

2006 Special Issue

V4 receptive field dynamics as predicted by a systems-level model of visual attention using feedback from the frontal eye field

Fred H. Hamker*, Marc Zirnsak

Allgemeine Psychologie, Psychologisches Institut II, Westf. Wilhelms-Universität, Münster, Germany

Received 15 July 2006; accepted 1 August 2006

Abstract

Visual attention is generally considered to facilitate the processing of the attended stimulus. Its mechanisms, however, are still under debate. We have developed a systems-level model of visual attention which predicts that attentive effects emerge by the interactions between different brain areas. Recent physiological studies have provided evidence that attention also alters the receptive field structure. For example, V4 receptive fields typically shrink and shift towards the saccade target around saccade onset. We show that receptive field dynamics are inherently predicted by the mechanism of feedback in our model. According to the model an oculomotor feedback signal from an area involved in the competition for the saccade target location, e.g. the frontal eye field, enhances the gain of V4 cells. V4 receptive field dynamics can be observed after pooling the gain modulated responses to obtain a certain degree of spatial invariance. The time course of the receptive field dynamics in the model resemble those obtained from macaque V4.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Attention; Model; Receptive field; Oculomotor feedback; Biased competition; Saliency; V4; FEF

1. Introduction

Attention refers to the net effect of multiple mechanisms that leads to a focusing of the available processing resources. It is known to improve visual perception and action in a number of ways, such as speeding up the reaction time towards a stimulus (Posner, Snyder, & Davidson, 1980), improving change detection (Rensink, O'Regan, & Clark, 1997), enhancing the perceived contrast (Carrasco, Ling, & Read, 2004) and increasing the spatial resolution (Yeshurun & Carrasco, 1998). A large number of models have been inspired by the classical idea of a 'spotlight of attention' that highlights an area of interest by routing that information into higher areas for further processing. Such general approaches raise at least two fundamental issues.

First of all, what is the source that determines the location and shape of a spatially selective attentional focus? More than 10 years ago, due to the lack of detailed electrophysiological

data, attention has been described as a selection or winner-takes-all process within a saliency (or master) map (Koch & Ullman, 1985; Treisman & Gelade, 1980; Wolfe, 1994). Such a map has been defined to indicate potentially relevant locations by an enhanced activity at the corresponding spatial location. In the search for the saliency map a number of brain areas have been identified. Among those are the frontal eye field (Schall, 2002; Thompson & Schall, 2000), the superior colliculus (Ignashchenkova, Dicke, Haarmeier, & Thier, 2004; Muller, Philastides, & Newsome, 2005) and LIP (Bisley & Goldberg, 2006). However, area V4 has also been shown to reflect aspects of a saliency map (Bichot, Rossi, & Desimone, 2005; Mazer & Gallant, 2003; Ogawa & Komatsu, 2004), which suggests that saliency alone might not be a sufficient criterion for defining the source of spatial attention in the brain. In fact, we suggested a model in which the information of saliency in V4 can be task relevant immediately after the presentation of a visual scene regardless of spatial attention (Hamker, *in press*). This selective enhancement at intermediate levels of the cortical hierarchy could be used to guide visuomotor processes such as eye movements. Inspired by electrophysiological and behavioral observations (Deubel & Schneider, 1996; Kowler, Anderson, Doshier, & Blaser, 1995; Kustov & Robinson, 1996;

* Corresponding address: Allgemeine Psychologie, Psychologisches Institut II, Westf. Wilhelms-Universität, Fliednerstrasse 21, 48149 Münster, Germany. Tel.: +49 251 83 34171; fax: +49 251 83 34173.

E-mail address: fhamker@uni-muenster.de (F.H. Hamker).

Moore & Armstrong, 2003; Rizzolatti, Riggio, Dascola, & Umiltà, 1987), we have provided computational evidence that, at least one, spatially selective feedback signal arrives from premotor cells of the oculomotor system such as the frontal eye field movement cells (Hamker, 2005a). This assumption makes specific predictions, since it constrains the timing and the spatial location of the feedback signal. Movement cells show only little, if any, response related to the onset of the stimulus and they start to increase in firing prior to saccade onset (although there is a continuity of frontal eye field visuomovement cells from showing only little to having a strong onset response). Thus, a spatially selective feedback at the saccade target occurs just prior to saccade onset but not immediately after the saccade target onset, since the movement cells require time to build up. However, spatially selective processing can occur earlier, due to feature-specific top-down signals (Hamker, *in press*).

