynton et al., 1999; Olman et al., 2004; Mante and Carandini, 2005; Dumoulin et al., 2008). We speculate that this preference of early visual cortex underlies the ability to detect these changes. Alternatively, changes in contrast-energy might attract more (exogenous) attention. These two explanations are not mutually exclusive and both or either one may explain why changes in contrast might be detected earlier.

Most change blindness studies do not account for effects of low-level image statistics. Some studies did, but show deviating results. Verma & McOwen find earlier detection for

changes in a region with high salience (Verma and McOwan, 2010). On the other hand, Stirk & Underwood do not find any differences based on saliency (Stirk and Underwood, 2007). The definition of saliency in these studies is based on variations in color, luminance and orientation (Itti and Koch, 2000). Here, we show that low-level image features, and in particular contrast, can lead to differences in the detection of changes.

Interaction of change in contrast-energy and manual labeling

Last, we found a significant interaction effect for the amount of change in manual labeling and contrast-energy change. By using the combination of the manually labeled images simultaneously with the inherent contrast energy of the images, we were able to investigate not only the effect of low-level and high-level image statistics on the change blindness paradigm. Most importantly, we were able to simultaneously measure the effect of both because the levels of contrast change and change in manual labeling was independent in the images. The interaction we find cannot be explained by the added benefit of two independent observations (race-model). This suggests that the low-level and high-level image features are not independently processed in the visual system, but instead they facilitate each other.

Our result that the high-level image representation interacts with contrast-energy is in line with the study of Self et al., where they find differences in the perceived contrast between figure and ground. The contrast of Gabor probes that are placed on figures are perceived as higher contrast as the probes placed on the background (Self et al., 2015). Where they use non-naturalistic stimuli, we confirm their findings of this interaction effect with natural images.

The objects that we perceive are segmented by the visual system through a hierarchy of cortical areas that extract increasingly complex forms of information. Different theories propose that higher-order areas feedback to lower-order areas and modulate their responses according to the prior expectations about the visual world (Mumford, 1992; Koch and Poggio, 1999; Rao and Ballard, 1999; Friston, 2005). We speculate that this

interaction is the origin of our interaction effect of the high-level image representation and the low-level image statistics.

Attention, memory and awareness

Change blindness is often explained in terms of attention and memory. Attention is thought to have a role in the encoding of a coherent representation for retention across disruptions (Rensink, 2000). The storage of information in working memory is determined by attention. In change blindness, subjects have to compare two image representations. Working memory is supposed to play an important role in this comparison process. Besides the working memory, a different memory storage is proposed that stores the image features in greater detail and is not dependent on attention. This is called the iconic memory. Where the iconic memory decays fast and is easily overwritten by new stimuli, the working memory is longer lasting.

Another way of interpreting the difference between iconic memory and working memory is that the iconic memory is led by bottom-up attention, which is drawn to endogenous low-level features. Whereas the working memory is drawn by top-down attention driven by exogenous high-level cues. In line with this idea, studies showed that short term memory for low-level features such as contrast (Xing et al., 2013) and orientation (Harrison and Tong, 2009) resides in early visual cortex.

Lamme & Roelfsema make the distinction of feedforward and recurrent processing (Lamme and Roelfsema, 2000). The feedforward sweep is thought to be pre-attentive and encodes the visual information based on the low-level image statistics. The recurrent process plays a role in the high-level image interpretation and is driven by attention. It is supposed that the recurrent processes are needed for visual awareness. Only recurrent processing reaches a conscious state. The feedforward sweep and the recurrent processes are not independent, but influence eachother. Our internal representation of a visual scene is build up from these interactions.

These processes relate to the results of our study. Where the effects for the low-level feature contrast-energy can be acquainted to the iconic memory or the feedforward sweep, the effect for high-level features can be acquainted to the working memory. Furthermore, in the recurrent processes low-level and high-level image properties influence each other, which can be related to the interaction effect that we find, and in particular to the facilitating role that we find when comparing the RTs to the race-model.

Conclusion

Where earlier studies on change blindness conclude that our ability to detect changes in a visual scene is influenced by the higher-level image interpretation of the scene, we find that it is not only the higher-level image interpretation we also find that it is influenced by the low-level feature contrast-energy. Manipulations that are either high in their change for the high-level image interpretation or contrast-energy are detected earlier. We also found an interaction effect; manipulations that were both high in their change for manual labeling and contrast-energy were detected earliest.

Presumably, these parts of an image are processed in greater detail. The change blindness paradigm can be used to give more insight in how low-level image statistics and high-level image representations are processed in the visual system. Our results suggest that they are not processed independently, but instead interact in how they are represented in the visual system.

Chapter 3

Modeling center-surround configurations in population receptive fields using fMRI

Published as:

Wietske Zuiderbaan, Ben M. Harvey, Serge O. Dumoulin (2012). Modeling center-surround configurations in population receptive fields using fMRI. *Journal of Vision,* 12(3),10.

Acknowledgements of author contributions: WZ and SOD designed the experiment. WZ and BMH collected the data, WZ performed data analysis under supervision of SOD. WZ wrote the manuscript and the other authors provided critical comments.

Abstract

Antagonistic center-surround configurations are a central organizational principle of our visual system. In visual cortex stimulation outside the classical receptive field can decrease neural activity and also decrease functional Magnetic Resonance Imaging (fMRI) signal amplitudes. Decreased fMRI amplitudes below baseline -0% contrast- are often referred to as "negative" responses. Using neural model-based fMRI data-analyses, we can estimate the region of visual space to which each cortical location responds, i.e. the population receptive field (pRF). Current models of the pRF do not account for a center-surround organization or negative fMRI responses. Here, we extend the pRF model by adding surround suppression. Where the conventional model uses a circular-symmetric Gaussian function to describe the pRF, the new model uses a circular-symmetric difference of Gaussians (DoG) function. The DoG-model allows the pRF analysis to capture fMRI signals below baseline and surround suppression. Comparing the fits of the models, an increased variance explained is found for the DoG-model. This improvement was predominantly present in V1/2/3 and decreased in later visual areas. The improvement of the fits was particularly striking in the parts of the fMRI signal below baseline. Estimates for the surround-size of the pRF show an increase with eccentricity and over visual areas V1/2/3. For the suppression index, which is based on the ratio between the volumes of both Gaussians, we show a decrease over visual areas V1 and V2. Using non-invasive fMRI techniques, this method gives the possibility to examine assumptions about center-surround receptive fields in human subjects.

Introduction

Antagonistic receptive fields are found throughout the visual system. In the visual cortex, the responses to stimulation in the classical receptive fields can be modulated by stimulation in the extra-classical receptive field. These modulations can be excitatory or inhibitory and have been characterized in detail by electrophysiological and psychophysical studies (for review see: (Allman, Miezin, & McGuinness, 1985; Carandini, 2004; Cavanaugh, Bair, & Movshon, 2002; Fitzpatrick, 2000; Hubel & Wiesel, 1968)). Recent human brain imaging techniques have also found evidence of these center-surround organizations. Using functional magnetic resonance imaging (fMRI), center-surround configurations have been implicated in the amplitude changes of the fMRI signal due to stimulus size (Kastner, De Weerd, Pinsk, Elizondo, Desimone & Ungerleider, 2001; Nurminen, Kilpelainen, Laurinen, & Vanni, 2009; Press, Brewer, Dougherty, Wade, & Wandell, 2001), surround (Tajima et al., 2010; Williams, Singh, & Smith, 2003; Zenger-Landolt & Heeger, 2003) or contextual modulations (Dumoulin & Hess, 2006; Harrison, Penny, Ashburner, Trujillo-Barreto, & Friston, 2007; Kastner et al., 2001; Murray, Kersten, Olshausen, Schrater, & Woods, 2002).

In addition, when measuring recording sites in the vicinity of regions that are positively correlated with a stimulus manipulation, *negative Blood Oxygenation Level Dependent (BOLD) responses* (NBRs) have been reported in the visual cortex (Shmuel, Augath, Oeltermann, & Logothetis, 2006; Shmuel et al., 2002; Smith, Williams, & Singh, 2004). These NBRs are defined as BOLD responses below those elicited by viewing mean-luminance gray (0% contrast). Also, when identifying or reconstructing visual stimuli from the fMRI signals, local cortical sites may contribute either positively or negatively (Kay, Naselaris, Prenger, & Gallant, 2008; Miyawaki et al., 2008). The coupling between measured fMRI responses and the underlying neural activity is complex (for review see: (Logothetis & Wandell, 2004)) and neural inhibition per se may not necessarily result in a decrease of the BOLD response. Studies combining electrophysiology and fMRI report that these NBRs are caused by decreases in overall neural activation (Shmuel et al., 2006; Shmuel et al., 2002). They state that the NBR can be either caused by a suppression of the local neuronal

activity and/or a decrease in afferent input, both caused by the activation of the positively responding regions. These results suggest that the neural center-surround organization influences the fMRI signal, and also suggest that the spatial center-surround profile of the recorded neural population may be resolvable at the resolution of fMRI.

Here, we extended a computational method for estimating receptive fields of a population of neurons (Dumoulin & Wandell, 2008), to measure center-surround configurations throughout the visual cortex. Since a receptive field measured with fMRI is estimated from a neuronal population instead of a single neuron, we refer to it as the *population receptive field* (pRF) (Dumoulin & Wandell, 2008; Victor, Purpura, Katz, & Mao, 1994). This method fits a model of the pRF to the fMRI data. We compare two models of the pRF. The original pRF-method describes the pRF with one single circular symmetric Gaussian (OG) in the visual field. This model can only represent positive responses to stimulation in any region of the visual field, and can therefore not explain negative fMRI responses. We extended the current pRF model by incorporating a suppressive surround using a *Difference of Gaussian* (DoG) function to represent the pRF. This model can account for suppression effects.

The DoG-model yielded improved performance in predicting the fMRI time-series. This improvement was predominantly present in V1/2/3 and decreased in later visual areas. We attribute the absence of a measurable center-surround configuration in later visual areas to technical limitations, which would hide this configuration at the resolution of fMRI though it may be present at neural resolutions. The properties of the center-surround configurations varied systematically across V1/2/3. The pRF center sizes increased with eccentricity and visual field maps, similar to previous reports (Amano, Wandell, & Dumoulin, 2009; Dumoulin & Wandell, 2008; Harvey & Dumoulin, 2011; Kay et al., 2008; Winawer, Horiguchi, Sayres, Amano, & Wandell, 2010). In addition, we show that the surround size of the pRF increases with eccentricity and up the visual field map hierarchy. To compute the total amount of suppression we need to take into account the surround volume not just its size. Therefore we defined the suppression index by the volume ratio of the two individual Gaussian components of the DoG-model. The

suppression index decreases with eccentricity and also decreases up the visual field map hierarchy. These results extend the notion of center-surround configurations to the spatial scale of fMRI measurements. This gives the ability to measure and test assumptions of center-surround population receptive fields in human subjects.

Methods

Subjects

Four subjects (one female; ages 21-37 years) participated in this study. All subjects had normal or corrected-to-normal visual acuity. All studies were performed with the informed written consent of the subjects and were approved by the Human Ethics Committee of University Medical Center Utrecht.

Stimulus Presentation

The visual stimuli were generated in Matlab using the PsychToolbox (Brainard, 1997; Pelli, 1997) on a Macintosh Macbook Pro. Stimuli were displayed by a configuration where optics back-projected the imaged stimuli of the projector onto a screen located outside the MRI bore. The subjects viewed the screen by mirrors placed on top of the scanner head coil. The stimulus radius was 6.25° of visual angle.

Stimulus Description

For the experiment, we used moving bar apertures which revealed a moving checkerboard pattern (100% contrast) that moved parallel to the bar orientation (Figure 1). The width of the bar subtended $1/4^{th}$ of the stimulus radius (1.56°). Four bar orientations (0°, 45°, 90° and 135°) and two different motion directions for each bar were used, giving a total of 8 different bar configurations within a given scan. The bar sweeps across the stimulus aperture in 20 steps (each 0.625° and 1.5 seconds), each pass taking 30 seconds. After a

horizontally or vertically oriented sweep, a period of 30 seconds of mean-luminance (0% contrast) is presented. This gives a total of 4 blocks of mean-luminance during each scan, presented at evenly spaced intervals.

To make sure subjects fixated at the center of the screen, a small fixation dot (0.125° radius) was presented in the middle of the stimulus. This fixation dot changed color (red-green) at random time intervals and subjects were instructed to respond at this change of color. An air pressure button was used to record these responses. When performance on the fixation task was below 75% correct, the scan was discarded (one scan total).

Eye movement data examining fixation during the presentation of this stimulus were measured outside the scanner using a highly accurate Eyelink II system (SG Research, Mississauga, Ontario, Canada). Values obtained during moving bar presentations and the mean-luminance blank periods were not significantly different, and so presentation of the moving bar did not appear to affect eye movements. Subjects had a fixation position distribution with a standard deviation of 0.22°, indicating a highly accurate fixation. Because there is inevitably some noise in the measurement of eye positions, this value represents an upper bound of the actual variation of fixation positions (Harvey & Dumoulin, 2011).

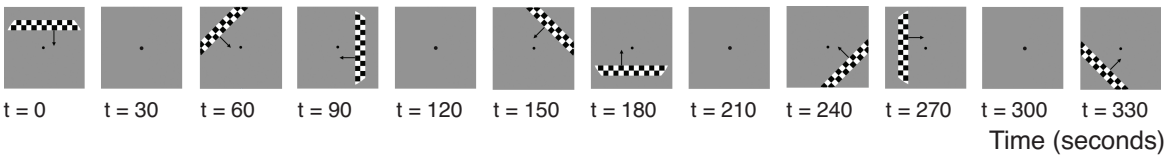


Figure 1 A schematic illustration of the stimulus sequence. The bar apertures revealed a high contrast checkerboard (100%). The checkerboard rows moved in opposite directions along the orientation of the bar. The bars moved trough the visual field with 8 different orientation-motion configurations as indicated by the arrows (arrows not present in actual stimulus). The total stimulus sequence lasted 360 seconds.

Functional imaging and processing

MRI data was acquired using a Philips Achieva 3T scanner with an 8-channel SENSE head-coil. The participants were scanned with a 2d-echo-planar-imaging sequence with 24 slices oriented perpendicular to the Calcarine sulcus with no gap. The following parameters were used; *repetition time* (TR) = 1500 ms, *echo time* (TE) = 30 ms and a flip angle of 70°. The functional resolution was 2.5x2.5x2.5 mm, given a *field of view* (FOV) of 224x224 mm.

The duration of each scan was 372 seconds (248 time-frames), of which the first 12 seconds (8 time-frames) were discarded due to start-up magnetization transients. For every subject 9 or 10 scans were acquired during the same session. These repeated scans were averaged to obtain a high signal-to-noise ratio.

Foam padding was used to minimize head movement. The functional images were corrected for head movement between and within the scans (Nestares & Heeger, 2000). For computation of the head movement between scans, the first functional volumes for each scan were aligned. Within scan motion correction was computed by aligning the frames of a scan to the first frame.

Anatomical imaging and processing

The T1-weighted MRI images were acquired in a separate session using an 8-channel SENSE head-coil. The following parameters were used: TR/TE/flip angle = 9.88/4.59/8. The scans were acquired at a resolution of 0.79x0.80x0.80 mm and were resampled to a resolution of $1mm^3$ isotropic. The functional MRI scans were aligned with the anatomical MRI using an automatic alignment technique (Nestares & Heeger, 2000). From the anatomical MRI white matter was automatically segmented using the *FMRIB's Software Library* (FSL) (S. M. Smith et al., 2004). After the automatic segmentation it was hand-edited to minimize segmentation errors (Yushkevich et al., 2006). The gray matter was grown from the white matter and to form a 4 mm layer surrounding the white matter. A smoothed 3D cortical surface can be rendered by reconstruction of the cortical surface at the border of the white and gray matter (Wandell, Chial, & Backus, 2000).

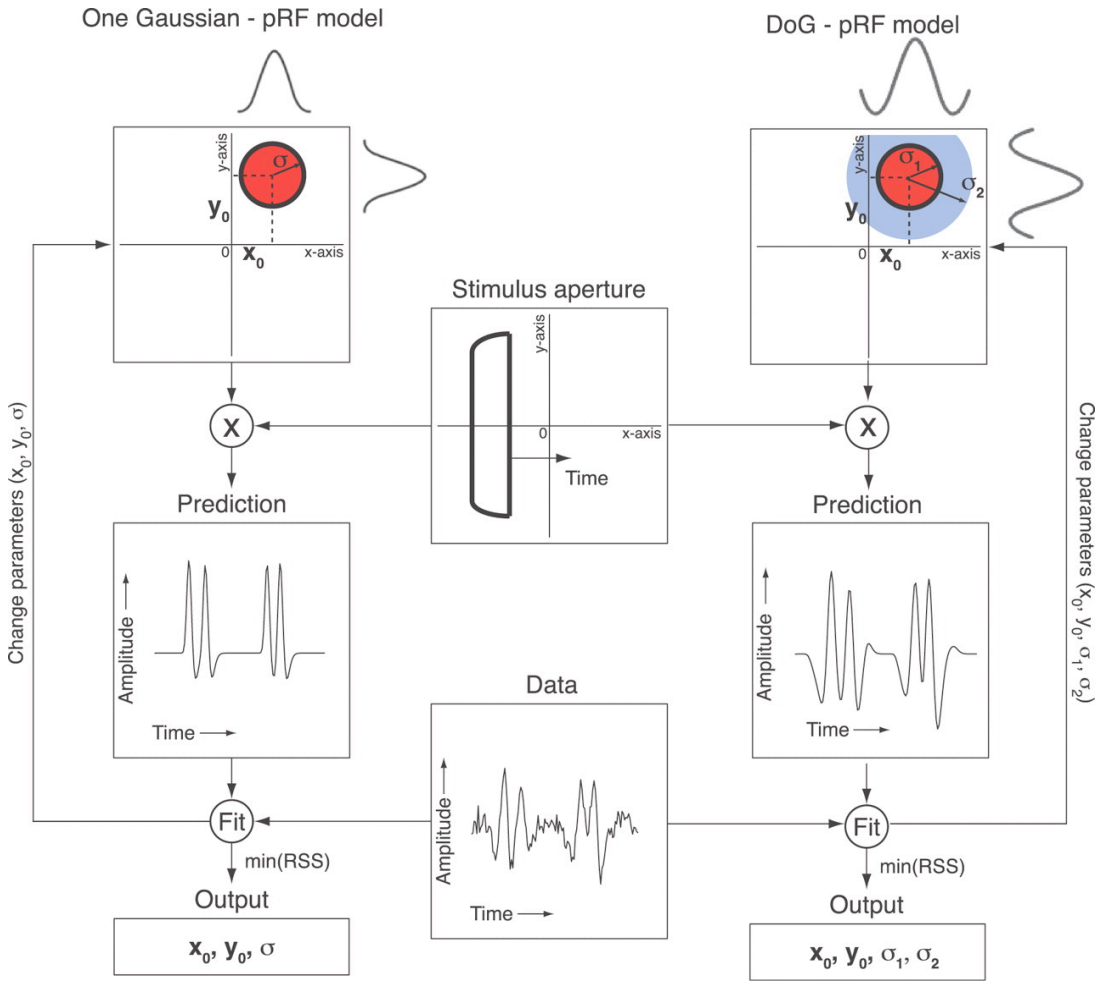


Figure 2 Flow chart of the fitting procedure of the two pRF models. The left panels illustrate the conventional one Gaussian pRF model, the right panels illustrate the difference of Gaussians pRF model. The middle panels indicate the analysis input, shared between both pRF models, consisting of the stimulus aperture and fMRI data. Convolution of a pRF model with the stimulus sequence predicts the fMRI time series. The pRF model parameters are estimated by minimizing the sum-of-squared errors between the predicted and the measured time-series.

pRF model-based analysis

The model estimates a pRF for every cortical location using a method previously described (Dumoulin & Wandell, 2008). In short, the method estimates the pRF by combining the measured fMRI time-series with the position time course of the visual stimulus. A prediction of the time-series is made by calculating the overlap of the pRF and the stimulus energy and a convolution with the haemodynamic response function (HRF). The optimal model parameters are chosen by minimizing the residual sum of squares between the predicted and measured time-series.

We compared two different models of the pRF. The conventional pRF model consists of one circular symmetric Gaussian (OG). The OG-model has four parameters: position (x_0,y_0), size (σ_1), and amplitude (β_1). Here we extend the pRF model to add an inhibitory surround. This model represents the pRF using a difference of Gaussians (DoG) function. The DoG-model is made up of a subtraction of 2 Gaussian functions, in which the Gaussian with the largest standard deviation is subtracted from the smaller one. The center of the two Gaussians is at the same position, and therefore the DoG adds two extra parameters to the model; the size of the negative surround (σ_2) and the amplitude of the surround (β_2). All parameters are in degrees of visual angle (°), except for the amplitudes (% BOLD/deg^2/sec). Restrictions of the DoG-model are that σ_2 is larger than or equal to σ_1, and β_2 is negative with an absolute value that is smaller than β_1. These restrictions ensure a center-surround configuration.

Thus the DoG pRF model ($g(x,y)$) is defined as a combination of two Gaussians ($g_+(x,y)$ and ($g_-(x,y)$):

$$g_+(x,y) = \exp - \left(\frac{(x-x_0)^2 + (y-y_0)^2}{2(\sigma_1)^2} \right) \quad (1)$$

$$g_{-}(x,y) = \exp -\left(\frac{(x-x_0)^2 + (y-y_0)^2}{2(\sigma_2)^2} \right) \quad (2)$$

$$g(x,y) = g_{+}(x,y) - g_{-}(x,y) \quad (3)$$

The next step is to define the stimulus. Assuming that all parts of the stimulus contribute equally to the fMRI response (Engel, Glover, & Wandell, 1997) the stimulus is defined as a binary indicator that marks over time each position of the stimulus in the visual field; s(x,y,t). Combining the position of the pRF in the visual field with the stimulus positions over time, the response of the pRF, r(t) is obtained by calculating the overlap for each Gaussian. We obtain two values of r(t); $r_{+}(t)$ for the positive Gaussian and $r_{-}(t)$ for the negative Gaussian.

$$r_{+}(t) = \sum_{x,y} s(x,y,t) g_{+}(x,y) \quad (4)$$

$$r_{-}(t) = \sum_{x,y} s(x,y,t) g_{-}(x,y) \quad (5)$$

The prediction of the time-series, p(t) is then calculated by convolution of the response of the pRF, r(t), with the HRF, h(t). Two different predictions of the time-series are made: one for the positive Gaussian and one for the negative Gaussian of the pRF. Since the negative BOLD responses exhibit a similar HRF as the positive BOLD response (Shmuel et al., 2006), the same HRF (h(t)) is used for both parts of the pRF.

$$p_{+}(t) = r_{+}(t) * h(t) \quad (6)$$

$$p_{-}(t) = r_{-}(t) * h(t) \quad (7)$$

Where * denotes convolution. Assuming there is a linear relationship between the blood oxygenation levels and the MR signal (Birn, Saad, & Bandettini, 2001; Boynton, Engel, Glover, & Heeger, 1996; Hansen, David, & Gallant, 2004), a scaling factor β_1 is used on p(t). This scaling factor accounts for the unknown units of the fMRI signal, which in turn represents the amplitude of the pRF. β_1 is calculated by using a *general linear model* (GLM) in which measurement noise, e, is taken into account as well. The two values for the amplitudes of the Gaussians are calculated using a GLM with two unknown values for β; β_1 for the positive amplitude and β_2 for the negative amplitude.

$$y(t) = p_+(t)\beta_1 + p_-(t)\beta_2 + e \tag{8}$$

The optimal parameters of the pRF are estimated by minimizing the RSS.

$$RSS = \sum_t (y(t) - (p_+(t)\beta_1 + p_-(t)\beta_2))^2 \tag{9}$$

To compare the performance of the two different models in predicting the fMRI time series, the variance explained predicted to measured fMRI time-series was computed.

Since it is possible that not all the voxels show a center-surround configuration, it should still be possible to obtain a configuration of the pRF without the suppressive surround. Therefore, we replaced all the voxels in the DoG-model where the OG-model fits better with the pRFs estimated by the OG-model.

For further technical and implementation details see (Dumoulin & Wandell, 2008).

Baseline estimation

The fMRI signal is without units, and the baseline activation must be indirectly deduced from the parts of the fMRI time-series where a baseline stimulus was shown (mean luminance, 0% contrast). The conventional pRF model-based analysis estimates this baseline level as a component within the GLM fit over the whole time-series. Initial

observations indicated different parameters for the baseline in the two different pRF models (Supplementary figure 1). Different baseline values suggest that the conventional pRF model may compensate for the negative fMRI signal by lowering its estimate of baseline activity. To ensure that differences in variance explained between the various models are not interfered by differences of the baseline estimation, the time-series of the mean luminance blocks are used to estimate the baseline. The standard HRF model (Glover, 1999) takes about 22.5 seconds to diminish to 0.55 % of its maximal amplitude. Therefore, to remove all fMRI signals caused by the stimuli, the first 22.5 seconds of the mean luminance blocks are removed, and the remainder of 7.5 seconds is used to estimate the baseline. Since every scan contains 4 sequences of mean-luminance (Fig 1), a total of 30 seconds of the fMRI time-series is used to estimate the level of the baseline activation. This baseline correction is required when comparing the two models using the same baseline, but is not necessary when estimating each model independently.