The second issue is the impact of a spatially selective feedback signal on visual processing. The most simple and common mechanism is that of a gating mechanism according to which the cells which receive feedback gate their input to higher areas for object perception. Even when this gating can be gradual with respect to the strength of the feedback signal, it nevertheless implements an on/off switch depending of the presence/absence of the feedback signal. Inspired by the observation that object recognition can be very fast and probably even possible without prior spatial selection (Li, VanRullen, Koch, & Perona, 2002; Rousset, Thorpe, & Fabre-Thorpe, 2004) we suggested that feedback affects just the gain (Hamker, 2003, 2004, 2005a) and showed that this mechanism is consistent with the frameworks of multiplicative scaling (McAdams & Maunsell, 1999) and biased competition (Desimone & Duncan, 1995) if, in addition, lateral interactions exist. The framework of biased competition is built upon the following observation: When two stimuli are presented within the receptive field of a neuron, the influence of the non-attended stimulus is suppressed, as if the receptive field shrinks around the attended stimulus (Moran & Desimone, 1985; Reynolds, Chelazzi, & Desimone, 1999). The idea of a receptive field shift has been supported by the observation that the response profile is distorted towards the attended location, even when attention is directed outside of the receptive field (Connor, Preddie, Gallant, & Van Essen, 1997). Receptive field shifts might be an indirect result from a multiplicative scaling further upstream (McAdams & Maunsell, 1999). However, little work has been done to directly measure the receptive field profiles in attended and non-attended situations. A direct mapping of the receptive field profile has been done peri-saccadically (Tolias, Moore, Smirnakis, Tehovnik, & Siapas, 2001) and it can be inferred from this study that similar effects occur in covert shifts of attention.

In this article we will demonstrate that our framework is able to qualitatively reproduce the above mentioned receptive field effects. We suggest that gain modulation occurs in the same area where the receptive field effects are observed, after the response of the feature detectors (“simple cells”) and before the activity is spatially pooled

onto “complex cells”. This hierarchical processing is consistent with feedforward models of object recognition (Fukushima, 1980; Riesenhuber & Poggio, 1999; Spratling, 2005), where in addition, feedback increases the gain prior to spatial pooling. Since the computation of area V4 in our present large scale model of visual attention (Hamker, 2005b, 2005c, 2005d, *in press*) has been simplified to a single layer in which the gain modulation takes place, we extend area V4 to three (functional) layers. The first layer combines the input from cells of earlier areas in the hierarchy — a layer of feature detectors. These cells project to the next layer in which feedback enhances the gain. The third layer spatially pools the responses of the second layer to implement a limited range of spatial invariance. The feedback signal originates in IT and in the FEF movement cells. We show here that this extended model shows peri-saccadic receptive field dynamics similar to that observed in V4 (Tolias et al., 2001).

2. Model

The present model is an extension of an earlier model which has been described in detail on tasks such as object detection in natural scenes, change detection, visual search, feature-based attention and other attentional experiments (Hamker, 2005b, 2005c, 2005d, *in press*). The full model description is located in Appendix A.

The model consists of visual areas V4, inferotemporal (IT) cortex, prefrontal areas that contain the frontal eye field (FEF) for saccade planning and more ventrolateral parts for implementing functions of working memory (Fig. 1).

If we present a visual scene to the model, features such as color, intensity and orientation are computed from the image. The fact that features that are unique in their environment ‘pop-out’ is accounted for by computing an initial stimulus-driven saliency which determines the input into V4. We consider this stage a simplification with respect to its location in the brain. Pop-out effects are not necessarily generated early in the visual pathway. They are probably also computed in later areas, such as IT.

In extension to the original model (Hamker, 2005a, 2005c) V4 is now simulated by 3 layers: V4in, V4gain and V4pool. Feedback from the FEF and IT increases the gain of the cells in V4 gain. Pooling these gain modulated responses results in a larger degree of spatial invariance. However, for simplicity the complexity of features is not increased from V4 to IT. We have shown earlier that such a hierarchy of processing in V4 allows to quantitatively replicate single cell recordings of experiments investigating the ‘biased competition’ of two stimuli in a receptive field of a V4 cell under variable attentional conditions (Hamker, 2004, 2005b).