Region of Interest

Using the estimated position parameters x_0, and y_0 of the pRF model, values for the polar angle ($\text{atan}(y_0/ x_0)$) and eccentricity ($\sqrt{(x_0^2 + y_0^2)}$) can be calculated. By rendering these polar angle and eccentricity maps onto the inflated cortical surface (Wandell et al., 2000), the borders of the visual areas can be drawn on the basis of their location in the visual field (DeYoe et al., 1996; Engel et al., 1997; Sereno et al., 1995; Wandell, Dumoulin, & Brewer, 2007). Visual areas V1, V2, V3, V3a, hV4 and LO-1 are defined as *regions of interest* (ROIs).

Surround Suppression Index

To indicate the effect of the suppressive surround on the pRF we use a measure for its suppression index that is based on the volumes under the two Gaussians that make the pRF (adapted from: (Sceniak, Hawken, & Shapley, 2001)). We only take into account the total volume of the two Gaussians that falls inside the stimulus range (6.25° radius).

$$SI = \frac{\beta_2 * (\sigma_2)^2}{\beta_1 * (\sigma_1)^2} \tag{10}$$

Results

The DoG pRF model captures systematic events of the fMRI time-series.

The pRF models give a prediction of the measured fMRI time-series for every voxel. Figure 3A shows the prediction of the time-series for the OG-model for an example voxel in V1. Comparing the time-series with the prediction of the model shows that the OG-model fails to explain many parts of the time-series that fall below the baseline activity (< 0% BOLD signal), these are indicated by the gray arrows. Aside from the parts of the prediction that show the negative undershoot of the HRF, the OG-model is unable to predict responses below the baseline. The prediction for the identical fMRI time-series made by the pRF of the DoG-model is shown in figure 3B. Where the OG-model misses the ability to predict the negative fMRI signal, the DoG-model shows a clear improvement of the fit. This improvement is particularly evident at the negative parts of the time-series.

To quantify these observations in the fits of the OG-model and the DoG-model, we computed the difference of the measured fMRI time-series and the predicted time-series (residuals). Figure 3C shows the distribution of these residuals for both models of the pRF from one subject's V1 (in percent BOLD signal change). The distribution height is normalized according to the peak of the DoG-model. A residual value of zero indicates that the prediction matched the measured fMRI signal perfectly. The arrows indicate where the means of both curves are. These are -0.23/-0.20% and -0.07/-0.05% for the mean/median of the OG and DoG-model respectively. The DoG-model's residuals are distributed tighter and with higher amplitude around zero than the OG-model, indicating a better fit to the time-series, whereas the residuals of the OG-model are balanced around a negative difference and with a slight skew.

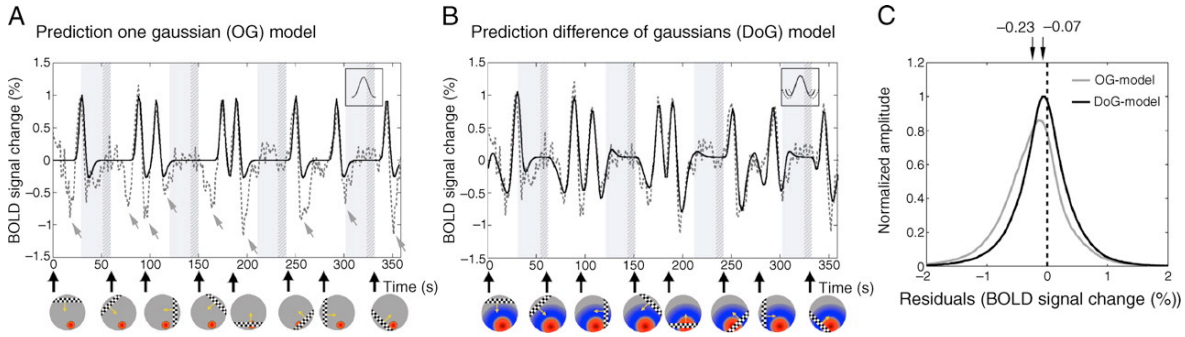


Figure 3 Example of the model fits for a cortical location (voxel) in V1 for both models of the pRF. The gray dotted lines indicate the fMRI time-series; the black solid lines show the fits of the models. Gray regions indicate mean-luminance presentations, whereas the dark gray regions indicate parts of the time-series used to estimate the baseline (0). The circular patches below indicate the visual field (gray), the estimated pRF for the voxel (the red part for the positive part of the pRF and the blue part for the negative part of the pRF) and the stimulus at a given time. A: The model fits of the OG-model of the pRF. The OG-model shows an accurate fitting for the positive parts of the time-series. The parts where the OG-model misses the ability to follow the time-series are particularly evident for the negative parts of the time-series. The gray arrows indicate these parts. B: The model fits of the DoG-model of the pRF. A clear improvement of the fit of the DoG-model can be seen at the negative parts of the fMRI signal. In parts where the OG-model misses the ability to give a correct prediction, the DoG-model that incorporated a suppressive surround is able to give an accurate prediction of the time-series. The models explain 49% and 72% of the variance, respectively. C: From the same subject this shows the distribution of the residuals from all the voxels in V1. To calculate the residuals, the predicted time-series is subtracted from the measured time-series (for all time-points). The gray line shows the residuals for the predicted time-series of the OG-model; the black line shows the residuals of the DoG-model. The arrows indicate the means of both curves. The mean of the residuals of the OG-model is -0.23%, for the DoG-model this value is -0.07%. Where the OG-model shows a negative shift in its residuals, the DoG-model shows a more balanced distribution of the residuals around zero.

The DoG-model explained more of the fMRI signal variance in V1 and decreasing in later visual areas

To quantify the goodness-of-fit of the pRF models, we compared the variance explained of the two models. The difference in variance explained of the DoG-model and the OG-model is depicted on the cortical surface in figure 4. For all subjects, only the values were plotted that have a minimal variance explained of 40% in the one Gaussian model. This figure illustrates that for all subjects the largest gain in variance explained for the DoG-model is found in V1. This difference in variance explained between the two models decreases in later visual areas.

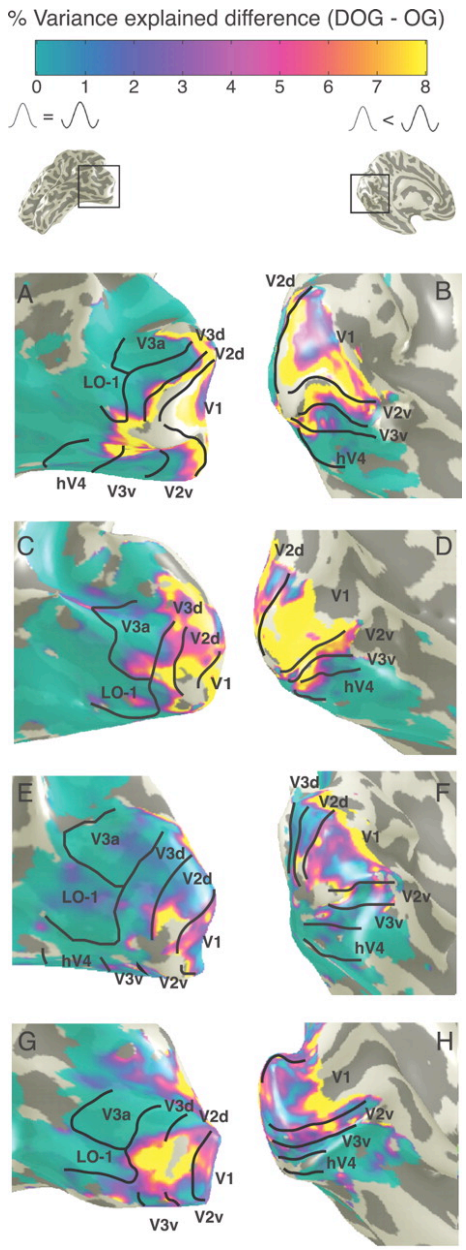


Figure 4 The difference in variance explained of the DoG-model and the OG-model depicted on inflated cortical surfaces. We show the left occipital lobe of all four subjects from a lateral (left panels) and medial (right panels) perspective. The largest difference in variance explained is found in visual area V1. This difference seems to decrease in later visual areas. Similar results were obtained in the right hemisphere (not shown). Only locations were the OG-model explained more than 40% of the variance in the time-series is shown.

To quantify these visual field map differences, the mean variance explained in the different visual areas is calculated for both models. Figure 5 shows the normalized variance explained of the different identified visual areas for both models. The averaged data is from all subjects for recording sites where both pRF models explained more than 30% of the variance in the fMRI time-series between 1.5-6 deg. eccentricity. We corrected for overall individual differences and then normalized the variance explained relative to the OG-model. Similar to Figure 4 this figure shows an increase of variance explained for the DoG-model. For V1 a relative improvement of ~14% is found for the DoG-model compared to the OG-model (absolute improvement is ~7%). This improvement decreases in V2 and V3 to ~12, and even more (~1%) in the other later visual areas V3a, hV4 and LO-1.

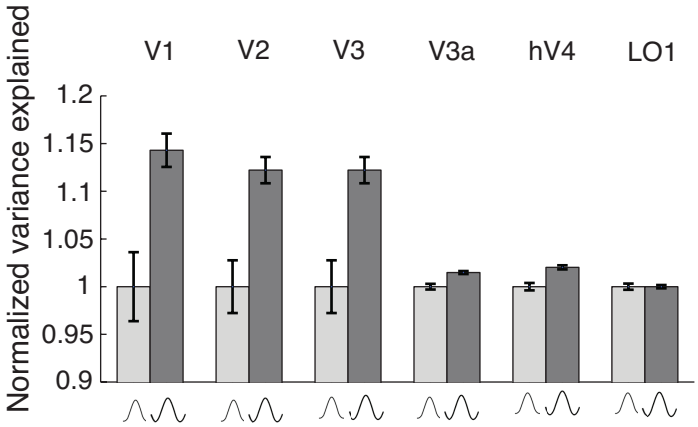


Figure 5 The variance explained for both the models of the pRF normalized to the OG-model. In V1 the increase of variance explained for the DoG-model is found to be ~14 % with respect to the OG-model. This improvement decreases in later visual areas. In V2 and V3, the improvement dropped to ~12%. The difference becomes even less in later visual areas V3a, hV4 and LO-1 (~1%). The averaged data is from all subjects, and the error bars reflect 1 standard error of the mean.

pRF size increases with eccentricity and in V1-V3

For both the OG-model as well as the DoG-model, estimates of the positive (center) pRF size are obtained. Both models estimate the standard deviation (σ_1) of a Gaussian to represent the positive part of the pRF. The DoG-model subtracts a second Gaussian from the positive one to obtain the total pRF. This subtraction leads to a change in the effective positive pRF size. For this reason the *full-width-half-max* (FWHM) was used to compare the effective positive pRF size. Figure 6 shows the relationship between eccentricity and FWHM in V1-V3. The averaged data is from the voxels of all subjects that have a variance explained > 30% and with eccentricity values between 1.5-6 deg. The lines are fit by a linear regression analysis. Both the linear regression and the averaging procedure are weighted by the variance explained of the individual voxels. The error bars represent one standard error of the mean. For both models of the pRF an increase of FWHM with eccentricity is shown as well as an increase of FWHM from V1-V3.

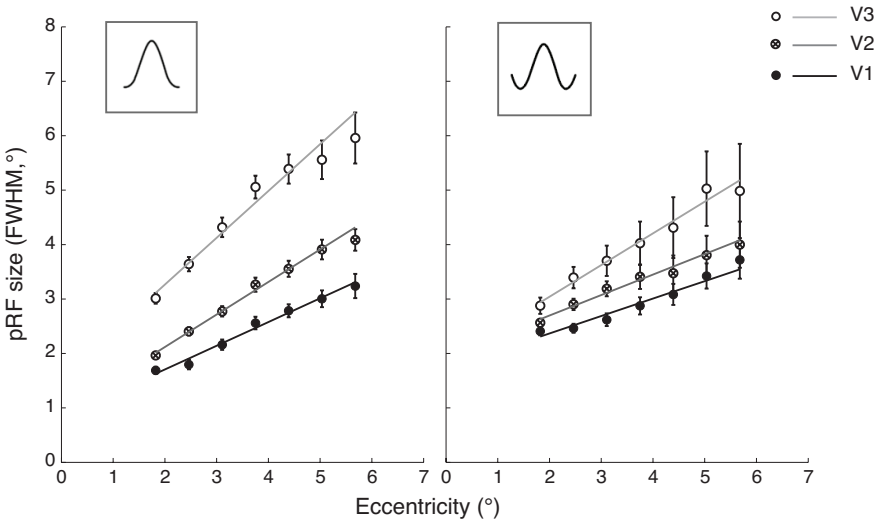


Figure 6 The relationship between eccentricity and pRF size in visual field maps V1-V3 for both models of the pRF. The averaged data is from all subjects, and the error bars reflect the standard error of the mean. The lines are fit from the data by a linear regression analysis. The pRF size increases with eccentricity and from V1-V3. These pRF sizes are similar between the two pRF models.

Measured sizes of the surround and suppression index

Besides producing measurements for the positive pRF size, the DoG-model gives estimates of the suppressive surround size as well. As a measure for the size of the surround we took the distance between the points where the pRF reaches its minimum amplitude. The relationship between eccentricity and the size of the negative surround is shown in Fig. 7A. We see an increase of the size of the surround with eccentricity and an increase of the surround from visual areas V1 to V3.

Fig. 7B shows the relationship between eccentricity and the suppression index of the volumes for visual field areas V1 to V3. The suppression index is calculated by dividing the volumes of the two estimated Gaussians that make the pRF (Formula 10). Since the pRFs can have a size that is beyond the size of the stimulus range, we only take into account the total volume of the two Gaussians that falls inside the stimulus range. For the suppression index we see a decrease over visual areas V1-V3. In visual areas V1 and V2 we see a decrease of the suppression index with eccentricity, where in V3 there is an increase of the suppression index with eccentricity. For both the surround size and the suppression index the averaged data is for the voxels that have a variance explained > 30% between 1.5-6 deg. eccentricity. The lines are fit from the data by a linear regression analysis; both the linear regression and the averaging procedure are weighted by the variance explained of the individual voxels. The error bars represent one standard error of the mean. For calculating the average surround size we only take the values of the voxels that show surround suppression. The voxels that do not show surround suppression have a suppression index of 0.

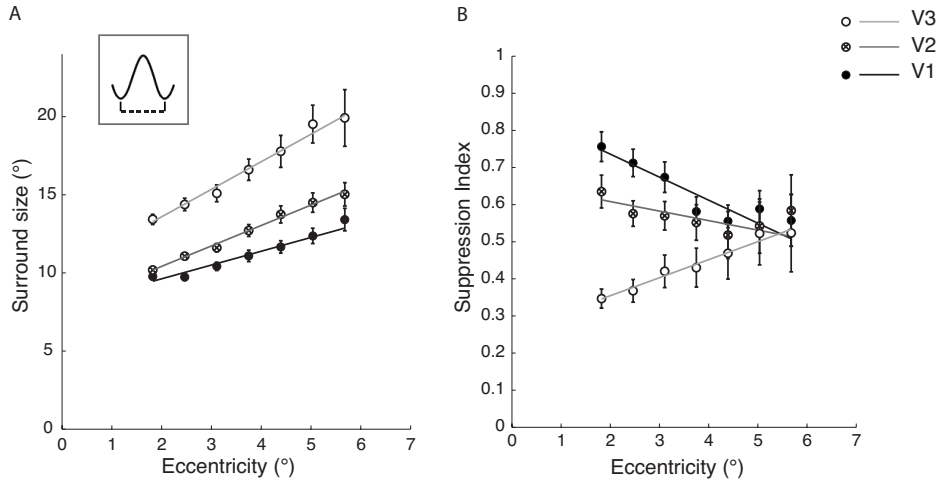


Figure 7 The relationship between eccentricity and pRF surround in visual field maps V1-V3 for both models of the pRF. *A:* The relationship between eccentricity and the size of the surround in visual field maps V1-V3. As a measure for the size of the surround we take the distance between the points of the pRF where it reaches its minimum amplitude. The averaged data is from all subjects, and the error bars represent 1 standard error of the mean. We see an increase of the size of the surround with eccentricity and an increase of surround-size over visual areas V1-V3. B: The relationship between eccentricity and the suppression index. The suppression index is calculated by dividing the volumes within the stimulus range of the two estimated Gaussians that make the pRF (Formula 10). The averaged data is from all subjects, and the error bars represent 1 standard error of the mean. We see a decrease of the suppression index over visual areas V1-V3. In visual areas V1 and V2 we see a decrease of the suppression index with eccentricity, where in V3 the suppression index increases with eccentricity.

Cross-validation

Where the OG-model represents the pRF by a single Gaussian, the pRF of the DoG-model is made up of the difference of two Gaussians. Adding this extra Gaussian leads to an increase of two extra parameters in the model; σ_2 and β_2. The addition of these extra parameters leads to a model with more degrees of freedom, which may result in an increase of the variance explained. Because the increase of variance explained is mostly

seen in V1-V3 but not in higher visual areas, we do not expect that the gain in variance explained is solely caused by the extra parameters. If it would be just the addition of the extra parameters causing the gain in variance explained, one would expect to see this gain throughout the visual cortex.

Nevertheless, we show that the gain in variance explained is a robust finding that is not only caused by the addition of extra parameters to the model. We do this by a cross-validation; in the training stage we fitted the model on a subset of the data and in the validation stage we evaluate these model parameters on the complementary part of the data. For this we split the data into two independent subsets. The time-series of the even scans were averaged for the first subset, and the time-series of the odd scans were averaged for the other subset of the data. Models were fitted on both subsets of the data. The parameters of these models are evaluated on their complementary dataset; the models that were fitted on the averaged time-series of the even scans are evaluated to the averaged time-series of the odd scans and vice versa. Figure 8 shows the variance explained for the evaluated predictions of both the OG-model and the DoG-model to its complementary dataset. The variance explained of the validation stage is comparable to the variance explained of the training stage. The averaged data is from the voxels that have a variance explained > 30% in both models and eccentricity values between 1.5-6 deg. The data is averaged over the subjects after normalization for individual differences. The error bars represent the 95% confidence interval. This result indicates that the DoG-model captures time-series characteristics that are systematically missed by the OG-model.

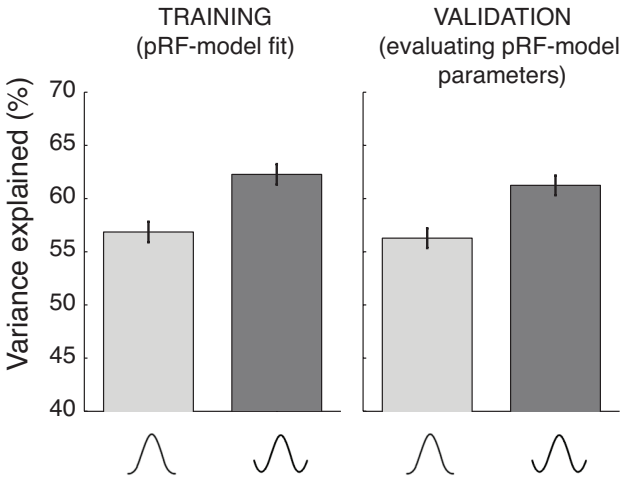


Figure 8 The variance explained of the model fits in the two stages of the cross-validation procedure: training and validation stage. The data was split into two different subsets. The data of the training stage gives the variance explained of the models on the data that it was fitted to, where for the validation stage we evaluated these model parameters on a different subset of the data. The error bars represent the 95% confidence interval. This result indicates that the improvement of the DoG-model fit captures systematic time-series characteristics missed by the OG-model, and that this improvement is not due to the increased number of model parameters.

Discussion

We introduce a computational model using fMRI that captures center-surround configurations in aggregate receptive fields of the underlying neuronal population. This model extends the conventional pRF model (Dumoulin & Wandell, 2008), where only the positively responding part of the pRF is estimated using one circular symmetric Gaussian (OG). The extension of the pRF model with a center-surround configuration provides a more biologically plausible description of the pRF. Whereas the conventional OG-model fails to explain parts of the time-series where the signal drops below the baseline, the DoG-model shows an improvement of the fits. The DoG-model shows a higher variance explained compared to the OG-model. This improvement was predominantly present in

V1/2/3 and decreased in later visual areas. The method provides new opportunities to measure center-surround configurations in the human visual cortex using fMRI and to test assumptions that are made on surround suppression.

Negative BOLD

In the cortex, many investigators have reported *negative BOLD responses* (NBRs) (Harel, Lee, Nagaoka, Kim, & Kim, 2002; Shmuel et al., 2006; Shmuel et al., 2002; A. T. Smith et al., 2004; Tootell, Mendola, Hadjikhani, Liu, & Dale, 1998; Wade & Rowland, 2010). Visual stimuli can elicit NBRs, typically adjacent to positively responding regions. Studies combining electrophysiology with fMRI in early visual cortex couple this NBR to a decrease in neural activation (Shmuel et al., 2006; Shmuel et al., 2002). Furthermore, the NBR shows a similar onset time and time course as compared to the positive BOLD response. Consequently, we believe that the negative signals are neuronal in origin. We extend these observations by showing that the negative parts of the measured time-series can be explained by a center-surround configuration of the pRF.

Center-surround configurations

Suppression – a reduction in response arising from the introduction of a surround to a visual target – has been found to decrease fMRI signals when using narrowband stimuli; (Beck & Kastner, 2007; McDonald, Seymour, Schira, Spehar, & Clifford, 2009; Wade & Rowland, 2010; Williams et al., 2003; Zenger-Landolt & Heeger, 2003), broadband stimuli patches (Kastner et al., 2001) and contextual modulations (Dumoulin & Hess, 2006; Harrison et al., 2007; Murray et al., 2002). Alternatively, fMRI responses elicited by viewing differently sized stimuli may also reveal surround suppression (Kastner et al., 2001; Nurminen et al., 2009; Press et al., 2001). As the stimulus size increases, the fMRI responses fail to increase linearly, which is interpreted as evidence for increasing suppressive contributions. These suppressive interactions may be local or long-range, also referred to as overlay and surround-suppression (Petrov, Carandini, & McKee, 2005; Petrov & McKee, 2006). Although our stimulus was not specifically designed to reveal suppressive interactions, we do reveal similar long-range suppressive contributions to the

pRF consistent with a center-surround configuration. Therefore we propose that our DoG-model captures the suppressive surround of the underlying neural population.

We found evidence for center-surround pRF configurations in V1 to V3, but not in later visual areas. Why did we not find a similar configuration in later visual areas? There are several possible explanations. First, the neuronal populations in later visual areas do not show surround suppression, or may show less surround suppression. Yet, many other studies suggest surround suppression beyond V3 (for review see (Allman et al., 1985)). To a certain extent our inability to reveal center-surround configurations beyond V3 may be a consequence of the exact stimulus layout, however, we do not believe this is the dominant reason. Second, the stimulus size may be too small. Given the increasing pRF size in later visual areas in combination with the maximum size of the stimulus (6.25° radius), this may not be enough to reconstruct the surround in later visual areas. Last, we may not be able to reconstruct the population center-surround configuration at the resolution of fMRI in later visual areas. Simulations (see appendix) indicate that increasing the position scatter of individual neuronal receptive fields leads to decreased surround effects at the population level. Position scatter is supposed to increase with increasing receptive fields though typically proportional to the RF size (Albright & Desimone, 1987; Dow, Snyder, Vautin, & Bauer, 1981; Fiorani, Gattass, Rosa, & Sousa, 1989; Gattass & Gross, 1981; Hetherington & Swindale, 1999; Hubel & Wiesel, 1974). However, if neuronal position scatter is proportional to RF size, this effect would not be seen. Therefore, increasing position scatter at later visual areas may cause the neuronal center-surround configuration to be lost at the resolution of fMRI. This is a limitation of our population receptive field approach, and other techniques may still reveal effects of the surround in later visual areas.

For the suppression index we see a decrease over visual areas V1-V3. Where we see a decrease of the suppression index with eccentricity in visual areas V1 and V2, in V3 we see an increase of the suppression index with eccentricity. When calculating the suppression index, we only take into account the total volume of the two Gaussians that fall inside the stimulus range. This might account for the increase of the suppression index over eccentricity in V3. When taking the total volumes of the two Gaussians into account

without considering the stimulus range, we still see a decrease of the suppression index over visual areas V1-V3 and we see a decrease over eccentricity for all visual areas V1-V3. On the other hand, taking the entire volume of the Gaussians includes the long tails of the Gaussian far beyond the stimulus range leading to unlikely high suppression indexes. Thus, as the surround sizes increase relative to the stimulus size, the suppression index becomes unstable.

Center-surround size estimates

The pRF is defined as the region of the visual field that elicits an fMRI response (Dumoulin & Wandell, 2008). The conventional pRF models only considered positive fMRI responses. The DoG-model estimates both the regions of visual space that elicits a positive (center) and negative (surround) response. The conventional OG-model and the DoG-model yield similar estimates for the center pRF sizes. Since the pRF size estimates of the OG-model are comparable to electrophysiological measurements, this observation remains valid for the DoG-model (Dumoulin & Wandell, 2008).

In addition, the DoG-model also estimates the size of the visual field that elicits negative fMRI responses. Our estimates of the DoG-model give a mean surround size of ~13° in V1. The mean surround size measured using electrophysiological approaches is ~5° (Angelucci, Levitt, Walton, Hupe, Bullier, & Lund, 2002; Levitt & Lund, 2002). Psychophysical methods combining fMRI measured a mean surround size of ~15° (Nurminen et al., 2009). Our measured sizes of the surround are most similar to the other fMRI study, whereas the electrophysiological studies show smaller surround sizes. This may be explained by a difference in the sampling sizes between fMRI and electrophysiology. FMRI measures from a larger neuronal population than electrophysiology. This leads to more variance in the visual field position, which in turn leads to an increase of the pRF size and surround estimates (see appendix).

Technical considerations

We found an average improvement of ~7% in variance explained (~14% relative) of the DoG compared to the conventional OG-model in V1. We believe that this improvement captures biologically relevant signal modulations, and that it is not simply due to the increased degrees of freedom. First, the DoG-model shows most improvement in V1, with a decreasing difference found in higher visual areas. If the increase in variance explained were due to more degrees of freedom of the DoG-model, one would expect to see this gain over the whole cortex. Second, testing of the DoG-model on a different subset of the data than the model was trained on, gives similar values for the variance explained. This suggests that the increase of variance explained for the DoG-model is not solely caused by the addition of the extra parameters to the pRF model, which could lead to over-fitting of noise in the fMRI signal, but instead has a biological origin. Third, the estimates for both the size of the positively responding region as well as the negatively responding region of the pRF are comparable to previous studies. Last, when comparing the goodness-of-fit of the two models, not only the difference in variance explained is important. Inspection of the predictions of the time-series for both models gives a more detailed illustration of the performance. The DoG-model explains specific parts of the time-series, which the OG-model systematically leaves unexplained. The parts of the time-series where the stimulus overlaps the surround of the pRF are the parts where the improvement for the DoG-model is seen. This might explain a gain of 'only' ~7% in variance explained for V1, as during a large part of the stimulus presentation preferentially the positive center or neither center nor surround get stimulated. In an experimental design where the surround of the pRF is stimulated to a greater extent, a larger difference in variance explained is expected between the two models of the pRF. These arguments suggest that the improvement in variance explained is not solely caused by more degrees-of-freedom of the DoG-model compared to the OG-model. The DoG-model gives a representation of the pRF that is biologically more plausible than the pRF of the OG-model.

Conclusion

The DoG-model makes it possible to measure center-surround configurations in the visual cortex using fMRI. This extended model of the pRF uses a biologically more plausible way to represent the pRF than the OG-model by taking surround suppression into account. During a scanning session of ~1 hour and using standard mapping stimuli we are able to estimate the parameters of the DoG-model. Several clinical conditions may alter pRF properties, and in particular center-surround properties (e.g. (Butler, Silverstein, & Dakin, 2008; Dakin & Frith, 2005; Marin, 2012; Yoon et al., 2009)). This method provides us with a direct measure on the properties of the pRFs throughout the visual system, which could be useful for studying both healthy and clinical populations.

Chapter 4

Image identification from brain activity using the population receptive field model

Wietske Zuiderbaan, Ben M. Harvey, Serge O. Dumoulin (under review).

Acknowledgements of author contributions: WZ and SOD designed the experiment. WZ and BMH collected the data, WZ performed data analysis under supervision of SOD. WZ wrote the manuscript and the other authors provided critical comments.

Abstract

A goal of computational models is not only to explain experimental data but also to make new predictions. A current focus of computational neuroimaging is to predict features of the presented stimulus from measured brain signals. These computational neuroimaging approaches may be agnostic about the underlying neural processes or may be biologically inspired. Here, we use the biologically inspired population receptive field (pRF) approach to identify presented images from fMRI recordings of the visual cortex, using an explicit model of the underlying neural response selectivity. The advantage of the pRF-model is its simplicity: it is defined by a handful of parameters, which can be estimated from fMRI data that was collected within half an hour. Using 7T MRI, we measured responses elicited by different visual stimuli: (i) conventional pRF mapping stimuli, (ii) semi-random synthetic images and (iii) natural images. The pRF mapping stimuli were used to estimate the pRF-properties of each cortical location in early visual cortex. Next, we used these pRFs to identify which synthetic or natural images was presented to the subject from the fMRI responses. We show that image identification using V1 responses is far above chance, both for the synthetic and natural images. Thus, we can identify visual images, including natural images, using the most fundamental low-parameter pRF model estimated from conventional pRF mapping stimuli. This allows broader application of image identification.

Introduction

Determining a human's mental state from their measured brain activity is a great challenge of neuroscience. Different decoding techniques have been used to determine what the subject was seeing, hearing, remembering or dreaming by analyzing fMRI activation patterns (for review see: (Tong & Pratte, 2012). Most of these studies use machine-learning techniques, where a classifier is trained on a set of activation patterns elicited by known stimuli. The classifier then compares a new activation pattern to this trained set of activation patterns to predict the state of an unknown stimulus. For example, it is possible to train a classifier using brain activation patterns elicited by sets of house and face images (Haxby et al., 2001). That classifier can then distinguish whether a new activation pattern was elicited by a house or a face image. However, this approach can only make predictions about a state that the classifier was explicitly trained on: a classifier trained to distinguish mental states elicited by house and face stimuli cannot distinguish between those elicited by, for example, animal and tree stimuli.

A different approach to make predictions about visually presented stimuli uses biologically inspired encoding models. These models are based on the brain's representation of stimuli (Dumoulin & Wandell, 2008; Kay, Naselaris, Prenger, & Gallant, 2008; Kriegeskorte, 2015; Mitchell et al., 2008; Naselaris, Prenger, Kay, Oliver, & Gallant, 2009; Thirion et al., 2006), and reveal not only what stimulus state is represented in the brain but also how the stimulus is represented. Kay and colleagues (2008) use a Gabor Wavelet Pyramid (GWP) as a biologically inspired encoding model for the early visual system. The model uses a training set of known natural images to determine the properties on the GWP. They identify presented images outside this training set using brain activation patterns measured with fMRI, and the properties of the GWP. The model predicts brain activity based on the inherent features of the stimulus. Since the model simulates the computations of the brain, it can make predictions for new stimuli. It can therefore identify new images that the model was not explicitly trained on. The downside of the GWP is that it uses many free parameters to capture the neural responses (2370), and requires long fMRI scan times (five hours per subject) for training.

Here we ask whether image identification is also possible using a basic biologically inspired encoding model that is already used in many conventional vision studies, i.e. the pRF model (Dumoulin & Wandell, 2008). The rationale of this study is two-fold. First, the study is a validation of the pRF-model on natural images. Using the pRF-model we can make a prediction of the measured brain activity to any image based on the contrast information of that image. Here we demonstrate that the representation of the image by the pRF-model corresponds to the actual representation based upon the measured brain activity in response to natural images. Second, this study allows a deeper insight into visual processing mechanisms operating on natural images. For example, we can evaluate if there are image features beyond contrast that are not included in the pRF-model but affect responses of visual cortex. So, where on the one hand this study shows us how contrast information is represented in V1, it also provides us information about the deviation of the representation in V1 from this contrast information.

The model was trained on responses to standard visual field mapping stimuli, which can be measured within half an hour of scanning and are already used in many conventional vision studies. We used these responses to estimate the pRF-properties of each cortical location (Dumoulin & Wandell, 2008). Next, we measured fMRI responses elicited by both synthetic stimuli and natural images. Synthetic images are widely used because they are more easily controlled. However, responses to synthetic images may not extrapolate to natural images (Carandini et al., 2005; David, Vinje, & Gallant, 2004; Kayser, Kording, & Konig, 2004; Li, VanRullen, Koch, & Perona, 2002). Therefore, by using both synthetic and natural images, we aim to bridge the gap between experimental settings and real-life situations. We identified the presented image based on the similarity of the measured fMRI signals to the predictions of the pRF-model. Using the voxels of visual field map V1, image identification from large image datasets (1000 synthetic stimuli or 200 natural images) is far above chance, both for the synthetic and natural images. As such, we show that it is possible to identify natural images using a pRF-model with minimal parameters.

Methods

Subjects

Two subjects (one female; ages 28-38 years) participated in this study. All subjects had normal or corrected-to-normal visual acuity. All studies were performed with the informed written consent of the subjects and were approved by the Human Ethics Committee of University Medical Center Utrecht.

Stimulus presentation

The visual stimuli were generated in Matlab using the PsychToolbox (Brainard, 1997; Pelli, 1997) on a Macintosh Macbook Pro. The stimuli were back-projected on a display inside the MRI bore. The subject viewed the display through mirrors inside the scanner. The size of the display was 15.0x7.9 cm with a resolution of 1024x538 pixels. The total distance from the subject's eyes to the display was 41 cm. The stimuli were constrained to a circular area with the size of the vertical dimension of the screen. The area outside this circle was remained at a constant mean luminance. This gave the stimulus a radius of 5.5° visual angle from the subject's point of view.

Stimuli used to estimate pRF properties

The pRF properties were estimated using conventional contrast-defined moving-bar apertures (Figure 1) (Dumoulin & Wandell, 2008). The width of the bar subtended 1/4th of the stimulus radius (1.375°). Four bar orientations (0°, 45°, 90° and 135°) and two different step directions for each bar were used, giving a total of 8 different bar configurations within a given scan. The bar stepped across the stimulus aperture in 20 steps (each 0.55° and 1.5 seconds) with each pass taking 30 seconds. A period of 30 seconds mean-luminance (0% contrast) was presented after every horizontally or vertically oriented pass. In total there were 4 blocks of mean-luminance during each scan, presented at evenly spaced intervals.

We used two different sets carrier-images within the moving bar apertures. The bar apertures used to train the pRF-model to identify the synthetic images were filled with a binarized bandpass filtered noise pattern (fundamental frequency is 1.5 cycles/deg). This pattern was presented with three different alternating high-contrast patterns, to obtain a full high-contrast response that is not based upon one specific high-contrast pattern. The bar apertures used to train the pRF-model to identify the natural images contain natural image content (Figure 1). Each bar aperture was presented for one TR (1.5s). For each bar aperture three different natural image carriers were used. The natural image content was replaced every 500ms. The image was shown for 300ms with a 200ms mean-luminance gap.

The natural image content for the bars came from images of the 'Berkeley Segmentation Dataset and Benchmark' database (Martin, Fowlkes, Tal, & Malik, 2001). The synthetic and natural images used in these bars did not include those used in the image identification test sets.

The subjects performed a fixation dot task to make sure they fixated at the center of the display. A small fixation dot (0.11° radius) was presented in the middle of the stimulus. The fixation dot changed its color from red to green at random time intervals and subjects were instructed to respond at this change of color using an air pressure button (average performance of 93.3%).

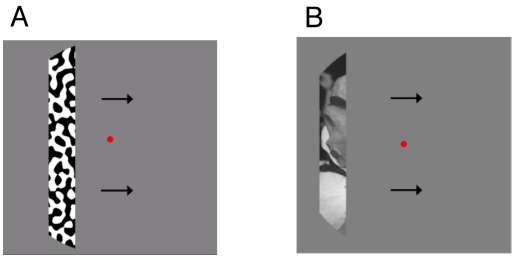


Figure 1 Examples of the pRF mapping stimuli used to estimate the pRF properties with a binarized bandpass filtered noise pattern (A) and natural image content (B).

Stimuli used for image identification

For the image identification process we used two different sets of images. We used both synthetic images and natural images. The synthetic images were scanned for one subject only, as a proof of concept for the image identification process. The images were presented in a block design. Each image was presented during a 9-second block. Within this block the same image was shown 18 times for a duration of 300ms followed by 200ms mean-luminance. The block where the image was presented was followed by a 12 second mean-luminance presentation. The synthetic images were presented with 3 alternating different high-contrast patterns, to obtain a full high-contrast response that is not based upon one specific high-contrast pattern. A fixation dot was presented at the center of the stimulus in both stimulus sets. The same fixation dot task as in the scans of the mapping stimuli was used (average performance of 91.5%).

Synthetic images

The synthetic images were semi-randomly generated patterns consisting of a hexagonal grid partially filled with binarized bandpass-filtered noise. The grid contains 60 small hexagons (cells), of which a random selection were filled with binarized bandpass filtered noise, and the rest filled with mean-luminance gray. This random selection gives 2^60 (>> trillion) possible images. We used 3 sets of these images in different scanning runs, with each set containing 15 different synthetic images (45 in total) and one full-field binarized bandpass-filtered noise stimulus. Figure 2B-C shows an example of the synthetic images used in the experiment. We used 5 non-random patterns (Figure 2B) and 40 random patterns (Figure 2C), they were combined in the image identification process.

Natural images

The natural images for the image identification process came from the 'Berkeley Segmentation Dataset and Benchmark' database (Martin et al., 2001). The original resolution of the images was 321x481 pixels (both landscape and portrait). We selected a

squared part of 321x321 pixels from the images and upsampled this to a resolution of 538x538 pixels, which corresponds to a stimulus of 11x11° diameter of visual angle. The images were masked by a circle with a raised cosine faded edge (width of 0.9°), and the areas outside this circle were set to mean luminance. The images were gamma-linearized and the mean contrast was set to 50%. We used 3 image sets in different scanning runs, each containing 15 different natural images (45 in total) and one full-field binarized bandpass-filtered noise stimulus. Figure 2D shows an example of the natural images.

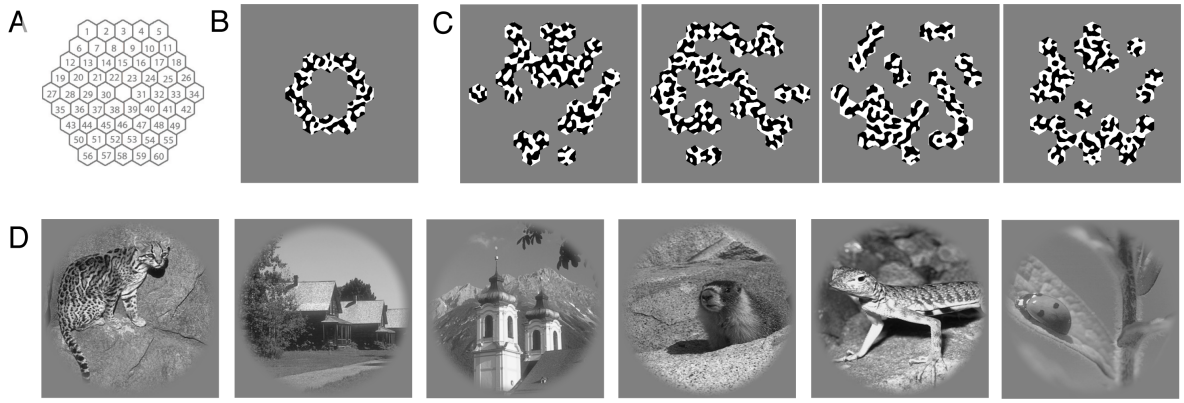


Figure 2 Image identification stimuli. (A) The synthetic stimuli were made using a grid of 60 hexagons (cells) that could be filled with either binarized bandpass-filtered noise or mean-luminance gray. Examples of the stimuli used in the experiment; (B) a non-random synthetic stimulus, (C) random synthetic stimuli and (D) natural images.

Functional imaging and processing

The MRI data was acquired with a Philips 7T scanner using a 32-channel head-coil. We scanned the participants with a 2d-echo-planar-imaging sequence with 26 slices oriented perpendicular to the Calcarine sulcus with no gap. The following parameters were used; *repetition time* (TR) = 1500 ms, *echo time* (TE) = 25 ms and a flip angle of 80°. The functional

resolution was 2x2x2 mm and the *field of view* (FOV) was 190x190x52 mm.

We used foam padding to minimize head movement. The functional images were corrected for head movement between and within the scans (Nestares & Heeger, 2000). For computation of the head movement between scans, the first functional volumes for each scan were aligned. Within scan motion correction was then computed by aligning the frames of a scan to the first frame.

The duration of the pRF mapping scans was 372 seconds (248 time-frames), of which the first 12 seconds (8 time-frames) were discarded due to start-up magnetization transients. During the three sessions for the synthetic images we acquired 10 pRF mapping scans in total. For the natural image sessions we scanned 2 subjects and acquired during the three scanning sessions 6 or 7 pRF mapping scans in total per subject. To obtain a high signal-to-noise ratio, we averaged the repeated scans. The duration of the image identification scan (both synthetic and natural) was 432 seconds (288 time-frames). The first 12 seconds (8 time-frames) were discarded due to start-up magnetization transients.

During the three sessions for the synthetic images we acquired three scans for each of the three different image sets. During the three sessions for the natural images we acquired two scans each for the three natural image sets.

Anatomical imaging and processing

The T1-weighted MRI images were acquired in a separate session using an 8-channel SENSE head-coil. The following parameters were used: TR/TE/flip angle = 9.88/4.59/8. The scans were acquired at a resolution of 0.79x0.80x0.80 mm and were resampled to a resolution of $1mm^3$ isotropic. The functional MRI scans were aligned with the anatomical MRI using an automatic alignment technique (Nestares & Heeger, 2000). From the anatomical MRI, white matter was automatically segmented using the *FMRIB's Software Library* (FSL) (S. M. Smith et al., 2004). After the automatic segmentation it was hand-edited to minimize segmentation errors (Yushkevich et al., 2006). The gray matter was grown from the white matter to form a 4 mm layer surrounding the white matter. A smoothed 3D cortical surface can be rendered by reconstruction of the cortical surface at the border of the white and gray matter (Wandell, Chial, & Backus, 2000).

pRF model-based analysis

The first step for image identification is the estimation of the pRF-model from the measured fMRI signal that was elicited by the pRF mapping bar stimuli (Figure 3A). The model estimates a pRF for every cortical location using a method previously described (Dumoulin & Wandell, 2008). For technical and implementation details see (Dumoulin & Wandell, 2008) but in short, the method estimates the pRF by combining the measured fMRI time-series with the position time course of the visual stimulus. A prediction of the time-series is made by calculating the overlap of the pRF and the stimulus energy for each time frame convolved with the haemodynamic response function (HRF). We estimated the parameters of the HRF that best describes the data of the whole acquired fMRI volume (Harvey & Dumoulin, 2011). The optimal parameters of the pRF-model are chosen by minimizing the residual sum of squares between the predicted and the measured time-series. We used the simplest pRF-model which consists of a circular symmetric Gaussian. This model has four parameters: position (x_0,y_0), size (σ_1), and amplitude (β_1).

Using the pRF-method, we estimated position parameters x_0, and y_0 of the pRF per voxel. From these values, the polar angle ($\text{atan}(y_0 / x_0)$) and eccentricity ($\sqrt{(x_0^2 + y_0^2)}$) values can be calculated. We drew the borders of the visual areas on the basis of their location in the visual field (DeYoe et al., 1996; Engel, Glover, & Wandell, 1997; Sereno et al., 1995; Wandell, Dumoulin, & Brewer, 2007) by rendering these polar angle and eccentricity maps onto the inflated cortical surface (Wandell et al., 2000). We defined visual area V1, V2 and V3 as our *region of interest* (ROI).

Image identification

Analysis of signal responses elicited by viewing synthetic and natural images

We measured fMRI responses to 45 synthetic images and to 45 natural images. We first determined the distribution of the responses at each recording site (voxel) elicited by each of these images. To estimate the responses we used standard GLM-analysis (Friston et al., 1998; Friston et al., 1995). Briefly, we fitted a block-design to the stimulus presentation

convolved with the HRF. We summarized every voxel's response by its t-value, which reflects the goodness of fit between the predicted time series and the measured data for each cortical location for that image. For image identification, we only used voxels with positive t-values, pRF eccentricity values from 0.5-4.5° and pRF variance explained above 55%.

Prediction response profiles for synthetic and natural images using the pRF model

In the identification process, we had a large candidate set of images from which to choose the image that was presented. To do this, we compared the measured response profiles against the response profiles predicted by the pRF-model. We determined these predicted response profiles from the pRF-model as follows (Figure 3B).

We converted the synthetic images to binary images where cells filled with bandpass-filtered noise were set to a binary value of 1 and cells filled with mean-luminance gray were set to 0. The prediction response profile is calculated by the summed overlap of the stimulus with the pRF of each cortical location and is normalized by the total volume of the pRF:

$$voxelprediction\_to\_synthetic\_image = \frac{\sum_{i=1}^{N} w_i \cdot S_i}{\sum_{i=1}^{M} w_i} \quad (1)$$

Where N is the number of pixels in the spatial window of the pRF and M is the total number of pixels in the stimulus area. S is the binary stimulus, where S_i is the binary value of whether the pixel of the stimulus was on (1) or off (0). The pRF weighting function is defined by w_i:

$$w_i = \exp - \left(\frac{(x_i - x_c)^2 + (y_i - y_c)^2}{2(\sigma)^2} \right) \quad (2)$$

Where x_c and y_c define the location of the center of the pRF in the visual field, σ is the size

of the pRF and x_i and y_i define the location of the *i*th pixel.

For the natural image we followed a slightly different approach as the stimulus cannot be binarized in the same way as the synthetic images. We predicted each voxel's response to each candidate natural image by calculating the Root-Mean-Squared (RMS) contrast (Bex & Makous, 2002; Pelli, 1997) weighted by the corresponding pRF. RMS contrast is defined as the standard deviation of the luminance intensities relative to the mean. The RMS-contrast is weighted by the pRF-function to obtain the local contrast-energy value per pixel:

$$local\_contrast\_energy = \sqrt{\frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \frac{(L_i - L)^2}{L^2}} \quad (3)$$

Where N is the number of pixels in the spatial window of the pRF. L is the mean luminance from the pixels inside the spatial window, and L_i is the luminance of the *i*th pixel.

The predicted response profile for each candidate image consisted of the predicted response amplitudes of all voxels within a given visual area. These response amplitudes are not further convolved with the HRF, as this would yield a linear amplitude transformation that leaves relative response amplitudes within the overall response profile intact.

Correlating predicted with measured responses profiles

To identify the image that was shown to the subject we compared the measured response profiles elicited by the presented image (Im_i) to the predicted response profiles for each candidate image {Im_1, Im_2, Im_3,..., Im_n}. The presented image was identified by choosing the candidate image that gives the highest correlation (Pearson's *r*) between its predicted and the measured response profile (Figure 3C).

We calculated the percentage of correct image identifications within each set of 15 images. We also increased the set size of candidate images by including images that we never showed. We increased the set of candidate images to 1000 for synthetic images, and

to 200 for natural images. This makes correct image identification by chance less likely. We bootstrapped this analysis for each set size, by making 1000 different combinations of different candidate images. From resulting proportions of correct image identifications among these candidate image sets, we computed the mean image identification performance and the 95% confidence interval of this mean.

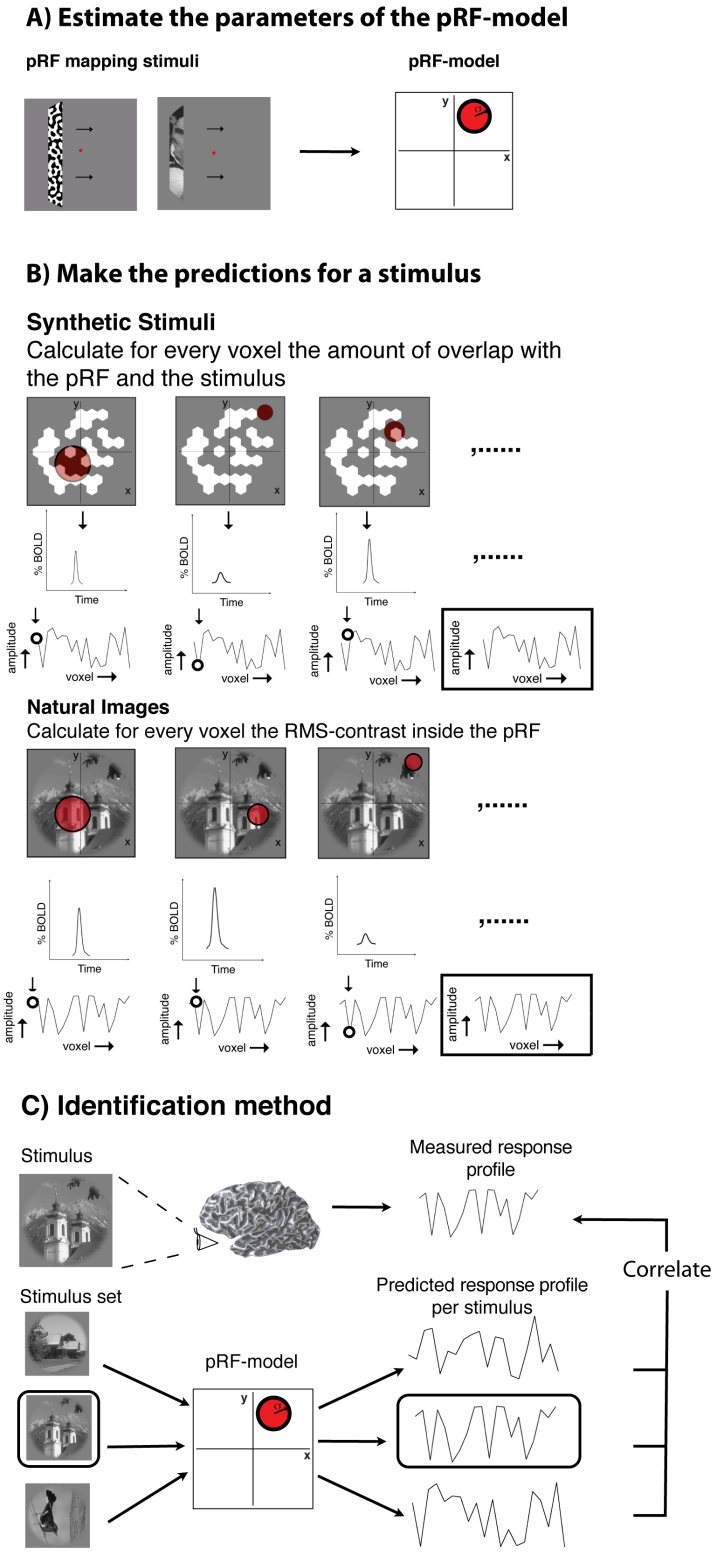

A) Estimate the parameters of the pRF-model
pRF mapping stimuli
pRF-model
y
x
B) Make the predictions for a stimulus
Synthetic Stimuli
Calculate for every voxel the amount of overlap with the pRF and the stimulus
% BOLD
Time
amplitude
voxel
Natural Images
Calculate for every voxel the RMS-contrast inside the pRF
C) Identification method
Stimulus
Measured response profile
Correlate
Stimulus set
Predicted response profile per stimulus
pRF-model

Figure 3 Schematic diagram of the image identification pipeline. A: First, we estimated the parameters of the pRF for every voxel based on the pRF bar stimuli. The pRF is modeled as a circular symmetric Gaussian function of which we use the parameters for position and size. B: Second, we predicted the response profiles for a large set of candidate images by either summing the overlap of the stimulus with each voxel's pRF (for the synthetic images), or by calculating the RMS-contrast inside each voxel's pRF (for the natural images). C: Finally, we predicted which candidate image elicited a measured response profile by finding which candidate image's predicted response profile was most strongly correlated to this measured response profile (i.e. which had the highest Pearson's *r*).

Results

Successful image identification for the synthetic images

To demonstrate that images could be identified using pRF-models, we first performed a proof of concept experiment using synthetic, high contrast images in a single subject. We identified the synthetic images using the pRF-model that summarized responses to bar apertures that revealed binarized bandpass filtered noise (Figure 1A). We used this model to predict the response profiles for a set of possible candidate images. We then generated a correlation matrix by correlating the predicted response profiles for every image to the measured response profiles. Figure 4 shows the correlation matrix for the voxels of primary visual cortex (V1) for one of the stimulus sets. The y-axis gives the correlation values of the stimulus set's measured response profiles to the predicted response profiles (on the x-axis). The black outlines show the candidate image whose predicted response profile was most highly correlated to the measured response profile, which we predicted was the presented image. The prediction accuracy for this subject was 93.3% (14 out of 15).

The correlation matrix shows the prediction accuracy using only the predicted response profiles for images within the presented set, giving the image identification performance a chance level of 1/15 images (chance level = 6.7%). Using the pRF-model,

we can make predicted response profiles for any image, including those that were never presented. This makes correct image identification less likely, and more accurately quantifies the likelihood of identifying the correct image by chance. Figure 5A shows the performance when we make predicted response profiles for up to 1000 different randomly generated synthetic stimuli. This reduces the chance level to 1/1000 (chance level = 0.1%). Nevertheless, the correct image was still identified correctly from 1000 candidate images for 89.0% of the presented images for visual area V1. The performance drops for later visual areas V2 (~56%) and V3 (~18%), but still remains high above chance. The thickness of the line includes the 95% confidence intervals, determined by bootstrapping the mean.

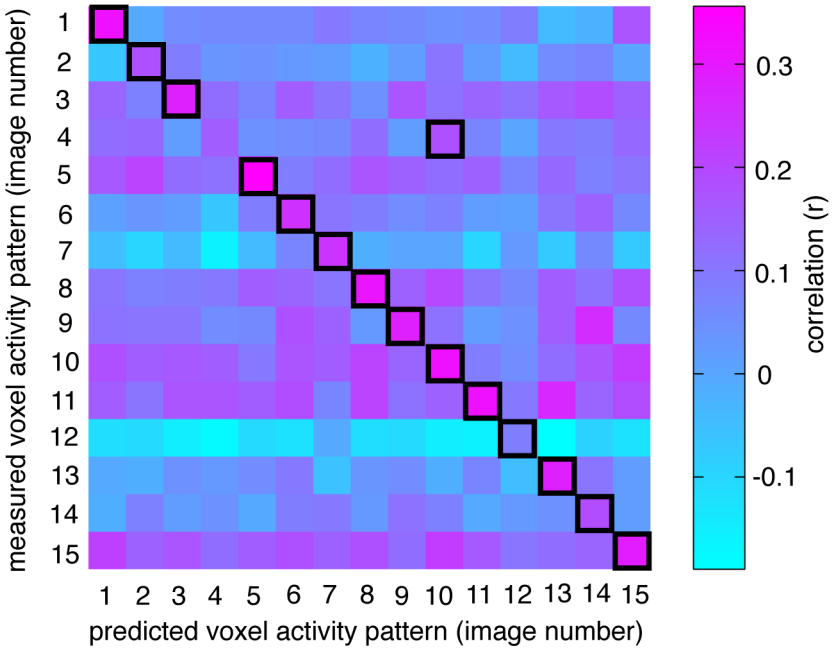


Figure 4 The correlation matrix shows the prediction accuracy using the voxels of V1 from the image identification process for an example set of the synthetic images. The colors represent the correlation (Pearson's *r*) of the measured response profiles from all the images with their predicted response profile (from the pRF-model). For this image set, 14 out of 15 images were identified correctly, giving a prediction accuracy of 93.3%.

Successful image identification for natural images

Next, we extended this approach to natural images. To predict response profiles for the natural images, we used the pRF-model that summarized responses to bar apertures containing natural image content (Figure 1B). The natural image content in these bars was independent from the natural images used for the image identification process. We scanned 45 different natural images for two subjects. For the 45 images the performance is 33% (15 out of 45 images) both for subject 1 and 2, the performance for a random selection of 45 images is ±38%. Both are very different from chance (2.2%). We choose to show the correlation with a larger set-size because idiosyncrasies of the chosen set size of 45 images are removed and results are more generalizable.

Figure 5B shows the identification performance for the natural images for both subjects (closed and dashed colored lines) as a function of candidate image set size, up to 200 candidate natural images. The thickness of the lines includes the 95% confidence intervals, determined by bootstrapping the mean. In each subject, image identification performance was far above chance, with approximately 29% correct image identification performance with 200 candidate images (chance level = 0.5%). The identification performance for visual areas V2 (~14%) and V3 (~12%) is lower compared to V1, but remains far above chance.

Besides the pRF-model that was estimated using bars with natural image content, we also used a pRF-model that was estimated on the standard pRF mapping stimulus consisting of contrast-defined bar apertures containing moving checkerboards (Figure 5C) (Dumoulin & Wandell, 2008). The latter stimulus is widely used to model pRF properties (Anderson et al., 2017; Baseler et al., 2011; Brewer & Barton, 2014; DeSimone, Viviano, & Schneider, 2015a, 2015b; Dumoulin & Wandell, 2008; He, Mo, Wang, & Fang, 2015; Hoffmann et al., 2012; Hummer et al., 2016; Kok, Bains, van Mourik, Norris, & de Lange, 2016; Papanikolaou et al., 2014a, 2014b; Schwarzkopf, Anderson, de Haas, White, & Rees, 2014a, 2014b; Thomas et al., 2015; Winawer, Horiguchi, Sayres, Amano, & Wandell, 2010; Zuiderbaan, Harvey, & Dumoulin, 2012). In addition, the pRF estimates were acquired on different days as the natural image data-set, as many researchers already have

such pRF models. In these circumstances, we still find that it is possible to identify natural images above chance level (~15%).

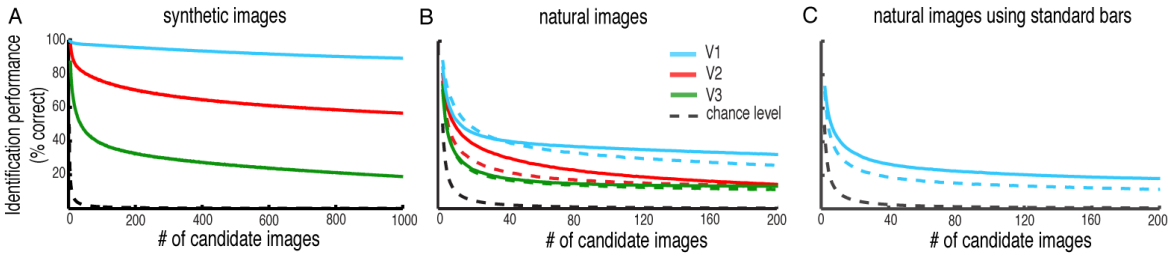


Figure 5 Image identification performance both for synthetic (A) and natural image stimuli (B, C). The pRFs were estimated in a separate scan where the bars were filled with synthetic (A, C) and natural (B) stimuli either in the same session (A,B) or in different sessions on different days (C). The blue, red and green lines show the performance for visual field maps V1, V2 and V3 respectively, for 2 subjects (closed and dashed colored lines). The thickness of the lines includes the 95% confidence intervals. For the synthetic image stimuli we increased the set size up to 1000 different images, for the natural images we increased the set size up to 200 different images. The black dashed line indicates the chance level. The high-contrast synthetic images were most accurately identified (A), and the natural images were also identified far above chance for all candidate image set sizes (B). Furthermore, the identification of the natural images is also possible with the pRF-model that was estimated using standard bar stimuli containing moving checkerboards, and estimated from a separate scanning session on a separate day (C).

Image identification accuracy depends on image content

We examined image identification accuracy for individual natural images. This reveals which images are most and least accurately identified. Figure 6 shows identification confidence for each image, based upon the mean difference of the correlation score of the presented image and all other candidate images.

$$confidence\_score_i = \frac{-\sum_{j=1}^{N} corr(i_m, j_p) - corr(i_m, i_p)}{N} \quad (4)$$

Where i is the image to calculate the score for, corr(i_m,j_p) is the correlation score (pearson's r) of the measured (m) response profile of the ith image and the predicted (p) response profile of the jth image. corr(i_m,i_p) is the correlation score of the measured and predicted response profiles of the presented image and N are the number of images in the set. The confidence score represents how hard it is to distinguish the presented image from other candidate images. A high difference in correlation scores will give a high confidence score. Here we see that for both subjects the same images are harder to identify than other images. We find a significant correlation (Pearson's r) of $r(43)=0.58$, $p<<0.001$.

Images could be harder to identify because they are more similar to other images or the visual representation deviates more from their contrast-energy content. To distinguish these possibilities, we checked the images for image similarity in contrast-energy using the predicted response profiles of the pRF-model. We found a significant effect for similarity in contrast between the images and how well we were able to predict the presented image. Thus images that were more similar in contrast-energy information as the other images in the candidate set of images to choose from were harder to identify. Nevertheless, after removing the relation with contrast (using GLM-analysis), we still found the effect across subjects that the same images were harder to identify than other images ($r = 0.53$ $p << 0.001$). This suggests that image performance for certain images is deteriorated because both (i) certain images are more in terms of contrast information and (ii) the measured responses of certain images hold information about image features that are not captured by the pRF-model.

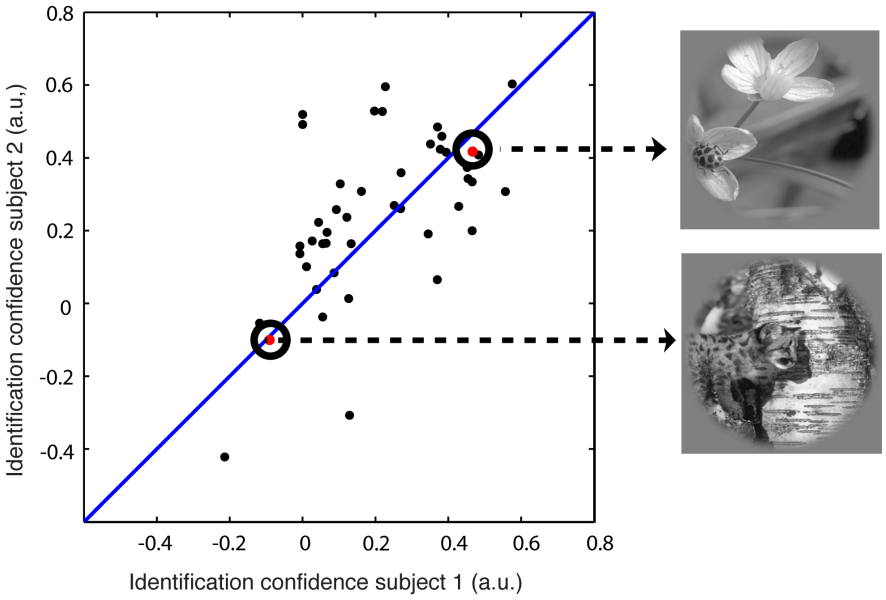


Figure 6 The identification confidence per natural image for our two subjects. Every dot represents an individual image. We see similar confidence of the individual natural images across the two subjects. Some images are identified less accurately using the pRF-model than others. This is explained by two factors: (i) certain images are more similar in terms of their contrast-energy content and (ii) responses to these images depend more on features that are not captured by the contrast-energy pRF-model predictions.

Discussion

We describe an effective method to identify presented natural images from the fMRI responses of V1 voxels using pRF-models. We used this method to identify both synthetic and natural images with accuracy far above chance. The synthetic images were easier to identify than the natural images, suggesting that image identification performance is affected by the spatial distribution of contrast within the identified images. The pRF model we employed had minimal parameters (3) and was estimated with very different stimuli. Even when using a pRF-model derived from a different scanning session and using the standard moving checkerboard stimuli, it is possible to identify natural images. This allows the image identification method to be more broadly applied.

We believe that this method is generally applicable despite our small sample size. First, previous papers have demonstrated image identification using similar approaches tested over small sample sizes. For example Kay et al. also use 2 subjects ((Kay et al., 2008; Thirion et al., 2006). Second, we have shown successful image identification in each subject with very high statistical confidence. Data from further subjects may show better or worse on image identification performance, but we believe this will reflect differences in data acquisition quality rather than reflecting any methodological limitation. Third, while conventional fMRI studies average data across large subject numbers, most collect less data from each subject. Our approach is to collect a large dataset for a small number of subjects and to analyze our data per subject in a voxel-wise modeling approach. The advantage of this approach is that detailed spatial information and individual variability is retained. This information is lost in the conventional fMRI studies by averaging of the data across subjects.

Image identification was possible for both the synthetic and natural images. In both cases, identification accuracy was far above the chance level. However, the identification accuracy was higher for synthetic than natural images. There are several possible reasons for these differences. The synthetic images were simple binary patterns of high contrast, designed to elicit a strong fMRI responses. Natural images have a broader spectrum of lower contrasts, and so elicit lower amplitude fMRI responses, with a lower signal to noise ratio (t-test on %BOLD signal change of the natural vs. synthetic images: $t(85379) = 26.73$, $p<<0.01$).

Furthermore, pRF-models as employed here only represents contrast energy. Early visual cortex is known to respond very strongly to differences in contrast (Boynton, Demb, Glover, & Heeger, 1999; De Valois & De Valois, 1988; Dumoulin, Dakin, & Hess, 2008; Mante & Carandini, 2005; Olman, Ugurbil, Schrater, & Kersten, 2004). The synthetic images are dominated by their contrast content, so responses to synthetic images are better captured by the pRF model than responses to natural images. While the synthetic images contained all orientations and spatial frequencies within any local area, natural images had highly variable spatial distributions of both orientation and spatial frequency. Orientation

content is known to affect the responses of each voxel (Freeman, Brouwer, Heeger, & Merriam, 2011; Haynes & Rees, 2005; Kamitani & Tong, 2005; Kay et al., 2008), and spatial frequency content is also likely to affect voxel responses (Henriksson, Nurminen, Hyvarinen, & Vanni, 2008; Kay et al., 2008; Olman et al., 2004; Singh, Smith, & Greenlee, 2000). Our pRF models do not account for these factors. Furthermore, responses to natural images are also likely to be affected by high-level image features and global image context (Petro, Vizioli, & Muckli, 2014), which are also not captured by the pRF-model.

Also, where image identification was possible using the voxels of visual areas V1, V2 and V3, we see that the accuracy drops for the later visual areas V2 and V3. The reason for this drop in accuracy can have several reasons. First, pRF sizes get bigger in higher visual areas (Dumoulin & Wandell, 2008; A. T. Smith, Singh, Williams, & Greenlee, 2001). This increase in receptive field sizes might decrease the resolution necessary for image identification. Second, the visual field maps get smaller for later visual areas (Dougherty et al., 2003), giving less measurement sites to perform the analysis. Third, the pRF-model only captures the contrast energy. As described above, other features of the images can be captured in the measured responses that are not represented by the responses predicted using the pRF-model. This effect can be bigger for later visual areas that are proposed to respond to more complex features (Felleman & Van Essen, 1991; Hubel, 1982).

Figure 6 shows the identification confidence of the individual natural images for both subjects for visual area V1. We see that some images are identified with less confidence than others and that these images are similar across subjects. We propose that this is due to two reasons. First, images that are more similar in terms of their content in contrast-energy are harder to distinguish. Second, responses to images that are harder to identify depend more on image content that is not captured by the pRF model, as discussed above. Differences in identification confidence can also indicate differences in the perception of images across subjects. This method can potentially be used to investigate these differences.

The pRF model we employ represents the most basic encoding model. Encoding models give information about how the information is represented in the voxels (Brouwer & Heeger, 2009; Dumoulin & Wandell, 2008; Kay et al., 2008; Kok & de Lange, 2014; Kriegeskorte, 2015; Mitchell et al., 2008; Naselaris et al., 2009; Thirion et al., 2006). These models use biologically inspired computations that make a translation between the stimulus and the response. The pRF model uses two parameters for visual field position (x and y) and a third for pRF size. PRF position and size are inherently necessary here, since we calculate the contrast within a specified area of the image to predict the response of each recording site. This area must have a size because a single image location (pixel) has no contrast. Since encoding models predict the activation pattern in response to a certain stimulus, these models can be inverted to decode and hence can be used for the purpose of image identification.

Alternatively, decoding methods define a categorical relationship between the activation patterns and the stimulus, and can be used to indicate whether a voxel contains information about a certain property of a (visually) presented stimulus (Cox & Savoy, 2003; Haxby et al., 2001; Haynes & Rees, 2005; Kamitani & Tong, 2005; Mitchell et al., 2004; Tong & Pratte, 2012). However, since they do not give a further description of how this information is represented in the voxels, these models are only able to make a prediction about the predefined categories that the model was explicitly trained on. It is not possible to make a prediction about any other property of the stimulus. This makes these models unsuitable for image identification, where every possible image should be able to be identified.

Last, both decoding and encoding models can reconstruct the image from the fMRI signal. The goal of image reconstruction is to reproduce the presented image. Reconstructions have been made for simple high contrast patterns (Miyawaki et al., 2008; Thirion et al., 2006), imagined (Thirion et al., 2006) and illusory contours (Kok & de Lange, 2014), dreams (Horikawa, Tamaki, Miyawaki, & Kamitani, 2013) as well as natural images (Naselaris et al., 2009). The studies of Thirion et al., 2006 and Miyawaki et al., 2008 also used the reconstructions to perform synthetic image identification by comparing the

reconstructed image with a set of candidate images. The study of Naselaris et al., 2009 used their image reconstructions to select images that were most similar to the presented image in terms of image structure and semantic content.

In the study of Kay et al., 2008, they use an encoding model of the visual system to identify natural images. They model the internal representation of the visual cortex using a Gabor-Wavelet-Pyramid (GWP). The GWP has parameters for position, orientation and spatial frequency. The Gabor filters of the model are applied to an image, and the combined outputs of the filters make the response prediction for that image. The main advantage of this approach (and for the pRF-model) is that it can predict the activation pattern for any image.

Using the GWP for image identification instead of the pRF-model leads to a better identification performance (Kay et al., 2008). Several differences between these studies may account for performance differences. First, Kay et al presented images for 1 second at a time, while we presented the images for 9 seconds. Longer presentations might lead responses to reflect more influences from higher-level image representations, which neither model represent. Second, Kay et al. used larger images (20x20° of visual angle) than we did (11x11°). Larger images would cause responses in a larger extend of visual cortex, which provides more data for identification. Third, the models differed greatly in their complexity. Where the standard pRF model uses only 3 free parameters to capture each recording site's response, the GWP uses up to 2,370. Kay et al. also included a retinotopic-only (RO)-model for image identification. In this model they removed the parameters for orientation and spatial frequency. Therefore, this model has fewer parameters compared to the original GWP. This model is conceptually similar to the pRF-model we use, but there are important differences. First, our pRF-model describes the pRF using one single circular Gaussian, while the RO-model is free in the shape of the responsive area. Second, the RO-model therefore has more free parameters than our pRF model. More parameters in the model will allow more accurate representation of voxels responses, but the drawback of using a model with a large amount of free parameters is that a large dataset has to be used to train the model. The estimation process of the model uses the fMRI data elicited by the example images of the training set. This means that for using models with more free parameters, more scanning time is needed as well. The GWP was estimated using a set of

1,750 different natural images. To obtain the data for estimating the GWP takes about 5 hours. To compare this with the pRF-model, this was estimated using standard mapping stimuli with 81 differential images, which only take about half an hour of scanning.

Conclusion

The pRF-method is a fast and simple biologically inspired model. We show that, even when training it on conventional pRF mapping stimuli, it can be used for the identification of visual images, including natural images. The advantage of using the pRF-method for image identification is that the model has a minimal amount of free parameters. Collecting the fMRI data for the estimation of the pRF-model can be done within half an hour of scanning using standard mapping stimuli. This makes the pRF-model a convenient method to be used for the identification of untrained images, including natural images.

Chapter 5

Enhanced responses in early visual cortex to subjectively important aspects of natural scenes

Wietske Zuiderbaan, Serge O. Dumoulin (in preparation)

Acknowledgements of author contributions: WZ and SOD designed the experiment. WZ collected the data, WZ performed data analysis under supervision of SOD. WZ wrote the manuscript and SOD provided critical comments.

Abstract

Our visual percept is not solely based on the sensory information that we receive from the light falling on our retina, but is also influenced by our knowledge of the world. The interaction between sensory information and our world knowledge is the basis of perception, yet there is no consensus about how this interaction influences the neural signal in early visual areas. Current theories implement this knowledge of the world as a perceptual hypothesis that is compared with the sensory input. According to different inference theories, the perceptual hypothesis can either be suppressing or boosting the sensory information represented by early visual areas. In our study we investigated the effect that the knowledge-based perceptual hypothesis has on the neural signal in early visual cortex and how this interacts with the responses to the sensory-driven image feature contrast.

We used 7T MRI to measure responses to different type of stimuli. First, we measured the responses elicited by conventional pRF mapping stimuli. From the responses to the pRF mapping stimuli we estimated the pRF-properties in the early visual areas. The pRF-model estimates for every cortical location the region of the visual field it responds to. Second, we measured the responses elicited by natural images. We used the pRFs to quantify both the amount of RMS-contrast (sensory-driven) and the amount of subjective importance (knowledge-based) in the pRF. Based on the inherent variations of contrast in the natural images we show that we can derive the contrast response function (CRF). Furthermore, we show how the CRF is modulated by the perceptual hypothesis as we defined it by subjective importance. We see that the CRF was boosted in visual areas V1-V2-V3 when the responses were elicited by parts of the image of high subjective importance. Thus, the sensory-driven image representation of a scene in early visual areas is boosted by the knowledge-based perceptual hypothesis.

Introduction

Our visual percept comes from the inference that the visual system makes from the sensory information it gets of the retina. The visual system infers the percept of our visual environment based on the sensory information and our knowledge of the world. Different computational and statistical theories link the percept to the neural representation. These theories propose that higher-order areas modulate the responses of early visual areas according to the perceptual hypothesis. This modulation is achieved through feedback from the higher-order areas to earlier visual areas. The difference of the theories lies in the effect that the feedback has on the neural representation of the earlier visual areas; it can be either suppressive or boosting.

Figure 1 illustrates two different theories. The predictive coding theory states that the feedback of higher-order areas subtracts the perceptual hypothesis from the neural representation of earlier visual areas (K. Friston, 2002; Mumford, 1992; Rao & Ballard, 1999). This causes the earlier visual areas to signal only the unpredicted signals; the residuals. In this case, the perceptual hypothesis has a suppressive effect on the neural representation of early visual areas (Fig 1A). On the other hand, the efficient coding theory (Barlow, 2001; K. Friston, 2002; Rao, 2005; Series, Lorenceau, & Fregnac, 2003; Simoncelli & Olshausen, 2001) proposes that higher-level areas will not subtract, but instead will increase the responses according to the perceptual hypothesis of the scene (Fig 1B). Opposite effects on the neural representation in early visual areas have been found for the predictability and task-relevance of a stimulus (for a review, see: (Rauss, Schwartz, & Pourtois, 2011). In short, there is no decisive answer to how the neural representation is influenced by the perceptual hypothesis and how this interacts with the responses to sensory-driven information. Furthermore, these effects are typically measured using synthetic images that are systematically manipulated. Ultimately, these measures are assumed to extrapolate to natural images.

In this study we investigated how the sensory-driven and knowledge-based information influence the neural representation in visual cortex using natural images. We measured the

responses using 7T fMRI to different sets of stimuli: standard mapping stimuli, 45 natural images and a full-field stimulus (100% contrast). The mapping stimuli were used to estimate for every cortical location the region of visual space it responds to: the population receptive field (Dumoulin & Wandell, 2008). For every cortical location we estimated the properties of its pRF (location and size). The pRF-model gives us the opportunity to select voxels that respond to certain parts of the image. Therefore, we can make use of the inherent variations of the sensory-driven and knowledge-based information of the natural images to investigate how both aspects influence the neural response.

First, we show how the neural responses are modulated by the sensory-driven image feature contrast. Neural responses in early visual cortex are known to be strongly modulated by contrast (Boynton, Demb, Glover, & Heeger, 1999; De Valois & De Valois, 1988; Dumoulin, Dakin, & Hess, 2008; Mante & Carandini, 2005; Olman, Ugurbil, Schrater, & Kersten, 2004). These neural responses do not sum linearly to an increase in contrast, but saturate at levels of high contrast (Albrecht & Hamilton, 1982). Here, we used the inherent variation in contrast of natural images to derive the contrast response function (CRF) from the measured fMRI responses to natural images. Second, we investigated the role of the perceptual hypothesis on the sensory-driven representation of the early visual areas. We used images from the Berkeley Segmentation Dataset and Benchmark database (D. Martin, Fowlkes, Tal, & Malik, 2001). In this dataset, Martin and colleagues asked human observers to identify the most important aspects of the image (Fig. 2C and Supplementary Fig. 1). We used these manually labeled aspects of the scene that represent the subjective importance of the parts of the image to define and quantify our measure for the perceptual hypothesis in the pRF.

We investigated how the perceptual hypothesis influences the sensory-driven representation of the early visual areas. For this, we show how the CRF is modulated by the subjective importance. The predictive coding theory would expect to find a suppressed CRF (Fig. 1C) for parts of the image of high subjective importance (purple) compared to the CRF elicited by parts of the image of low subjective importance (green). On the other hand, the efficient coding theory would expect the opposite; a boosted CRF (Fig. 1D) for

high subjective importance (purple) compared to the CRF of low subjective importance (green).

We see that the CRF is modulated by the amount of subjective importance. The CRFs of high subjective importance are boosted with respect to the CRF of low subjective importance. This result is in support of the efficient coding theory that states that the perceptual hypothesis boosts the sensory-driven image representation in early visual areas.

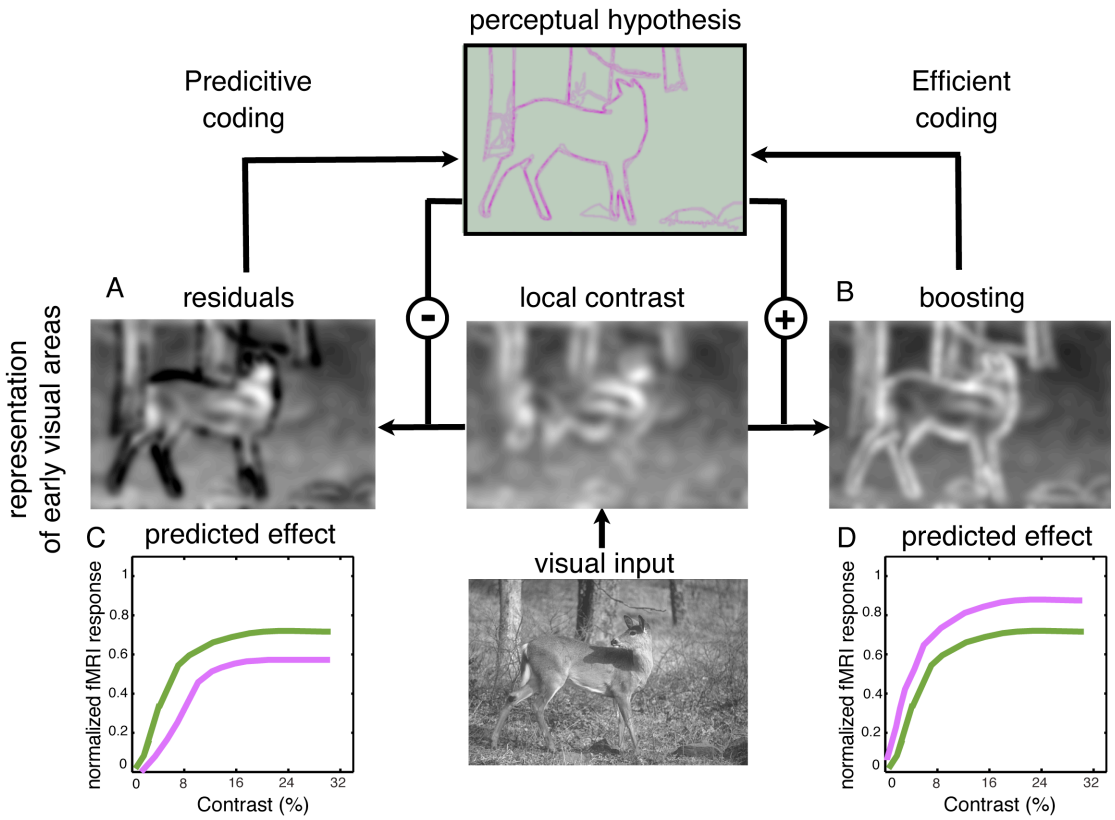


Figure 1 How does the knowledge-based perceptual hypothesis interact with the sensory-driven image representation in early visual areas? This figure illustrates two different theories about inference. The predictive coding theory states that the early visual areas signal the residuals; the difference between the sensory-driven image representation and the knowledge-based perceptual hypothesis. The perceptual hypothesis has a suppressive effect on the sensory-driven image representation (A). On the other hand, the efficient coding states that the perceptual hypothesis boosts the sensory-driven representation according to the perceptual hypothesis (B). Both theories predict a different effect on the contrast response function (CRF). The predictive coding theory predicts a suppressed CRF (C) for responses to parts of the image of high subjective importance (purple) compared to CRF made from responses to parts of the image of low subjective importance (green). The efficient coding theory predicts the opposite. This theory predicts a boosted CRF (D) for responses to parts of the image of high subjective importance (purple) compared to CRF made from responses to parts of the image of low subjective importance (green). We investigated how the neural representation in early visual areas is affected by both the sensory-driven and the knowledge based image information.

Methods

Subjects

Four subjects (male; ages 29-41 years) participated in this study. All subjects had normal or corrected-to-normal visual acuity. All studies were performed with the informed written consent of the subjects and were approved by the Human Ethics Committee of University Medical Center Utrecht.

Stimulus presentation

The visual stimuli were generated in Matlab using the PsychToolbox (Brainard, 1997; Pelli, 1997) on a Macintosh Macbook Pro. The stimuli were back-projected on a display inside the MRI bore. The subject viewed the display through mirrors inside the scanner. The size of the display was 15.0x7.9 cm with a resolution of 1024x538 pixels. The total distance from the subject's eyes to the display was 41 cm. The stimuli were constrained to a circular area with the size of the vertical dimension of the screen. The area outside this circle was maintained at a constant mean luminance. This gave the stimulus a radius of 5.5° visual angle from the subject's point of view.

Stimuli

pRF mapping stimulus

For the pRF mapping stimuli, we used bar apertures to train the pRF-model, these apertures were filled with natural image content (Fig. 2A). The natural image content for the bars came from images of the 'Berkeley Segmentation Dataset and Benchmark' database (D. Martin et al., 2001).

The width of the bar subtended 1/4th of the stimulus radius (1.375°). Four bar orientations (0°, 45°, 90° and 135°) and two different step directions for each bar were used, giving a total of 8 different bar configurations within a given scan. The bar stepped across the stimulus aperture in 20 steps (each 0.55° and 1.5 seconds) with each pass taking

30 seconds. A period of 30 seconds mean-luminance (0% contrast) was presented after every horizontally or vertically oriented pass. In total there were 4 blocks of mean-luminance during each scan, presented at evenly spaced intervals.

The subjects performed a fixation dot task to make sure they fixated at the center of the display. A small fixation dot (0.11° radius) was presented in the middle of the stimulus. The fixation dot changed its color from red to green at random time intervals and subjects were instructed to respond at this change of color using an air pressure button.

Natural images

The natural images came from the 'Berkeley Segmentation Dataset and Benchmark' database (D. Martin et al., 2001). The original resolution of the images was 321x481 pixels (both landscape and portrait). We selected a squared part of 321x321 pixels from the images and upsampled this to a resolution of 516x516 pixels, which corresponds to a stimulus of 11x11° diameter of visual angle. The images were masked by a circle with a raised cosine faded edge (width of 0.9°), and the areas outside this circle were set to mean luminance. The images were gamma-linearized and the mean contrast was set to 50%. We used 3 image sets in different scanning runs, each containing 15 different natural images (45 in total) and one full-field binarized bandpass-filtered noise stimulus (Fig. 2B). Figure 2C shows an example of the natural images. A fixation dot was presented at the center of the stimulus. The same fixation dot task as in the scans of the mapping stimuli was used.

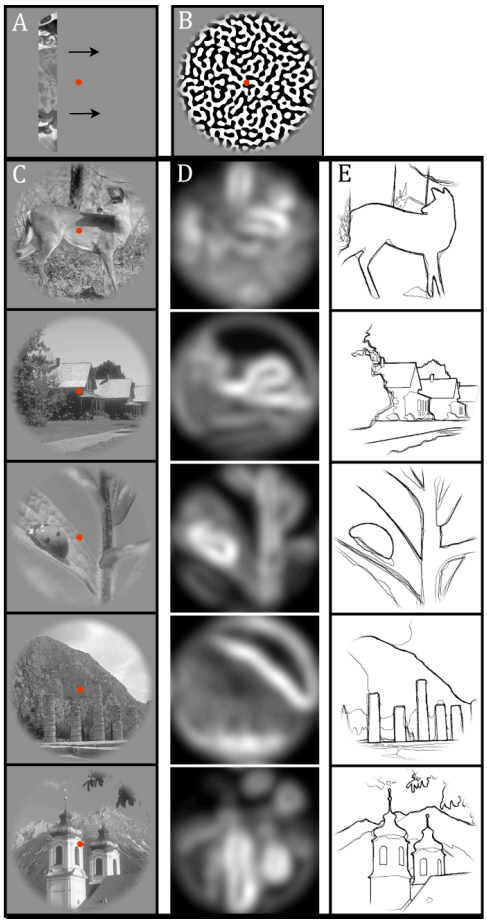


Figure 2 We used different stimuli in our study. A: The pRF mapping stimulus consists of a bar sweeping through the visual field in 8 different moving directions. B: We used the response to the full-field stimulus (100% contrast) to normalize the voxel's responses. C: We derived the CRFs from the responses to natural images. D: From the natural images we calculated the amount of local contrast. E: We quantified subjective importance by the manual labeling of the natural images. The images came from a database (D. Martin et al., 2001) in which observers drew lines on the image to highlight the parts of the image they considered to be important for the representation of the scene.

Functional imaging and processing

The MRI data was acquired with a Philips 7T scanner using a 32-channel head-coil. We scanned the participants with a 2d-echo-planar-imaging sequence with 25 slices oriented perpendicular to the Calcarine sulcus with no gap. The following parameters were used;

repetition time (TR) = 1500 ms, *echo time* (TE) = 25 ms and a flip angle of 80°. The functional resolution was 2x2x2 mm and the *field of view* (FOV) was 190x190x50 mm.
We used foam padding to minimize head movement. The functional images were corrected for head movement between and within the scans (Nestares & Heeger, 2000). For computation of the head movement between scans, the first functional volumes for each scan were aligned. Within scan motion correction was then computed by aligning the frames of a scan to the first frame.

The duration of the pRF mapping scans was 372 seconds (248 time-frames), of which the first 12 seconds (8 time-frames) were discarded due to start-up magnetization transients. During the three sessions we acquired 6-8 pRF mapping scans in total per subject. To obtain a high signal-to-noise ratio, we averaged the repeated scans. During the three sessions for the natural images we acquired 6-7 scans each for the three natural image sets.

The duration of the scans with the natural images was 432 seconds (288 time-frames). The first 12 seconds (8 time-frames) were discarded due to start-up magnetization transients. The images were presented in a block design. Each image was presented during a 9-second block. Within this block the same image was shown 18 times for a duration of 300ms followed by 200ms mean-luminance. The full-field stimuli were presented with 3 alternating different high-contrast patterns, to obtain a full high-contrast response that is not based upon one specific high-contrast pattern. The block in which the stimulus was presented was followed by a 12 second mean-luminance presentation. Four longer blank periods were inserted during the scan, each of 33 seconds.

Anatomical imaging and processing

The T1-weighted MRI images were acquired in a separate session using an 8-channel SENSE head-coil. The following parameters were used: TR/TE/flip angle = 9.88/4.59/8. The scans were acquired at a resolution of 0.79x0.80x0.80 mm and were resampled to a resolution of $1mm^3$ isotropic. The functional MRI scans were aligned with the anatomical MRI using an automatic alignment technique (Nestares & Heeger, 2000). From the anatomical MRI, white matter was automatically segmented using the *FMRIB's Software Library* (FSL) (Smith et al., 2004). After the automatic segmentation it was hand-

edited to minimize segmentation errors (Yushkevich et al., 2006). The gray matter was grown from the white matter to form a 4 mm layer surrounding the white matter. A smoothed 3D cortical surface can be rendered by reconstruction of the cortical surface at the border of the white and gray matter (Wandell, Chial, & Backus, 2000).

pRF model-based analysis

The first step for image identification is the estimation of the pRF-model from the measured fMRI signal that was elicited by the pRF mapping bar stimuli (Fig 2A). The model estimates a pRF for every cortical location using a method previously described (Dumoulin & Wandell, 2008). In short, the method estimates the pRF by combining the measured fMRI time-series with the position time course of the visual stimulus. A prediction of the time-series is made by calculating the overlap of the pRF and the stimulus energy convolved with the haemodynamic response function (HRF). We estimated the parameters of the HRF that best describes the data of the whole acquired fMRI volume (Harvey & Dumoulin, 2011). The optimal parameters of the pRF-model are chosen by minimizing the residual sum of squares between the predicted and the measured time-series. We used the simplest pRF-model which consists of a circular symmetric Gaussian. This model has four parameters: position (x_0,y_0), size (σ_1), and amplitude (β_1). For further technical and implementation details see (Dumoulin & Wandell, 2008).

Region of interest

Using the pRF-method, we estimated position parameters x_0, and y_0 of the pRF per voxel. From these values, the polar angle ($\text{atan}(y_0 / x_0)$) and eccentricity ($\sqrt{(x_0^2 + y_0^2)}$) values can be calculated. We drew the borders of the visual areas on the basis of their location in the visual field (DeYoe et al., 1996; Engel, Glover, & Wandell, 1997; Sereno et al., 1995; Wandell, Dumoulin, & Brewer, 2007) by rendering these polar angle and eccentricity maps onto the inflated cortical surface (Wandell et al., 2000). We defined visual areas V1-V2-V3 as our *regions of interest* (ROI).

Deriving the contrast response function

Analysis of voxel responses to the natural images

We measured fMRI responses to 45 natural images and 1 full-field stimulus (100% contrast). We first determined the voxel response amplitudes in %BOLD signal change as they were elicited by each of these images. The voxel responses were calculated using a general linear model (GLM) (K. J. Friston et al., 1998; K. J. Friston et al., 1995).

To reduce the noise from the individual voxel differences in response amplitudes, we normalized the responses according the voxel's response to the full-field (100% contrast) stimulus. Therefore, we divided the voxel's responses of the natural images (in %BOLD signal change) by the voxel's response to the full-field stimulus.

Voxel selection

To make the CRFs, we only used the voxels that showed an overall significant response (t-values>4.0), pRF eccentricity values from 0.5-4° and pRF variance explained above 40%. Furthermore we used a threshold for the sizes of the pRF, dependent on the visual area (V1 sigma: 0.25-0.8 V2 sigma: 0.25-1.1 V3 sigma: 0.25-1.75).

Contrast values

To derive the CRF, we computed the amount of contrast of every natural image within each pRF (Fig. 3B). We determined these contrast values from the pRF-model as follows. The pRF is modeled as a circular symmetric Gaussian function, described by parameters for position (x_0,y_0) and size (σ).

The Gaussian weighting function of the pRF is defined by w_i:

$$w_i = \exp - \left(\frac{(x_i - x_c)^2 + (y_i - y_c)^2}{2(\sigma)^2} \right) \quad (2)$$

Where x_c and y_c define the location of the center of the pRF in the visual field, σ is the size of the pRF and x_i and y_i define the location of the *i*th pixel.

We computed each voxel's contrast value to each natural image by calculating the Root-Mean-Squared (RMS) contrast (Bex & Makous, 2002; Pelli, 1997) of the part of the image inside the voxel's pRF. RMS contrast is defined as the standard deviation of the luminance intensities relative to the mean. The RMS-contrast is weighted by the pRF Gaussian function to obtain the local contrast-energy value per pixel:

$$local\_contrast\_energy = \sqrt{\frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \frac{(L_i - L)^2}{L^2}} \quad (3)$$

Where N is the number of pixels in the stimulus window. L is the mean luminance from the pixels inside the spatial window, and L_i is the luminance of the *i*th pixel.

Plotting of the CRF

The CRFs were made by plotting the fMRI responses as a function of the amount of contrast inside the pRF of the according voxels. The voxel's responses were binned based on the amount of contrast inside the pRF. This binning is per percentile such that every bin contains 12.5% of the total amount of voxels. We plotted the CRFs using the voxels of visual area V1, V2 and V3.

We fitted CRFs using the following equation (modified from(Boynton et al., 1999)):

$$R(C) = a\frac{C^q}{C^q + Q^q} \tag{4}$$

Where R is the fMRI response and C is the amount of RMS-contrast inside the pRF. The variables for q and Q define the shape of the CRF. The parameters of this function were fit to give the highest variance explained to the fMRI data.

These CRFs can be modulated by different mechanisms (Carandini & Heeger, 2011). Here we investigated the role of the knowledge-based perceptual hypothesis on the CRF.

Quantifying the knowledge-based perceptual hypothesis

The natural images came from the Berkeley Segmentation Dataset and Benchmark database (D. Martin et al., 2001). In this dataset, Martin and colleagues asked human observers to identify the most important aspects of the image (Fig. 2C). In each manually labeled image, a human observer drew lines on the image to highlight the parts of the image they considered to be important for the representation of the scene. We use these measurements of the subjective important aspects of the scene to define and quantify the knowledge-based perceptual hypothesis of the image. We took the average of 5 manually labeled images as our definition of the perceptual hypothesis.

Along with the contrast of the image in the pRF (Fig 2B), we also calculated the amount of subjective importance in the pRF (Fig 2C). The pixels of the manually labeled images have values between 0 (not labeled) and 1 (pixel labeled by all 5 observers). The amount of subjective importance in the pRF is calculated by a weighted sum of the pixels in the averaged manually labeled images and the pRF. This weighted sum was normalized by the total volume of the pRF:

$$subjective\_importance = \frac{\sum_{i=1}^{N} w_i \bullet S_i}{\sum_{i=1}^{N} w_i} \tag{5}$$

Where N is the total number of pixels in the stimulus area. S is the manually labeled image, where S_i is the value that specifies how many observers labeled the pixel of the image as important for the visual percept. A value of 0 indicates that none of the observers labeled the pixel as important for the visual percept of the scene. With increments of 0.2 for every observer labeling the pixel to a maximum of 1 where every observer labeled the pixel of the image. The variable w_i is the Gaussian weighting function of the pRF defined by formula 2.

Based on the amount of subjective importance inside the pRFs we split the voxel's responses. We made two CRFs; based on responses that were elicited by voxels that were either 'low' or 'high' in the amount of subjective importance. The definition of 'low' and 'high' was based on the 50th percentile of the values for the amount of subjective importance inside the pRF. We took the first 50th percentile of the changes for the condition 'low' and the second 50th percentile for the condition 'high'. We did not observe any differences either for sigma, eccentricity and the fMRI responses to the full-field stimulus between the voxels of 'low' and 'high' subjective importance at all reported contrast bins.

Statistics

The CRFs were compared for the difference in the amount of subjective importance. To see whether the CRF obtained from the 'low' subjective important responses was statistically different than the one obtained with 'high' subjective important responses. We bootstrapped the area under the fitted curve of the CRFs to obtain the 95% confidence interval and the p-values to test for a statistical difference in the CRFs. We tested the CRFs of visual areas V1-V2-V3 for statistical differences caused by the amount of subjective importance.

Results

CRF can be derived from the inherent contrast of natural images

We plotted the fMRI responses according to the amount of RMS-contrast inside the pRFs. This shows how the fMRI response changes with increasing contrast; the CRF. Figure 2B shows the CRF made from the responses of the voxels in visual area V1. The individual points represent the mean normalized fMRI response from the data of all our subjects, and the errorbars represent the standard error of the mean. The line was fit from the data by the CRF-function described in formula 4. The variance explained for the line to the data was > 87%. The CRFs show a non-monotonic increase for the response amplitudes elicited by high-contrast regions of the image as it is known to occur based on synthetic stimuli (Albrecht & Hamilton, 1982; Boynton et al., 1999; Kay, Winawer, Rokem, Mezer, & Wandell, 2013; Legge, 1981; Ohzawa, Sclar, & Freeman, 1985; Sclar, Maunsell, & Lennie, 1990).

Enhanced responses for subjective importance in visual area V1

To investigate the effect of the knowledge-based perceptual hypothesis on the CRF, we calculated besides the amount of contrast inside the pRF (Fig. 2B), also the amount of reported subjective importance inside the pRF (Fig. 2D). In Fig. 2A four example pRFs are shown; the green pRFs are (in this image) at a location of low subjective importance and the pink pRFs are at a location of high subjective importance. The pRFs with the closed black line around it are at a location of low contrast, and the pRFs with the striped black line are at a location of high contrast. These four pRFs represent the combination of the different levels (low vs. high) of both the sensory-driven aspect and the knowledge-based aspects.

Fig. 2E shows the CRFs in visual area V1 split for the amount of subjective importance inside the pRFs. The green line represents the CRF for the responses to parts of the image that were of low subjective importance, and the pink line represents the CRF elicited by responses to parts of the image with high subjective importance. Again, the

individual points represent the mean (normalized) fMRI response from the data of all our subjects and the errorbars represent the standard error of the mean. The line was fit from the data by the CRF-function described in formula 4. The variance explained for the lines to the data was >85%.

We see that the CRF is modulated by the amount of subjective importance of the parts of the images (falling inside the pRF). Both the CRF of low- and high-subjective importance show a saturating effect with an increase of contrast-energy. Furthermore, the CRF is boosted when it is elicited by voxels that respond to a part of the image that is labeled as highly subjective important. Thus, high levels of subjective importance increase the neural response at all contrast levels.

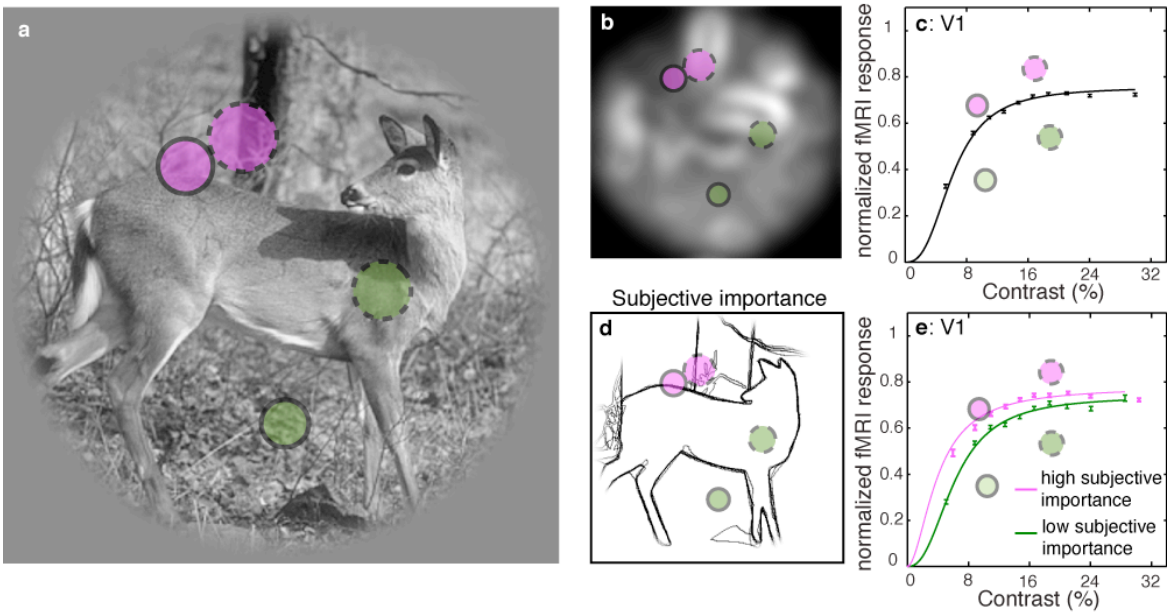


Figure 3 In this study we investigated how the sensory-driven and knowledge-based image information influence the neural representation. We calculated the amount of RMS-contrast inside the pRF (B) and plotted the fMRI response (normalized for the response elicited by the full-field stimulus) according to the amount of contrast inside the voxel's pRF; the CRF (C). The individual points represent the mean (normalized) fMRI response and the errorbars represent the standard error of the mean. The line was fit from the data by the CRF-function described in formula 4. Besides the amount of contrast-energy, we also calculated the amount of the subjective importance inside the pRF (D). Four example pRFs from visual area V1 are shown on the image, with the according responses in the graphs of the CRFs of visual area V1. The green pRFs are located on an area of low subjective importance, where the pink pRFs are located on an area of high subjective importance. Furthermore, the closed black line indicates that the pRF is on a location of the image of low contrast, where the striped black line indicates that the pRF is on a location of high contrast. We made two CRFs (E); one for the values with low subjective importance (green) and one for the high subjective importance (pink). We see that the CRFs are boosted for the responses to areas of high subjective important.

Enhanced responses for subjective importance in visual area V1-V2-V3

Figure 3 shows the CRFs in visual area V1, V2 and V3 split for the amount of subjective importance inside the pRFs. We do not only see the saturating effect of the fMRI response with increasing contrast in V1, but also in V2 and V3. Furthermore, we see the same effect for subjective importance on the CRF in V2 and V3 as we see it in V1. The CRF is boosted for high subjective importance in V1, V2 and V3. Again, the green line represents the CRF for low subjective importance and the purple line represents the CRF for high subjective importance. The individual points represent the mean normalized fMRI response for all our subjects and the errorbars represent the standard error of the mean. The variance explained for the fitted lines to the data was >80%.

These results cannot be explained by differences between the selected voxels of 'low' and 'high' subjective importance in voxel response (we checked for response to the full-field stimulus), sigma or eccentricity. To quantify the differences in the CRF according to the low and high amount of subjective importance, we used a bootstrapping procedure. We bootstrapped the distributions of the area under the fitted CRF of low- and high subjective importance. We found for visual area V1, V2 and V3 a statistical significant difference of the CRF for the high subjective importance vs. the low subjective importance ($p<0.001$). We interpret this as an increase of the CRF according to subjective importance.

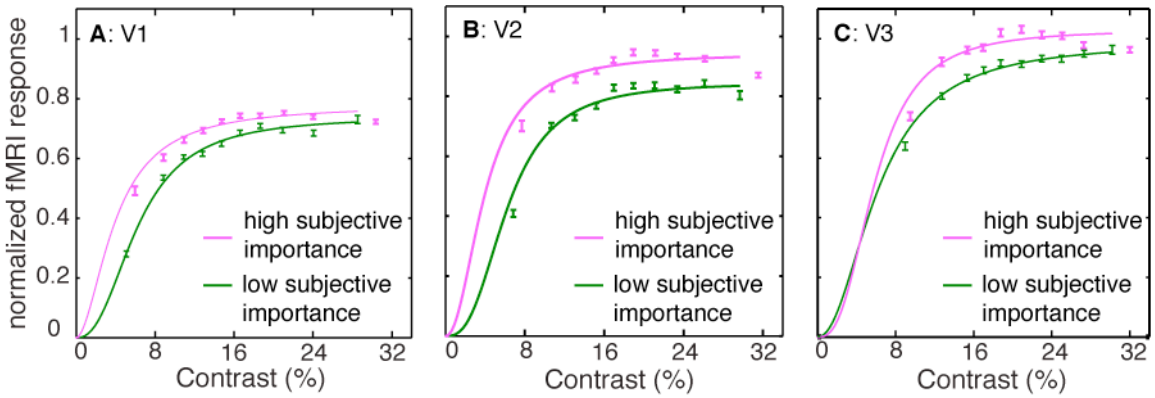


Figure 3 The CRFs for visual areas V1, V2 and V3 split for the amount of the perceptual hypothesis defined by subjective importance. We made two CRFs per visual area; one for the values with low subjective importance (green) and one for the high subjective importance (pink). We see that the CRFs are boosted for the responses to areas of high subjective important in visual areas V1, V2 and V3.

Discussion

In our study we show that we can derive the CRFs based on the inherent contrast of the images. Furthermore, we show that the CRF is boosted in visual areas V1-V2-V3 when it is elicited by responses to regions of high subjective importance. We suggest that this is evidence for the efficient coding theory, which proposes that the sensory-driven image representation in early visual areas is boosted according to the knowledge-based perceptual hypothesis.

Effects of local image contrast on the neural response

First, we analyzed the neural responses according to the (inherent) amount of contrast in the natural images. Plotting the amount of contrast in the pRF according to the normalized fMRI responses, we see that the responses rise with increasing contrast in the pRF for visual areas V1, V2 and V3. However, the responses do not rise linearly with increasing contrast, but saturate at high contrast. This non-monotonic increase in response amplitudes with contrast is a characteristic shape of the CRF. The CRFs as we measured

them using the natural images and the pRF method show a typical shape of the CRF as they are measured using earlier fMRI studies (Boynton et al., 1999; Kay et al., 2013; Olman et al., 2004), psychophysics (Boynton et al., 1999; Legge, 1981; Olman et al., 2004) as well as electrophysiological studies (Albrecht & Hamilton, 1982; Ohzawa et al., 1985; Sclar et al., 1990). CRFs are typically measured using synthetic images; we show that these results can be extrapolated to natural images.

Modulation of neural responses by subjective importance

After deriving the CRF according to the sensory-driven image feature contrast, we investigated how the neural responses in early visual areas (based upon the CRF) are modulated by the knowledge-based perceptual hypothesis. We quantified the knowledge-based perceptual hypothesis by the subjective importance within the image. Besides calculating the amount of contrast energy inside the pRF, we also calculated the amount of subjective importance inside the pRF.

Here we show for visual areas V1, V2 and V3 that when the CRF is derived from responses elicited by regions of high subjective importance, the CRF is boosted compared to the CRF of low subjective importance. This result is in support of statistical theories of inference where feedback connections have an excitatory role and boost responses in early visual cortex (Barlow, 2001; K. Friston, 2002; Rao, 2005; Series et al., 2003; Simoncelli & Olshausen, 2001).

Efficient versus predictive coding

The theories of efficient coding and predictive coding predict a different effect of the subjective importance on the CRF. In both theories, higher visual areas form a hypothesis on the cause of the sensory data. This hypothesis is sent back to lower visual areas via feedback to compare this hypothesis with the bottom-up representation. The theories differ in the effects that the top-down mechanisms have on the bottom-up processes. In the efficient coding theory the feedback is boosting, whereas in the predictive coding theory the feedback has a subtractive effect on the sensory-driven representation. In our study, we find a boosting effect on the CRF for subjective importance. Where our result is in line

with the efficient coding theory, the predictive coding theory predicts a suppressive effect on the CRF for subjective importance. According to the predictive coding theory, top-down responses cause the early visual areas to silence the expected and signal the unexpected. This is supported by neuroimaging studies that find suppressed responses in early visual areas when the stimulus matches the expectation (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; den Ouden, Friston, Daw, McIntosh, & Stephan, 2009; Kok, Jehee, & de Lange, 2012; Kok, Rahnev, Jehee, Lau, & de Lange, 2012; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008). Increased responses are found to regions inducing an illusory contour (no sensory information), together with decreased responses to regions receiving sensory information (Kok & de Lange, 2014). Furthermore, Murray et al. find increased fMRI activity in object sensitive area LOC together with a reduced activity in visual area V1 when presenting coherent shapes (Murray, Kersten, Olshausen, Schrater, & Woods, 2002). This is in agreement with the idea of predictive coding that higher visual areas subtract their hypothesis through feedback from the bottom-up representation in lower areas. Several neuroimaging studies have found suppressed responses in early visual areas for top-down influences.

Why do we not find this suppressive effect? There can be different reasons. There are important differences between our study and other studies that find suppressive effects in early visual areas for top-down influences; they are related to the definition of the knowledge-based perceptual hypothesis, attention, local image statistics and figure-ground segmentation. We will discuss each influence in turn below.

Definition of the knowledge-based perceptual hypothesis

Most studies that find supporting evidence for the predictive coding theory see 'expected' as their definition for the perceptual hypothesis and use this as a learned stimulus property specific for the experiment. This learned property can be either the frequency of the presented stimulus (Summerfield et al., 2008), or priming of the stimulus for location (Kok, Rahnev, et al., 2012), orientation (Kok, Jehee, et al., 2012) or presence of the stimulus (den Ouden et al., 2009). In these studies, the top-down effects contain information about a property of the stimulus that was explicitly learned for the experiment. What is different in

our study is that we quantify the influence of top-down effects using subjective importance in natural images. Whereas previous studies use synthetic images, these will lack the amount of knowledge-based information that is inherently available in natural images. Natural images differ along many dimensions, which may explain the different results. For example, attention, local image statistics and image content.

Attention

Deviating results have been found for the effect of predictability on the neural response. These deviating results are arguably explained by the task-relevance of the stimulus (Rauss et al., 2011) or in terms of attention (Kok, Rahnev, et al., 2012; Summerfield & Egner, 2009). A task-relevant stimulus enhances, rather than reduces the neural response of a predictable stimulus (Rauss et al., 2011). Kok et al. propose that this can be explained by attention (Kok, Rahnev, et al., 2012; Summerfield et al., 2008). Attention is known to increase the neural response in visual cortex (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999). When predictability is combined with attention, attention seems to reverse the suppressive effect of predictability (Kok, Rahnev, et al., 2012; Summerfield et al., 2008). In the study of Kok et al., predictability is a task-irrelevant cue and codes for the likelihood of a stimulus to appear, where attention is coded by a task-relevant cue. Recent models of predictive coding have been adapted to account for increasing effects of attention (Feldman & Friston, 2010; Spratling, 2008).
The difference with our study is that our subjects performed a fixation dot task that was irrelevant for the stimulus. We can therefore not explain our results using attention with the definition of task-relevance as it is used in previous studies.

Local image statistics

The manual labels that we use to define subjective importance can be made by using different cues, from low-level and mid-level to high-level knowledge based cues. Besides the sensory-driven property contrast, they can contain different statistical properties that can affect the responses in visual cortex such as orientation (Freeman, Brouwer, Heeger, & Merriam, 2011; Haynes & Rees, 2005; Kamitani & Tong, 2005; Kay, Naselaris, Prenger, & Gallant, 2008) and spatial frequency (Henriksson, Nurminen, Hyvarinen, & Vanni,

2008; Kay et al., 2008; Olman et al., 2004; Singh, Smith, & Greenlee, 2000). These statistical properties are shown to be confounded with results of shape-perception (Dumoulin & Hess, 2006). Potentially any of these can contribute to the enhanced responses. This is a limitation of using natural images, where all possible interactions take place. On the other hand, our analysis can take this into account, though the possibilities in natural images are ultimately unlimited.

For the dataset of images used in the experiment, Martin et al. have compared the manual labels with the performance of local boundary detector algorithms. Where the algorithms detect parts of the manual labels, a significant portion was not detected (D. R. Martin, Fowlkes, & Malik, 2004). This suggests that the manual labels were not only based on local image properties, but also on global image context and higher-level prior knowledge.

Image content

The manual labels may represent (for instance) shape information or arguable, figure-ground representations. The inactivation of higher extra-striate visual area V5 can decrease responses in early visual areas (V1, V2 and V3) and impair figure-ground representations (Hupe et al., 1998). Also, figure-ground representations can increase the neural response in early visual areas (Lamme, 1995; Lee, Mumford, Romero, & Lamme, 1998). This is in line with our results that support the efficient coding theory in which responses in early visual areas are increased according to the perceptual hypothesis of later visual areas.

Supporting evidence for the predictive coding theory has been found for the perception of contours and shape. Murray et al. find increased fMRI activity in object sensitive area LOC together with a reduced activity in visual area V1 when presenting coherent shapes (Murray et al., 2002). However, they looked at an overall activity reduction of all the recording sites in V1. Possibly, there is an increase for recording sites responding to locations coding for the shape, but a reduced net activity in V1.

Furthermore, suppressed responses in V1 are found for contours matching the perceptual hypothesis. Responses to regions where the sensory information matched the percept were decreased, whereas the responses to regions that induced the illusion (no matching sensory

information) were increased (Kok & de Lange, 2014). Where both previous studies used synthetic images, we used natural images.

From synthetic to natural images

Commonly, the effects of different image features on the neural response are measured using synthetic stimuli that can be easily manipulated. However, results from studies using synthetic images do not always extrapolate to natural environments (Carandini et al., 2005; David, Vinje, & Gallant, 2004; Kayser, Kording, & Konig, 2004; Li, VanRullen, Koch, & Perona, 2002). Synthetic images do not contain the same knowledge-based information as we encounter it in the real world. Therefore, there is a need to use more naturalistic stimuli in experiments.

Here, we use the inherent variation of image properties in natural images to analyze how these modulate the fMRI responses. The strength of the pRF-model is that it allows the selection of voxels to specific parts of the image. By selecting voxels that respond to certain image properties we can analyze how these properties modulate the fMRI response. Future studies can disentangle the effects of other image properties. This analysis technique gives us the chance to extrapolate results from synthetic images to natural images.

Conclusion

We measured the interactive effects that the sensory-driven image property contrast and the knowledge-based perceptual hypothesis have on the neural response. Using the pRF method, we are able to quantify different image properties in the pRF. We show that by using the inherent contrast in the natural images we are able to derive the CRF. Furthermore, we show that the responses to parts of the images of high subjective importance are boosted with respect to the responses to parts of the images that observers consider being less important for the representation of that scene. Where these effects are commonly measured using synthetic stimuli, we show how these results extrapolate to the use of natural images.

Chapter 6

General discussion

The aim of this thesis was to gain more insight in how the sensory-driven information and knowledge-based information influence visual perception. We describe a number of studies where we investigate how these information sources affect the neural signal (Chapter 3, 4 and 5) and our behavior (Chapter 2). We used a combination of behavioral methods with computational neuroimaging techniques. In our experiments we used both synthetic and natural images. Where the synthetic images contained mainly sensory-driven information, the natural images contained both sensory-driven and knowledge-based information. In this thesis, we present three new analysis techniques that extend the analysis of the pRF-method (Dumoulin & Wandell, 2008) to measure contextual influences of both sensory-driven and knowledge-based information in visual cortex (Chapter 3, 4 and 5). Furthermore, we investigated the interaction of the sensory-driven and knowledge-based information of a scene on our behavior in a change detection task (Chapter 2). The experiments in this thesis provide a quantitative link between physical sensory stimulation and our subjective perceptual experience on the neural signal and our behavior.

The effect of sensory-driven information on the neural response

Sensory-driven contextual influences are measurable using the pRF-method

In Chapter 3 we show that we are able to measure sensory-driven contextual influences of surround suppression using the pRF-method. Where the original pRF-model describes the pRF with only a positive responding region, we see that negative responses remain unexplained by the model. We extended the original pRF-model by incorporating a suppressive surround in the description of the pRF. The pRF is now modeled as a Difference of Gaussians (DoG) function (Angelucci & Bressloff, 2006; DeAngelis, Freeman, & Ohzawa, 1994; Rodieck, 1965; Sceniak, Ringach, Hawken, & Shapley, 1999). The DoG-model captures a center-surround organization and can modulate its prediction according to information outside the classical receptive field. We show that the DoG-model can explain negative fMRI responses and explains the measured responses better in early visual areas (Zuiderbaan, Harvey, & Dumoulin, 2012).

A strength of the pRF-model is that the parameters of the model can be compared. This allows us to study different properties of the underlying neuronal population. Where electrophysiology is an invasive method that is mainly used to study neural properties in animals, fMRI is a non-invasive method that can be used to study neural properties in human subjects. Using the pRF-method, results from electrophysiology can now be extrapolated to pRF measurements from human subjects. Also, clinical populations show abnormal perceptual functioning, presumably caused by altered center-surround properties (e.g. (Butler, Silverstein, & Dakin, 2008; Dakin & Frith, 2005; Marin, 2012; Yoon et al., 2009)). This method provides a direct measure to compare properties of the pRFs in both healthy and clinical populations. A nice example that shows the clinical relevance of our method is the study of Anderson et al., where they showed reduced surround suppression in people with schizophrenia (Anderson et al., 2017).

Deriving contrast response functions from the inherent contrast of natural images

In Chapter 5 we present a new analysis technique to derive contrast response functions (CRFs) from the inherent contrast of natural images using the pRF-method. Where CRFs are typically measured using synthetic images of different contrast levels (Albrecht & Hamilton, 1982; Boynton, Demb, Glover, & Heeger, 1999; Kay, Winawer, Rokem, Mezer, & Wandell, 2013; Legge, 1981; Ohzawa, Sclar, & Freeman, 1985; Sclar, Maunsell, & Lennie, 1990), we present a technique that can derive the CRF from natural images. Natural images contain different levels of contrast depending on the region of the image. Using the pRF-method, we can select responses caused by different levels of contrast in the natural images. Where CRFs are typically measured using synthetic images, we show that these results can be extrapolated to natural images. We derived the CRFs using the responses to 45 different images, but in theory, given that any natural image will contain different levels of inherent contrast, it is possible to derive the CRF from a single image presentation. CRFs can be modulated by several mechanisms (Carandini & Heeger, 2011). Using this method we are now able to measure modulations of the CRFs in natural images. In Chapter 6 we showed how the CRF is modulated by the subjective importance.

Predicting the presented stimulus from the measured fMRI signal

In Chapter 4 we present a new analysis technique to predict the visually presented stimulus from the measured fMRI signal using the most basic pRF-model. We show that the model can be used to identify images that the model was not explicitly trained on. Image identification has been demonstrated before (Kay, Naselaris, Prenger, & Gallant, 2008; Miyawaki et al., 2008; Thirion et al., 2006), but not using the pRF model. The advantage of using the pRF-model for image identification is its simplicity: the model is defined by a handful of parameters and can be estimated using standard mapping stimuli. Despite the synthetic nature of the mapping stimuli, we show that our results also transfer to the identification of natural images. We identified both synthetic and natural images based on the contrast information in the images. Where the identification accuracy was far above chance for both the synthetic and the natural images, the performance was lower for the natural images.

Interestingly, we see that (using the voxels of V1) some images are harder to identify than other images, and that this pattern is similar across subjects. We propose this is due to different reasons. The predictions of the pRF-model are made using only the local contrast-energy of the images. We show that images that are more similar in the information of contrast-energy were identified less accurately. However, after removing the relationship to image similarity defined by contrast-energy, we still see an image-specific correlation of identification accuracy across subjects. We suggest this is caused by the knowledge-based information present in natural images. Natural images are likely to be affected by complex image features and global image context (Petro, Vizioli, & Muckli, 2014). These properties are not captured by the pRF-model, which only captures the image contrast. We suggest that knowledge-based information of the scene is already present in the internal representation of V1. In Chapter 6 we investigated the effects of subjective importance on the internal representation in early visual areas.

Interaction of sensory-driven and knowledge-based information on the neural signal and behavior

Enhanced neural signal for the knowledge-based information

In Chapter 6 we show how the internal representation of early visual areas is encoded by both the sensory-driven and knowledge-based information. We first derived the sensory-driven representation by the fMRI responses to the local contrast-energy in the images. For this, we derived the CRF that shows how the neural response is affected by the local contrast-energy in the pRF. We investigated how the CRF is modulated by the knowledge-based image information as we defined it by subjective importance. We see that the CRF in early visual areas V1, V2 and V3 is boosted for parts of the image that are of high subjective importance. Thus, the internal representation of early visual areas not only encodes sensory-driven information, but also knowledge-based information.

This supports our view that in Chapter 5 some images were harder to identify since the neural responses to these images can be affected more by top-down contextual influences. For the identification process in Chapter 5, the predictions of the images were made using only the information of local contrast-energy of the images. When the internal representation of V1 also contains knowledge-based information, this will lead to an incomplete neural representation for the prediction of the pRF-model, since knowledge-based information is not taken into account in making the predictions using the pRF-model.

The study of Chapter 6 shows that the interaction of sensory-driven and knowledge-based information already takes place in early visual areas. Where previous studies use synthetic images, we now extrapolate these results to natural images.

In the Chapters 3, 4 and 5 we provide more insight in the information that is encoded in the internal representation of early visual areas. We see an interaction of sensory-driven and knowledge-based information in the neural signal of early visual areas. In Chapter 2 we investigated if this interaction is also reflected in our behavior.

Both the sensory-driven and knowledge-based information of a scene are important in change detection

In Chapter 2 we directly compare the role of sensory-driven and knowledge-based information on our ability to detect changes in a visual scene. We find that it is not only the knowledge-based image interpretation that facilitates change detection, as is suggested by previous literature (O'Regan, Rensink, & Clark, 1999; Rensink, ORegan, & Clark, 1997; Sampanes, Tseng, & Bridgeman, 2008; Shore & Klein, 2000; Stirk & Underwood, 2007). Sensory-driven information such as contrast-energy also influences our ability to detect changes. Most important, we find that the knowledge-based and sensory-driven information interacts, which suggests that these sources of information are not independently represented in the visual system.

This is supported by our results of Chapter 4 and 5. In Chapter 4 we show that similar images were harder to identify than other images using the fMRI responses of visual area V1. We suggested that this is caused by the internal representation of V1 not only capturing sensory-driven information, but also knowledge-based information. This is affirmed by the study in Chapter 5 where we show that the sensory-driven information (defined by the CRF) of early visual areas is enhanced according to the knowledge-based information (defined by subjective importance).

We can relate our results of the interaction found in early visual areas with our results of change detection. We show that the internal representation of V1 is modulated by the interaction of sensory-driven and knowledge-based information. This interaction is also reflected in our perceptual functioning, and more specifically in our ability to detect changes in a visual scene.

Neural and behavior interactions

Our result of the enhanced CRFs for high subjective importance is in line with the theory of efficient coding that proposes an enhancing role for feedback according to the perceptual hypothesis (Barlow, 2001; Friston, 2002; Rao, 2005; Series, Lorenceau, & Fregnac, 2003; Simoncelli & Olshausen, 2001). What would be the functional role of this feedback from higher visual areas? What is the purpose of increased responses in early visual areas to subjectively important information?

The efficient coding mechanism makes predictions about the external sources by disambiguating the neural signals. By making these predictions, the visual system can extract the behaviorally relevant information from the neural signal; the irrelevant activity is reduced, while the responses that are relevant for the visual percept are enhanced. This process is hypothesized to cause the information to be encoded in a sparser way, and this compression of information in the neural signal can be a way for the brain to deal with its capacity limits.

We find a similar interaction for contrast-energy and subjective importance reflected in the neural signal as we do in the performance of change detection. Both an enhanced neural signal and faster reaction times in change detection are found for high levels of contrast-energy and subjective importance.

The enhanced performance for the detection of changes that are high in contrast and/or high in subjective importance can be explained as an epiphenomenon of the enhanced neural signals in early visual areas for similar image features, and serve no further functional role.

However, the similarity of the effects that we find in the enhanced neural signals and enhanced detection performance for resembling image information can also be related to the view of V1 as a 'multiscale cognitive blackboard' (Roelfsema & de Lange, 2016). This view hypothesizes that the enhanced responses found in V1 are related to cognitive processes such as attention, where attention is directed to those parts of a scene that are encoded by an enhanced response in early visual areas. This view can explain why the changes of high contrast and/or high subjective importance were detected faster. We see an enhanced neural response to high levels of contrast and subjective importance of a scene. When an enhanced neural signal causes attention to be driven towards those parts of the scene that elicit this enhanced response, this might lead to a faster detection of changes that are high in contrast and/or high subjective importance.

Where Roelfsema & de Lange base their view on results from studies using synthetic stimuli and relate this to the performance of attention-based tasks, we now extrapolate this view to the use of natural images. Furthermore, the enhanced neural responses to contrast and subjective importance were found in the responses of subjects viewing natural images without a stimulus-related task.

We interpret the enhanced responses to subjective importance as the result of interactions between early and later visual areas. Possibly, the interaction between early and later visual areas is a mechanism to drive attention to potentially important parts of a scene. I suggest that feedback from later visual areas enhances the sensitivity for parts of the image of high subjective importance by modulating the responses in early visual areas. This boosts the CRF in early visual areas according to subjective importance, which causes smaller contrast changes to be necessary to detect differences in the scene for these regions of the image.

Synthetic versus natural images

In this thesis, we studied both stimulus-driven and knowledge-based contextual influences in the visual system. For this, we used both synthetic and natural images. The use of synthetic stimuli is very common when studying vision in the laboratory. The reason for this is that they are easily controlled. The drawback of using natural images is that they are difficult to control and they might contain an unlimited range of image properties. Although synthetic stimuli are easy to control, the drawback of using synthetic stimuli is that the responses to these stimuli do not always extrapolate to natural scenes (Carandini et al., 2005; David, Vinje, & Gallant, 2004; Kayser, Kording, & Konig, 2004; Li, VanRullen, Koch, & Perona, 2002). The synthetic images lack an important amount of knowledge-based contextual information as we encounter this in real life. The functioning of the visual system is mainly to solve contextual ambiguities of natural scenes. The use of natural images possibly drives a different population of neurons with neuronal properties that are necessary to disambiguate natural scenes. Therefore, there is a need to use more naturalistic stimuli in experiments.

A strength of the pRF-model is that it allows the selection of voxels that are stimulated by different parts of the stimulus. Where conventional fMRI studies ignore the possibility that different properties of the stimulus have deviating effects on the neural signal, we can investigate these effects with the pRF-model. This is especially useful for studying the neural responses to natural images. The pRF-model allows us to investigate the effects that different image properties have on the neural signal; these effects will be

different depending on the part of the image. Using the pRF-model we can select voxels responding to a certain part of the image and therefore select voxels that respond to specific image properties.

Where we investigated how the neural signal is affected by contrast-energy and subjective importance, future studies can disentangle the effects of other image properties using the presented methods. Using the newly developed analysis techniques we can study both synthetic and natural images. This approach introduces a new opportunity to extrapolate results from synthetic to natural stimuli.

Appendix

Supplementary material & figures

Supplementary material

Neuronal position scatter

With fMRI we cannot measure the activation of single neurons, instead activation of a population of neurons is measured. The size of the neuronal sampling population is dependent on the voxel size, and thus dependent on the resolution of fMRI.

The pRF is estimated from the total neuronal population, so eventually the properties of all the individual neurons influence the estimated pRF. In such a neuronal population, the position scatter of the individual neurons can vary (Hubel & Wiesel, 1974). The size of the estimated pRF has been reported to vary for both this neuronal position scatter, $\sigma^2_{position\_variance}$, as well as with the average receptive field size of the neuronal population, $\sigma^2_{neuronal\_RF}$ (Dumoulin & Wandell, 2008).

$$\sigma^2_{\text{population RF}} = \sigma^2_{\text{neuronal\_RF}} + \sigma^2_{\text{position\_variance}} + k \quad (11)$$

Where k is a constant factor for capturing non-neural contributions to the pRF.
A higher variance in the neuronal positions will increase the pRF size, despite unchanged neuronal RF sizes. The assumption we examine here is that the neuronal position scatter will not only influence the estimated pRF size, but will also influence the ability to measure center-surround configurations of the pRFs. Figure 9 illustrates this idea. When the position variance of a neuronal population is low (Fig. 9A), there is a clear circular receptive field noticeable as a pRF, which still shows a center-surround organization. With high position variance (Fig. 9B) this center-surround organization seems to have disappeared. Using formula 11, we calculated the pRF from a total of 100,000 neuronal receptive fields. While keeping the neuronal RF constant we varied position variance (and k=0). Figure 9C shows a representation of the pRFs that are calculated for neuronal populations with different position variances. Increasing position variance leads to a decrease in center-surround configuration of the pRF. Specifically, the amplitude of the negative surround becomes less pronounced (Figure 9D).

This simulation illustrates that the neuronal position scatter will not only affect the pRF size, but also the strength of the negative component of the pRF. For visual areas with a high neuronal position variance, the ability to measure center-surround configurations is therefore lost at the resolution of fMRI.

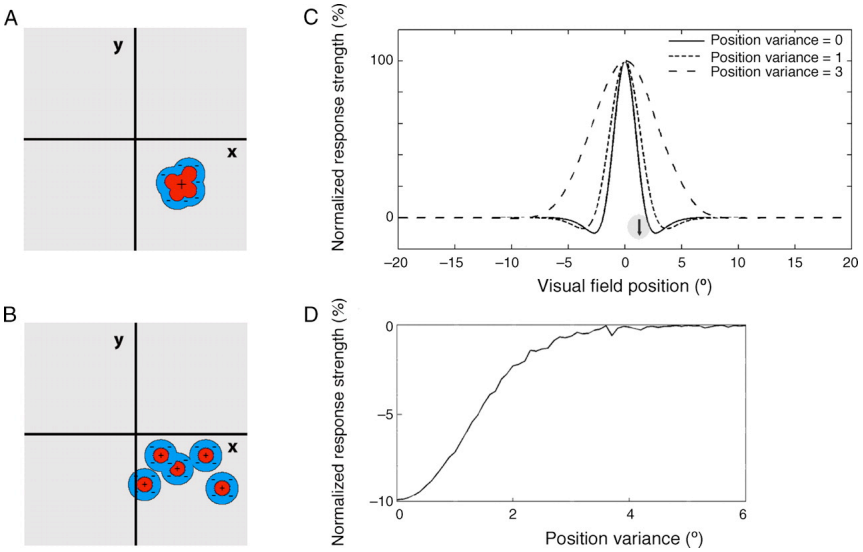
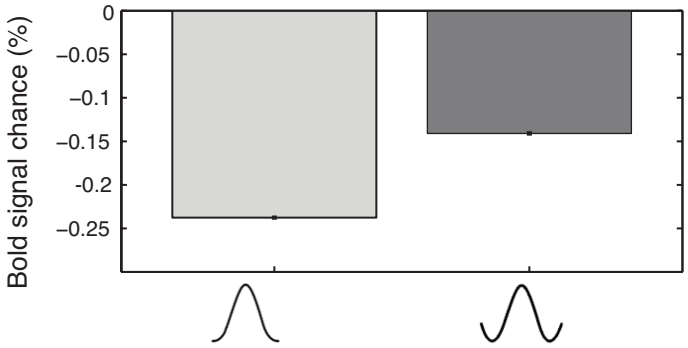


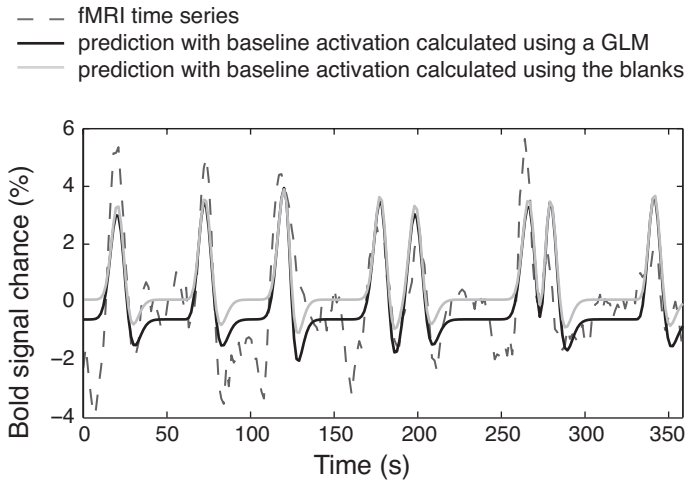
Figure 9 Simulations estimating the pRF from a population of individual neuronal receptive fields (RF). The pRFs are calculated from 100,000 individual neuronal receptive fields using equation 11. Individual RFs are identical each with their own center-surround configuration. Panels A and B illustrate a few individual RFs with low (A) and high (B) position variance. The actual simulations used a 100,000 RFs. These individual RFs were summed to give rise to the pRF (C). For the neuronal population with a low variance in its positions (A) the neuronal RF center-surround configuration is reflected in its pRF. However, the pRF of the neuronal population with high position scatter (B) has lost the center-surround configuration at the pRF level. The relationship between the position variance of a neuronal population to the response strength of the suppressive surround of the estimated pRF (arrow in panel C) is shown in panel (D). For a neuronal population with a high position variance the response strength of the suppressive surround approaches zero. These simulations suggest the ability to measure center-surround configurations is lost in neuronal populations with a high neuronal position scatter.

Supplementary figures

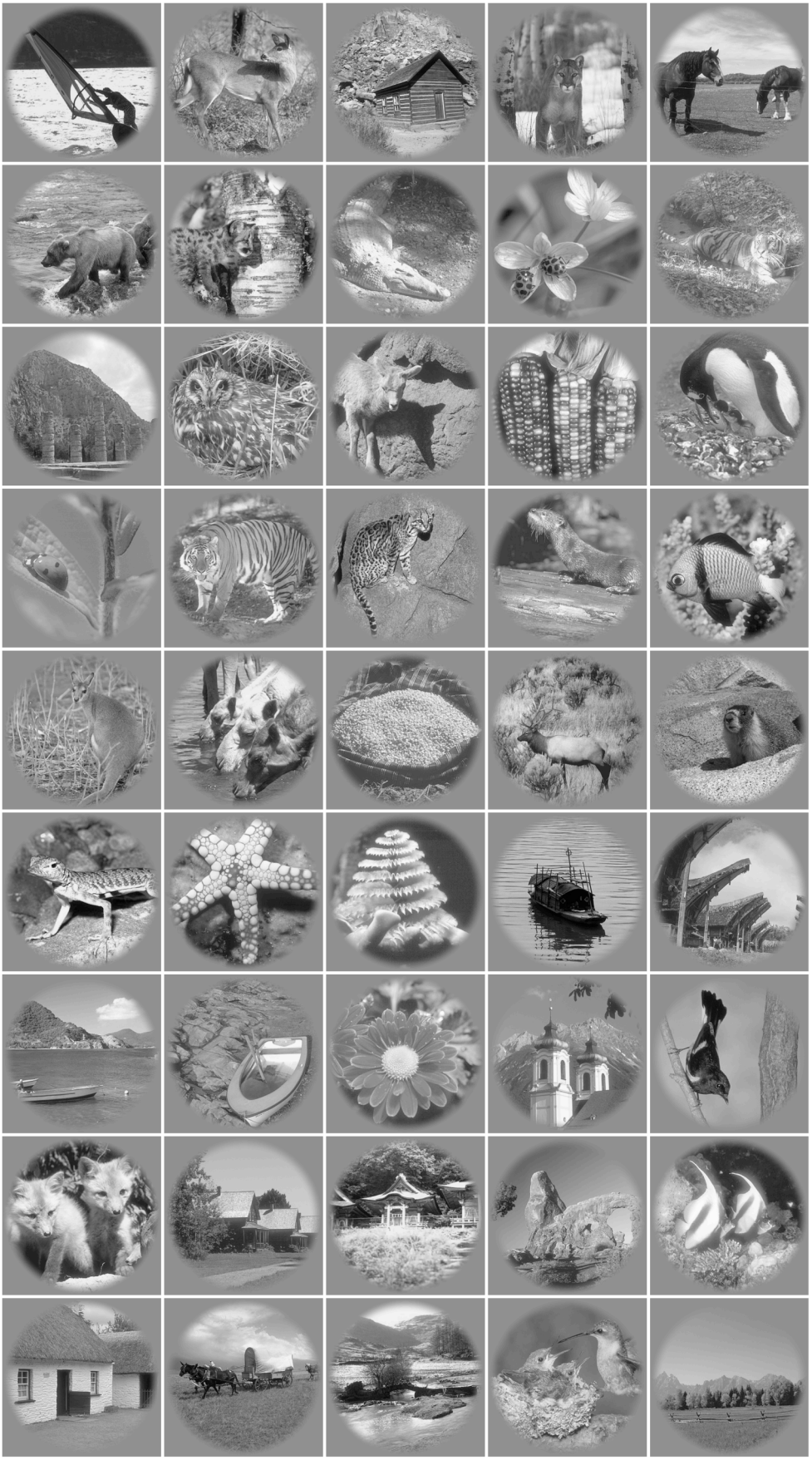
A



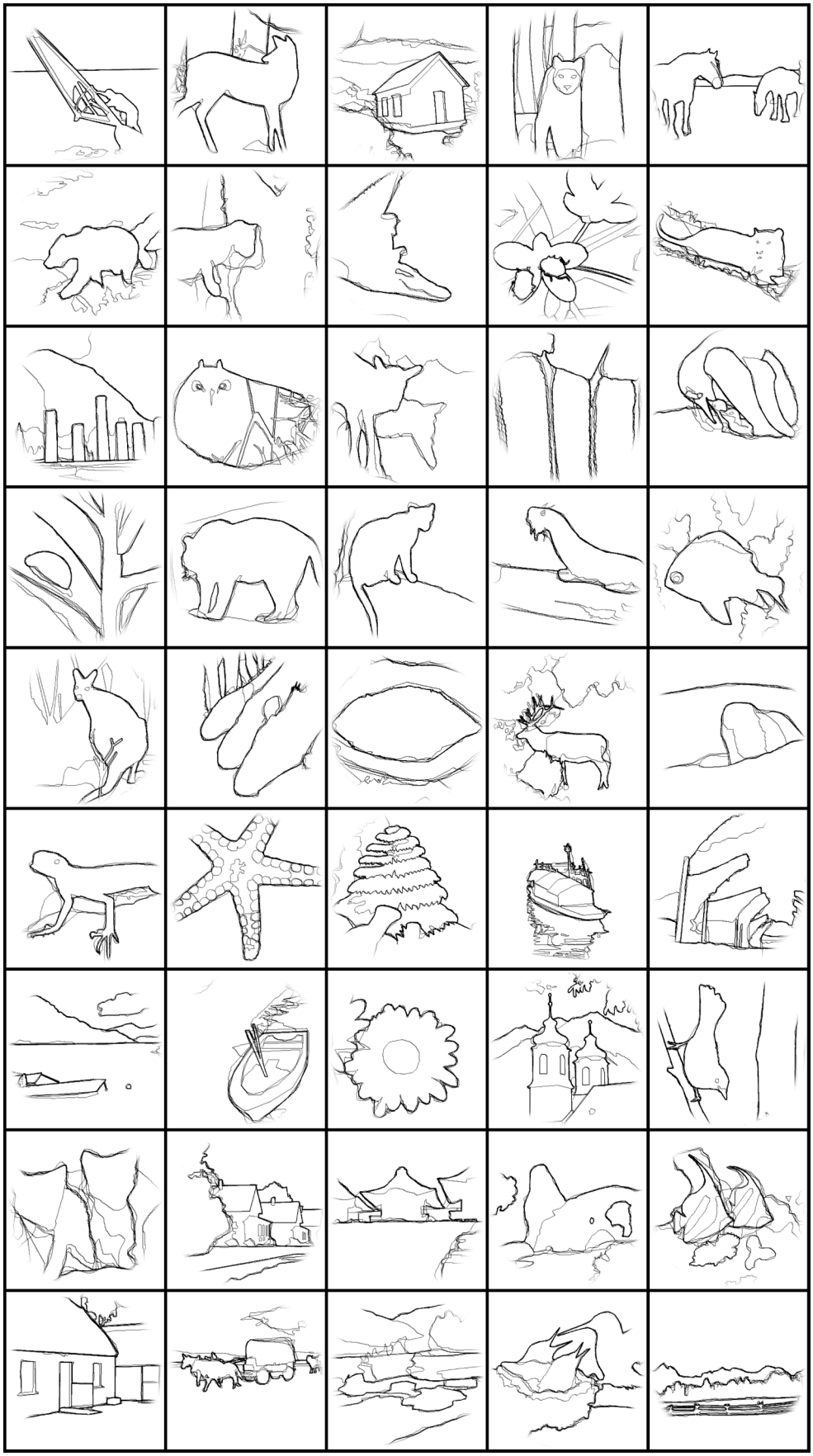
B



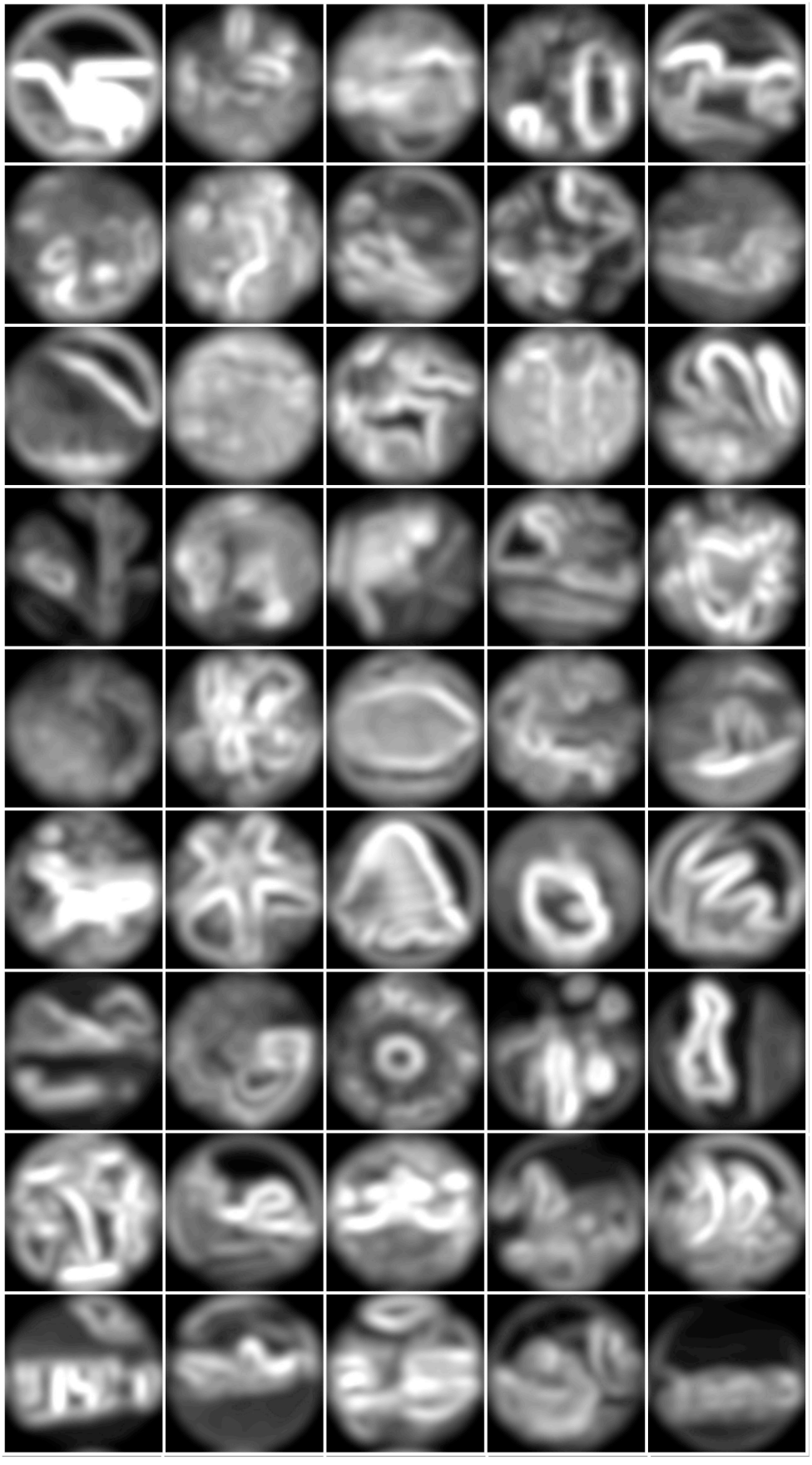
Supplementary figure 1 Differences in baseline estimation. A: The conventional OG-model estimates the parameters of the baseline activation during the fit of the pRF, according to the whole time-series. Comparing the baseline activity measurements for the OG-model and the DoG-model, we find different estimates for this baseline activation, which are measured to the same time-series. These data are taken from V1 of all subjects. The errorbar represents standard error of mean. This suggests the OG-model is able to compensate for the negative fMRI signal by lowering its baseline activation. B: An example time-series from a recording site in V1 (dashed lines) with the predictions of the conventional OG-model where the baseline activation is estimated using the two ways. First, the baseline is calculated during the GLM fit (black line); the second method calculates the parameters for the baseline using the time-series where no stimulus was shown (gray line). This example illustrates that the conventional OG-model is able to compensate for the negative fMRI signal to some extent by lowering the baseline level.



Supplementary figure 2 The natural images that were used as stimuli in the experiment. The images came from the Berkeley Segmentation Dataset and Benchmark database (D. Martin et al., 2001).



Supplementary figure 3 For all the images in the Berkeley Segmentation Dataset and Benchmark database there were manual labels available (D. Martin et al., 2001). In each manually labeled image, human observers identified the most important aspects of the image. We used these measurements of the subjective important aspects of the scene to define and quantify the knowledge-based perceptual hypothesis of the image. The average of 5 manually labeled images was used as our definition of the perceptual hypothesis.



Supplementary figure 4 For the sensory-driven image information we calculated the local contrast-energy by the RMS-contrast.

Appendix

References

References

Adelson, E. H. (1995). http://persci.mit.edu/_media/gallery/checkershadow_illusion4med.jpg.

Adelson, E. H. (2000). Lightness perception and lightness illusions *The New Cognitieve Neurosciences, 2nd ed (ed. M. Gazzaniga), MIT press, Cambridge*, 339-351.

Albright, T. D., & Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale, 65*(3), 582-592.

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of neurophysiology, 48*(1), 217-237.

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 30*(8), 2960-2966.

Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci, 8*, 407-430.

Anderson, E. J., Tibber, M. S., Schwarzkopf, D. S., Shergill, S. S., Fernandez-Egea, E., Rees, G., et al. (2017). Visual Population Receptive Fields in People with Schizophrenia Have Reduced Inhibitory Surrounds. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 37*(6), 1546-1556.

Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in brain research, 154*, 93-120.

Angelucci, A., Levitt, J. B., Walton, E. J., Hupe, J. M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 22*(19), 8633-8646.

Amano, K., Wandell, B. A., & Dumoulin, S. O. (2009). Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *Journal of neurophysiology, 102*(5), 2704-2718.

Barlow, H. (2001). Redundancy reduction revisited. *Network, 12*(3), 241-253.

Baseler, H. A., Gouws, A., Haak, K. V., Racey, C., Crossland, M. D., Tufail, A., et al. (2011). Large-scale remapping of visual cortex is absent in adult humans with macular degeneration. *Nat Neurosci, 14*(5), 649-655.

Beck, D. M., & Kastner, S. (2007). Stimulus similarity modulates competitive interactions in human visual cortex. *Journal of vision, 7*(2), 19 11-12.

Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America. A, Optics, image science, and vision, 19*(6), 1096-1106.

Birn, R. M., Saad, Z. S., & Bandettini, P. A. (2001). Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. *Neuroimage, 14*(4), 817-826.

Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision research, 39*(2), 257-269.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci, 16*(13), 4207-4221.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Brewer, A. A., & Barton, B. (2014). Visual cortex in aging and Alzheimer's disease: changes in visual field maps and population receptive fields. *Frontiers in psychology, 5*, 74.

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 29*(44), 13992-14003.

Butler, P. D., Silverstein, S. M., & Dakin, S. C. (2008). Visual perception and its impairment in schizophrenia. *Biol Psychiatry, 64*(1), 40-47.

Carandini, M. (2004). Receptive fields and suppressive fields in the early visual system. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (3 ed., pp. 312-326). Cambridge, MA: MIT Pres.

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., et al. (2005). Do we know what the early visual system does? *The Journal of neuroscience : the official journal of the Society for Neuroscience, 25*(46), 10577-10597.

Carandini, M., & Heeger, D. J. (2011). Normalization as a canonical neural computation. *Nature reviews. Neuroscience, 13*(1), 51-62.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J Neurophysiol, 88*(5), 2530-2546.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage, 19*(2 Pt 1), 261-270.

Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron, 48*(3), 497-507.

David, S. V., Vinje, W. E., & Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 24*(31), 6991-7006.

De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York ; Oxford: Oxford University Press.

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

den Ouden, H. E., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral cortex, 19*(5), 1175-1185.

DeSimone, K., Viviano, J. D., & Schneider, K. A. (2015a). Population Receptive Field Estimation Reveals New Retinotopic Maps in Human Subcortex. *J Neurosci, 35*(27), 9836-9847.

DeSimone, K., Viviano, J. D., & Schneider, K. A. (2015b). Population Receptive Field Estimation Reveals New Retinotopic Maps in Human Subcortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 35*(27), 9836-9847.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., et al. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc Natl Acad Sci U S A, 93*(6), 2382-2386.

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., & Wandell, B. A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *Journal of Vision, 3*(10), 586-598.

Dow, B. M., Snyder, A. Z., Vautin, R. G., & Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of the monkey. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale, 44*(2), 213-228.

Dumoulin, S. O., Dakin, S. C., & Hess, R. F. (2008). Sparsely distributed contours dominate extra-striate responses to complex scenes. *Neuroimage, 42*(2), 890-901.

Dumoulin, S. O., & Hess, R. F. (2006). Modulation of V1 activity by shape: image-statistics or shape-based perception? *Journal of neurophysiology, 95*(6), 3654-3664.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage, 39*(2), 647–660.

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex, 7*(2), 181-192.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience, 4*, 215.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex, 1*(1), 1-47.

Fiorani, M., Jr., Gattass, R., Rosa, M. G., & Sousa, A. P. (1989). Visual area MT in the Cebus monkey: location, visuotopic organization, and variability. *The Journal of comparative neurology, 287*(1), 98-118.

Fitzpatrick, D. (2000). Seeing beyond the receptive field in primary visual cortex. *Curr Opin Neurobiol, 10*(4), 438-443.

Freeman, J., Brouwer, G. J., Heeger, D. J., & Merriam, E. P. (2011). Orientation decoding depends on maps, not columns. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 31*(13), 4792-4804.

Friston, K. (2002). Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annual review of neuroscience, 25*, 221-250.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 360*(1456), 815-836.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. *Neuroimage, 7*(1), 30-40.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., et al. (1995). Analysis of fMRI time-series revisited. *Neuroimage, 2*(1), 45-53.

Gattass, R., & Gross, C. G. (1981). Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque. *Journal of neurophysiology, 46*(3), 621-638.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage, 9*(4), 416-429.

Hansen, K. A., David, S. V., & Gallant, J. L. (2004). Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. *Neuroimage, 23*(1), 233-241.

Harel, N., Lee, S. P., Nagaoka, T., Kim, D. S., & Kim, S. G. (2002). Origin of negative blood oxygenation level-dependent fMRI signals. *J Cereb Blood Flow Metab, 22*(8), 908-917.

Harrison, L. M., Penny, W., Ashburner, J., Trujillo-Barreto, N., & Friston, K. J. (2007). Diffusion-based spatial priors for imaging. *Neuroimage, 38*(4), 677-695.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458*(7238), 632-635.

Harvey, B. M., & Dumoulin, S. O. (2011). The Relationship between Cortical Magnification Factor and Population Receptive Field Size in Human Visual Cortex: Constancies in Cortical Architecture. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 31*(38), 13604-13612.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature neuroscience, 8*(5), 686-691.

He, D., Mo, C., Wang, Y., & Fang, F. (2015). Position shifts of fMRI-based population receptive fields in human visual cortex induced by Ponzo illusion. *Experimental brain research, 233*(12), 3535-3541.

Hetherington, P. A., & Swindale, N. V. (1999). Receptive field and orientation scatter studied by tetrode recordings in cat area 17. *Visual neuroscience, 16*(4), 637-652.

Henriksson, L., Nurminen, L., Hyvarinen, A., & Vanni, S. (2008). Spatial frequency tuning in human retinotopic visual areas. *Journal of Vision, 8*(10), 5 1-13.

Hoffmann, M. B., Kaule, F. R., Levin, N., Masuda, Y., Kumar, A., Gottlob, I., et al. (2012). Plasticity and stability of the visual system in human achiasma. *Neuron, 75*(3), 393-401.

Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science, 340*(6132), 639-642.

Hubel, D. H. (1982). Exploration of the primary visual cortex, 1955-78. *Nature, 299*(5883), 515-524.

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

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol, 195*(1), 215-243.

Hubel, D. H., & Wiesel, T. N. (1974). Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *The Journal of comparative neurology, 158*(3), 295-305.

Hummer, A., Ritter, M., Tik, M., Ledolter, A. A., Woletz, M., Holder, G. E., et al. (2016). Eyetracker-based gaze correction for robust mapping of population receptive fields. *Neuroimage, 142*, 211-224.

Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature, 394*(6695), 784-787.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research, 40*(10-12), 1489-1506.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature neuroscience, 8*(5), 679-685.

Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: implications for receptive field sizes in the human visual cortex. *Journal of neurophysiology, 86*(3), 1398-1411.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*(4), 751-761.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452*(7185), 352-355.

Kay, K. N., Winawer, J., Rokem, A., Mezer, A., & Wandell, B. A. (2013). A two-stage cascade model of BOLD responses in human visual cortex. *PLoS computational biology, 9*(5), e1003079.

Kayser, C., Kording, K. P., & Konig, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Current opinion in neurobiology, 14*(4), 468-473.

Koch, C., & Poggio, T. (1999). Predicting the visual world: silence is golden. *Nature neuroscience, 2*(1), 9-10.

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current biology : CB, 26*(3), 371-376.

Kok, P., & de Lange, F. P. (2014). Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Current biology : CB, 24*(13), 1531-1535.

Kok, P., Jehee, J. F., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265-270.

Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & de Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral cortex, 22*(9), 2197-2206.

Kriegeskorte, N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual review of vision science, 1*, 417-446.

Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 15*(2), 1605-1615.

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

Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research, 43*(2), 149-164.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision research, 38*(15-16), 2429-2454.

Legge, G. E. (1981). A power law for contrast discrimination. *Vision research, 21*(4), 457-467.

Levitt, J. B., & Lund, J. S. (2002). The spatial extent over which neurons in macaque striate cortex pool visual signals. *Visual neuroscience, 19*(4), 439-452.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99*(14), 9596-9601.

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual review of physiology, 66*, 735-769.

Mante, V., & Carandini, M. (2005). Mapping of stimulus energy in primary visual cortex. *Journal of neurophysiology, 94*(1), 788-798.

Marin, O. (2012). Interneuron dysfunction in psychiatric disorders. *Nature reviews. Neuroscience, 13*(2), 107-120.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proceedings Eighth International Conference on Computer Vision, Vol II, Proceedings*, 416-423.

Martin, D. R., Fowlkes, C. C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE transactions on pattern analysis and machine intelligence, 26*(5), 530-549.

Mazza, V., Turatto, M., & Umilta, C. (2005). Foreground-background segmentation and attention: A change blindness study. *Psychological Research-Psychologische Forschung, 69*(3), 201-210.

McDermott, J. (2004). Psychophysics with junctions in real images. *Perception, 33*(9), 1101-1127.

McDonald, J. S., Seymour, K. J., Schira, M. M., Spehar, B., & Clifford, C. W. (2009). Orientation-specific contextual modulation of the fMRI BOLD response to luminance and chromatic gratings in human visual cortex. *Vision research, 49*(11), 1397-1405.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive psychology, 14*(2), 247-279.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & psychophysics, 40*(5), 331-343.

Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., et al. (2004). Learning to decode cognitive states from brain images. *Machine Learning, 57*(1), 145-175.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science, 320*(5880), 1191-1195.

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M. A., Morito, Y., Tanabe, H. C., et al. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron, 60*(5), 915-929.

Moseley, M. E., & Glover, G. H. (1995). Functional MR imaging. Capabilities and limitations. *Neuroimaging clinics of North America, 5*(2), 161-191.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological cybernetics, 66*(3), 241-251.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 99*(23), 15164-15169.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron, 63*(6), 902-915.

Nestares, O., & Heeger, D. J. (2000). Robust multiresolution alignment of MRI brain volumes. *Magn Reson Med, 43*(5), 705-715.

Nurminen, L., Kilpelainen, M., Laurinen, P., & Vanni, S. (2009). Area summation in human visual system: psychophysics, fMRI, and modeling. *Journal of neurophysiology, 102*(5), 2900-2909.

O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of 'mudsplashes'. *Nature, 398*(6722), 34.

Ohzawa, I., Sclar, G., & Freeman, R. D. (1985). Contrast gain control in the cat's visual system. *Journal of neurophysiology, 54*(3), 651-667.

Olman, C. A., Ugurbil, K., Schrater, P., & Kersten, D. (2004). BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision research, 44*(7), 669-683.

Papanikolaou, A., Keliris, G. A., Papageorgiou, T. D., Shao, Y., Krapp, E., Papageorgiou, E., et al. (2014a). Population receptive field analysis of the primary visual cortex complements perimetry in patients with homonymous visual field defects. *Proc Natl Acad Sci U S A, 111*(16), E1656-1665.

Papanikolaou, A., Keliris, G. A., Papageorgiou, T. D., Shao, Y., Krapp, E., Papageorgiou, E., et al. (2014b). Population receptive field analysis of the primary visual cortex complements perimetry in patients with homonymous visual field defects. *Proceedings of the National Academy of Sciences of the United States of America, 111*(16), E1656-1665.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis, 10*(4), 437-442.

Petro, L. S., Vizioli, L., & Muckli, L. (2014). Contributions of cortical feedback to sensory processing in primary visual cortex. *Frontiers in psychology, 5*, 1223.

Petrov, Y., Carandini, M., & McKee, S. (2005). Two distinct mechanisms of suppression in human vision. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 25*(38), 8704-8707.

Petrov, Y., & McKee, S. P. (2006). The effect of spatial configuration on surround suppression of contrast sensitivity. *Journal of vision, 6*(3), 224-238.

Press, W. A., Brewer, A. A., Dougherty, R. F., Wade, A. R., & Wandell, B. A. (2001). Visual areas and spatial summation in human visual cortex. *Vision research, 41*(10-11), 1321-1332.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences, 24*, 574-590.

Rao, R. P. (2005). Bayesian inference and attentional modulation in the visual cortex. *Neuroreport, 16*(16), 1843-1848.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience, 2*(1), 79-87.

Rauss, K., Schwartz, S., & Pourtois, G. (2011). Top-down effects on early visual processing in humans: a predictive coding framework. *Neuroscience and biobehavioral reviews, 35*(5), 1237-1253.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision research, 40*(10-12), 1469-1487.

Rensink, R. A. (2002). Change detection. *Annual review of psychology, 53*, 245-277.

Rensink, R. A., ORegan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368-373.

Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision research, 5*(11), 583-601.

Roelfsema, P. R., & de Lange, F. P. (2016). Early visual cortex as a multiscale cognitive blackboard. *Annual review of vision science, 2*, 131-151.

Sampanes, A. C., Tseng, P., & Bridgeman, B. (2008). The role of gist in scene recognition. *Vision Research, 48*(21), 2275-2283.

Sceniak, M. P., Hawken, M. J., & Shapley, R. (2001). Visual spatial characterization of macaque V1 neurons. *J Neurophysiol, 85*(5), 1873-1887.

Sceniak, M. P., Ringach, D. L., Hawken, M. J., & Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nature neuroscience, 2*(8), 733-739.

Schwarzkopf, D. S., Anderson, E. J., de Haas, B., White, S. J., & Rees, G. (2014a). Larger extrastriate population receptive fields in autism spectrum disorders. *J Neurosci, 34*(7), 2713-2724.

Schwarzkopf, D. S., Anderson, E. J., de Haas, B., White, S. J., & Rees, G. (2014b). Larger extrastriate population receptive fields in autism spectrum disorders. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 34*(7), 2713-2724.

Sclar, G., Maunsell, J. H., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision research, 30*(1), 1-10.

Self, M. W., Mookhoek, A., Tjalma, N., & Roelfsema, P. R. (2015). Contextual effects on perceived contrast: Figure-ground assignment and orientation contrast. *Journal of Vision, 15*(2).

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268*(5212), 889-893.

Series, P., Lorenceau, J., & Fregnac, Y. (2003). The "silent" surround of V1 receptive fields: theory and experiments. *Journal of physiology, Paris, 97*(4-6), 453-474.

Shmuel, A., Augath, M., Oeltermann, A., & Logothetis, N. K. (2006). Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nat Neurosci, 9*(4), 569-577.

Shmuel, A., Yacoub, E., Pfeuffer, J., Van de Moortele, P. F., Adriany, G., Hu, X., et al. (2002). Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. *Neuron, 36*(6), 1195-1210.

Shore, D. I., & Klein, R. M. (2000). The effects of scene inversion on change blindness. *The Journal of general psychology, 127*(1), 27-43.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual review of neuroscience, 24*, 1193-1216.

Simons, D. J., & Ambinder, M. S. (2005). Change blindness - Theory and consequences. *Current Directions in Psychological Science, 14*(1), 44-48.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends in cognitive sciences, 9*(1), 16-20.

Singh, K. D., Smith, A. T., & Greenlee, M. W. (2000). Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage, 12*(5), 550-564.

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral cortex, 11*(12), 1182-1190.

Smith, A. T., Williams, A. L., & Singh, K. D. (2004). Negative BOLD in the visual cortex: evidence against blood stealing. *Hum Brain Mapp, 21*(4), 213-220.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage, 23 Suppl 1*, S208-219.

Spratling, M. W. (2008). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in computational neuroscience, 2*, 4.

Stirk, J. A., & Underwood, G. (2007). Low-level visual saliency does not predict change detection in natural scenes. *Journal of vision, 7*(10), 3 1-10.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in cognitive sciences, 13*(9), 403-409.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature neuroscience, 11*(9), 1004-1006.

Tajima, S., Watanabe, M., Imai, C., Ueno, K., Asamizuya, T., Sun, P., et al. (2010). Opposing effects of contextual surround in human early visual cortex revealed by functional magnetic resonance imaging with continuously modulated visual stimuli. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 30*(9), 3264-3270.

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J. B., Lebihan, D., et al. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage, 33*(4), 1104-1116.

Thomas, J. M., Huber, E., Stecker, G. C., Boynton, G. M., Saenz, M., & Fine, I. (2015). Population receptive field estimates of human auditory cortex. *Neuroimage, 105*, 428-439.

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual review of psychology, 63*, 483-509.

Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Liu, A. K., & Dale, A. M. (1998). The representation of the ipsilateral visual field in human cerebral cortex. *Proc Natl Acad Sci U S A, 95*(3), 818-824.

Ulrich, R., Miller, J., & Schroter, H. (2007). Testing the race model inequality: an algorithm and computer programs. *Behavior research methods, 39*(2), 291-302.

Verma, M., & McOwan, P. W. (2010). A semi-automated approach to balancing of bottom-up salience for predicting change detection performance. *Journal of vision, 10*(6), 3.

Victor, J. D., Purpura, K., Katz, E., & Mao, B. (1994). Population encoding of spatial frequency, orientation, and color in macaque V1. *J Neurophysiol, 72*(5), 2151-2166.

Wade, A. R., & Rowland, J. (2010). Early suppressive mechanisms and the negative blood oxygenation level-dependent response in human visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 30*(14), 5008-5019.

Wandell, B. A. (1999). Computational neuroimaging of human visual cortex. *Annual review of neuroscience, 22*, 145-173.

Wandell, B. A., Brewer, A. A., & Dougherty, R. F. (2005). Visual field map clusters in human cortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 360*(1456), 693-707.

Wandell, B. A., Chial, S., & Backus, B. T. (2000). Visualization and measurement of the cortical surface. *J Cogn Neurosci, 12*(5), 739-752.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron, 56*(2), 366-383.

Wandell, B. A., Winawer, J., & Kay, K. N. (2015). Computational modeling of responses in human visual cortex. *Acquisition Methods, Methods and Modeling, 1*, 651-659.

Williams, A. L., Singh, K. D., & Smith, A. T. (2003). Surround modulation measured with functional MRI in the human visual cortex. *Journal of neurophysiology, 89*(1), 525-533.

Winawer, J., Horiguchi, H., Sayres, R. A., Amano, K., & Wandell, B. A. (2010). Mapping hV4 and ventral occipital cortex: the venous eclipse. *J Vis, 10*, 1-22.

Xing, Y., Ledgeway, T., McGraw, P. V., & Schluppeck, D. (2013). Decoding working memory of stimulus contrast in early visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 33*(25), 10301-10311.

Yoon, J. H., Rokem, A. S., Silver, M. A., Minzenberg, M. J., Ursu, S., Ragland, J. D., et al. (2009). Diminished orientation-specific surround suppression of visual processing in schizophrenia. *Schizophr Bull, 35*(6), 1078-1084.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., et al. (2006). User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage, 31*(3), 1116-1128.

Zenger-Landolt, B., & Heeger, D. J. (2003). Response suppression in v1 agrees with psychophysics of surround masking. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 23*(17), 6884-6893.

Zuiderbaan, W., Harvey, B. M., & Dumoulin, S. O. (2012). Modeling center-surround configurations in population receptive fields using fMRI. *Journal of Vision, 12*(3), 10.

Appendix

Nederlandse samenvatting

Nederlandse samenvatting

Zicht is een van de belangrijke zintuigen die ons in staat stellen om onze omgeving waar te kunnen nemen en hiermee een interactie te hebben. Wanneer we rondkijken in de wereld zien we veel verschillende objecten en vormen. Voor onze interactie met de wereld om ons heen, is het nodig dat we objecten kunnen zoeken, vinden en herkennen. Hiervoor moet het visuele systeem objecten kunnen segmenteren en detecteren.

Ons brein krijgt via onze ogen een enorm grote hoeveelheid aan informatie binnen. Aangezien de informatie die vanuit onze ogen komt, ruizig en ambigu is, maakt het visuele systeem van deze informatie een interpretatie. Voor het maken van deze interpretatie worden niet alleen de zintuiglijke eigenschappen van de stimulus gebruikt, maar ook contextuele informatie die bestaat uit onze kennis van de wereld. Met behulp van deze voorkennis voorspelt het visuele systeem wat de meest waarschijnlijke oorzaak is van het beeld op het netvlies.

Hoe het beeld van het netvlies in onze hersenen wordt verwerkt en hoe dit ons gedrag beïnvloedt, is een belangrijke vraag in de neurowetenschappen. In dit proefschrift is onderzocht wat de contextuele invloed van zintuiglijke en kennisgebaseerde eigenschappen is op zowel ons gedrag (*hoofdstuk 2*) als de neurale stimulus representatie van de visuele cortex (*hoofdstuk 3, 4 en 5*). Hiervoor hebben we een combinatie van gedragsmethoden en computationele neuroimaging technieken gebruikt.

In de experimenten beschreven in dit proefschrift hebben we zowel synthetische als natuurlijke afbeeldingen gebruikt. Waar de synthetische afbeeldingen voornamelijk zintuiglijke informatie bevatten, bevatten de natuurlijke afbeeldingen zowel zintuiglijke als op kennis gebaseerde informatie. Van de natuurlijke afbeeldingen kwantificeren we zowel de lokale contrast-energie (zintuiglijke informatie) als de subjectief belangrijke informatie (kennisgebaseerde informatie).

In *hoofdstuk 2* laten we zien dat er een interactie van zintuiglijke en kennisgebaseerde informatie terug te vinden is in ons gedrag. Wanneer we om ons heen kijken in de wereld, hebben we de ervaring dat we onze omgeving waarnemen met veel details, zoals een camera. Echter, vaak merken we grote veranderingen in onze omgeving

helemaal niet op. Dit onvermogen wordt ook wel veranderingsblindheid genoemd. Dit verschijnsel illustreert dat niet elk deel van de visuele scene zo gedetailleerd wordt verwerkt als we intuïtief ervaren. Het visuele systeem moet van de grote hoeveelheid informatie die het ontvangt via onze ogen een coherent beeld maken. De hersenen hebben slechts een beperkte verwerkingscapaciteit, en hierdoor kan niet alle informatie van de visuele scene in detail verwerkt worden. Hierbij zien we dat sommige veranderingen gemakkelijker op te sporen zijn dan andere.

In het tweede hoofdstuk onderzoeken we wat de invloed van zintuiglijke informatie (lokale contrast-energie) en kennisgebaseerde informatie (subjectief belangrijk) in natuurlijke afbeeldingen is op ons vermogen om veranderingen te detecteren. We meten reactietijden, terwijl we veranderingen maken in zowel de zintuiglijke als de kennisgebaseerde informatie van de afbeelding. Onze resultaten laten zien dat de mogelijkheid om veranderingen in een visuele scene te detecteren, zowel door de zintuiglijke als de kennisgebaseerde informatie wordt beïnvloed. Bovendien vinden we een interactie; manipulaties in een visuele scene, die zowel groot zijn in de verandering van zintuiglijke als kennisgebaseerde informatie worden het snelst gevonden. Deze resultaten suggereren dat de zintuiglijke en kennisgebaseerde informatie niet onafhankelijk van elkaar worden verwerkt in het visuele systeem.

In *hoofdstuk 3, 4 en 5* hebben we vervolgens onderzocht hoe de interactie tussen de zintuiglijke en kennisgebaseerde informatie van de stimulus is gerepresenteerd in het neurale signaal van de visuele cortex. Hiervoor hebben we gebruik gemaakt van functionele MRI. Met behulp van fMRI kunnen we hersenactiviteit meten in reactie op visuele stimuli. We kunnen hierdoor onderzoeken hoe de visuele cortex reageert op verschillende visuele prikkels.

Neuronen in de visuele cortex reageren op een bepaald gebied in het gezichtsveld, dit wordt ook wel het receptieve veld van dit neuron genoemd. Bij het stimuleren van dit deel van het gezichtsveld, wordt dit neuron geactiveerd. Met fMRI kunnen we per corticale locatie de regio in het visuele veld schatten waarop het reageert; het populatie receptieve veld (pRF) (Dumoulin & Wandell, 2008). In dit proefschrift presenteren we drie nieuwe analysetechnieken die gebruik maken van de pRF-analyse

om contextuele invloeden van zowel zintuiglijke als kennisgebaseerde informatie in de visuele cortex te meten.

Neuronen in de vroege visuele cortex hebben kleine receptieve velden en zullen daarom alleen informatie van een klein gedeelte van de visuele scene bevatten. De informatie van neuronen met kleine receptieve velden zal hierdoor veel ambiguïteiten bevatten. Een manier om te disambigueren, is om de informatie in context te plaatsen. Dit kan bijvoorbeeld door de informatie te integreren over een groter gedeelte van het visuele veld. In de visuele cortex kan stimulatie buiten het klassieke receptieve veld de hersenactiviteit verminderen en een negatief fMRI signaal veroorzaken.

In *hoofdstuk 3* presenteren we een methode waarmee we laten zien dat neuronen in de vroege visuele cortex rekening houden met de context van lokale stimulatie. Hier breiden we het originele pRF-model uit door rekening te houden met stimulatie buiten het klassieke receptieve veld. Waar het originele model geen negatieve fMRI responsen kan verklaren, zien we dat dit nieuwe model dat wel kan. We laten zien dat dit nieuwe model voornamelijk in vroege visuele gebieden een betere voorspelling van het gemeten fMRI signaal geeft.

Verschillende klinische aandoeningen kunnen de eigenschappen van de pRF veranderen. Deze nieuwe methode geeft ons een directe maat van de eigenschappen van de pRFs in het visuele systeem, die nuttig zijn voor het bestuderen van zowel gezonde als klinische populaties. Een mooi voorbeeld hiervan is de studie van Anderson en collegae (Anderson et al., 2017), waarbij ons model is gebruikt om te laten zien dat er minder suppressie is buiten het klassieke receptieve veld bij mensen met schizofrenie.

Met behulp van het pRF-model kunnen we een voorspelling maken van de hersenactiviteit voor elke visuele stimulus. In *hoofdstuk 4* beschrijven we een fMRI studie waarin we het pRF-model gebruiken om de visueel gepresenteerde afbeelding te identificeren. We hebben zowel synthetische als natuurlijke afbeeldingen geïdentificeerd op basis van de zintuiglijke lokale contrast-energie informatie.

We laten zien dat het model kan worden gebruikt om afbeeldingen te identificeren, waarmee het model niet expliciet is getraind. Waar zowel de

synthetische als de natuurlijke afbeeldingen kunnen identificeren, was de prestatie minder goed voor de natuurlijke afbeeldingen.

Wat hierbij interessant is, is dat we zien dat (met behulp van de voxels van het eerste visuele gebied V1) bepaalde natuurlijke afbeeldingen moeilijker zijn te identificeren dan andere, en dat dit patroon vergelijkbaar is tussen verschillende proefpersonen. Wij denken dat een belangrijke reden hiervoor is, dat de natuurlijke afbeeldingen naast de zintuiglijke ook kennisgebaseerde informatie bevatten. Het pRF-model baseert zijn voorspelling alleen op de zintuiglijke contrast informatie, en wij zien dit als een aanwijzing dat de kennisgebaseerde informatie reeds aanwezig is in de interne representatie van V1.

In *hoofdstuk 5* hebben we verder onderzocht wat het effect is van de op kennis gebaseerde informatie in natuurlijke afbeeldingen op de interne stimulus representatie van vroege visuele gebieden. De interactie tussen zintuiglijke informatie en onze kennis van de wereld is de basis van perceptie, maar er is geen consensus over hoe deze interactie het neurale signaal in vroege visuele gebieden beïnvloedt. Huidige theorieën implementeren de kennis van de wereld als een perceptuele hypothese die wordt vergeleken met de zintuiglijke input. Volgens deze theorieën kan de perceptuele hypothese de zintuiglijke informatie die door vroege visuele gebieden gerepresenteerd word, onderdrukken of juist versterken. In deze studie vinden we dat de neurale respons op contrast-energie (zintuiglijk) in vroege visuele gebieden (V1, V2 en V3) wordt versterkt wanneer deze respons wordt veroorzaakt door een gedeelte van de afbeelding dat we als subjectief belangrijk beschouwen (kennisgebaseerd).

In dit proefschrift laten we zien dat perceptie zowel door zintuiglijke stimulatie als onze kennis van de wereld bepaald wordt. We hebben aangetoond dat de interactie van de zintuiglijke en kennisgebaseerde informatie zowel in ons gedrag als in het neurale signaal van vroege visuele gebieden terug te vinden is. Waar deze effecten gewoonlijk worden bestudeerd met behulp van synthetische stimuli, extrapoleren we deze resultaten naar het gebruik van natuurlijke afbeeldingen.

Appendix

List of publications

List of publications

Journal articles

Zuiderbaan W, Dumoulin SO. (in preparation). Enhanced responses in early visual cortex to subjectively important aspects of natural scenes.

Zuiderbaan W, Harvey BM, Dumoulin SO. (under review). Image identification from brain activity using the population receptive field model.

Zuiderbaan W, van Leeuwen, J., Dumoulin SO. (under review). Change blindness is influenced by both low-level image properties and high-level image representation

Dumoulin, S.O., Harvey, B.M., Fracasso, A., **Zuiderbaan, W.**, Luijten, P.R., Wandell, B.A., Petridou, N (2017). In vivo evidence of functional and anatomical stripe-based subdivisions in human V2 and V3. *Scientific reports* 7(1), 733.

Harvey BM, Vansteensel MJ, Ferrier CH, Petridou N, **Zuiderbaan W**, Aarnoutse EJ, Bleichner, MG, Dijkerman HC, van Zandvoort MJE, Leijten FSS, Ramsey NF, Dumoulin SO (2013). Frequency-specific spatial interactions in human electrocorticography: V1 alpha oscillations reflect surround suppression. *NeuroImage*. 65: 424-432. (*Editor's Choice Award for the best article in the field of Systems Neuroscience (2013) NeuroImage*)

Zuiderbaan W, Harvey BM, Dumoulin SO (2012). Modeling center-surround configurations in population receptive fields using fMRI. *Journal of Vision*. 12(3):10.

Peer-Reviewed Conference Abstracts (Talk)

Zuiderbaan W, Harvey BM, Dumoulin SO (2012). Using the population receptive field model to identify images from fMRI signals. *VSS 12.* [Vision Sciences Society]

Zuiderbaan W, Harvey BM, Dumoulin SO (2011). Image identification from brain activity using the population receptive field model. *SFN. Abstr.* 851.06 [Society for Neuroscience]

Zuiderbaan W, Ress D, Harvey BM, Green CA, Dumoulin SO (2010). Modeling center-surround configurations in population receptive fields using fMRI. *SFN. Abstr.* 325.10 [Society for Neuroscience]

Peer-Reviewed Conference Abstracts (Poster)

Zuiderbaan W, van Leeuwen, J. Dumoulin SO (2016). Change detection: the role of low-level versus high-level image representations *VSS 16.* [Vision Sciences Society]

Zuiderbaan W, Dumoulin SO (2015). Deriving contrast response functions from fMRI responses to natural images. *SFN. Abstr.* 700.03 [Society for Neuroscience]

Zuiderbaan W, Harvey BM, Dumoulin SO (2010). Measuring center-surround configurations using fMRI. *FENS. Abstr.* 081.42. [Federation of European Neuroscience Societies]

Appendix

Curriculum Vitae

Curriculum Vitae

Wietske Zuiderbaan was born on June 12th 1982 in Wijckel, the Netherlands. In 2000 she graduated from her secondary education at Bogerman College in Sneek. After a detour of studying Clinical Child and Educational Studies at Utrecht University and working outside science, she started her study Cognitive Artificial Intelligence at Utrecht University, of which she completed the masters '*cum laude*' in 2010. During her masters she performed her internship in the Dumoulin lab, where she continued as a PhD student at the department of Experimental Psychology of Utrecht University under the direct supervision of Prof. Dr. Serge Dumoulin. She will now start as a postdoctoral researcher at the Brigham and Woman's Hospital-Harvard Medical School in Boston, USA, in the lab of Dr. Stelios Smirnakis and Dr. Lucia Vaina.