The growing receptive field size along the processing hierarchy requires that a number of V4 pool cells project to a single IT cell. Search in this model can be goal directed since IT receives feature-specific feedback from the prefrontal memory (PFmem) cells.

The planning of an eye movement is implemented as follows. The FEF visuomovement (FEFv) neurons receive convergent afferents from V4in and V4pool. The input activity at each location is summed across all dimensions (e.g. color,

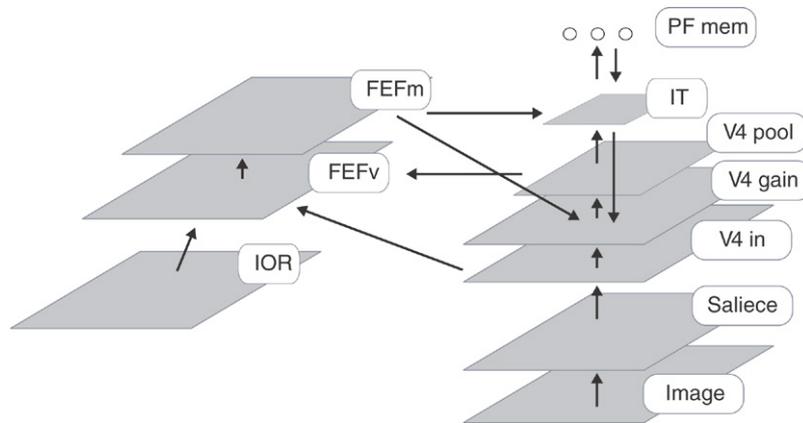


Fig. 1. Model for visual attention. First, information about the content and its low level stimulus-driven saliency is extracted, as indicated by the map “Saliency”. This information is sent further downstream to V4 and to IT cells which are broadly tuned to location. A target template is encoded in PF memory (PFmem) cells. Feedback from PFmem to IT increases the strength of all features in IT matching the template. Feedback from IT to V4gain sends the information about the target downwards to cells with a higher spatial tuning. FEF visuomovement (FEFv) cells combine the feature information across all dimensions and indicate salient or relevant locations in the scene. The FEF movement (FEFm) cells compete for the target location of the next eye movement. The activity of the FEF movement cells is also sent to V4 gain and IT for gain modulation. The IOR map memorizes recently visited locations and inhibits the FEF visuomovement cells. However, this map is only required for the simulation of a scanpath but not for the receptive field dynamics simulated here.

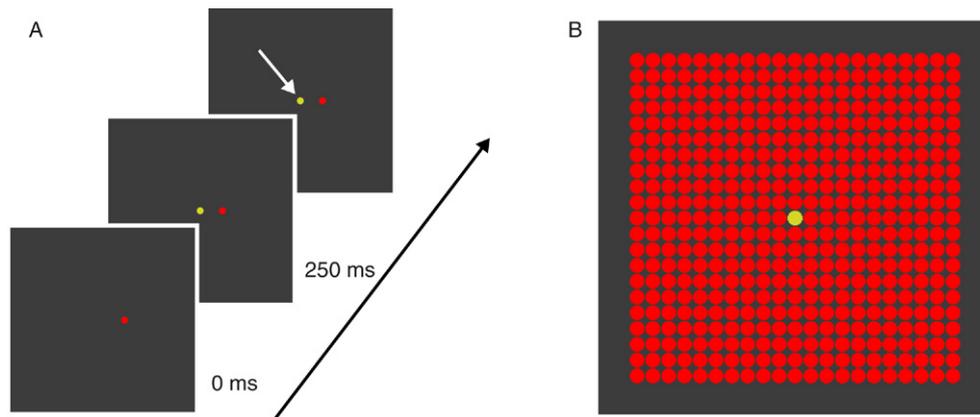


Fig. 2. We used images of 400×400 pixels in size, where each dot is 16 pixels in diameter. (A) Sequence of stimulus presentation. A probe stimulus was presented for 250 ms followed by a saccade target. The trial was finished when the model indicated the onset of an eye movement towards the saccade target as illustrated by the arrow. Since the model does not come with a foveal representation, a specific fixation stimulus is not required. (B) A grid of 21×21 probe stimuli is used for measuring the spatiotemporal sensitivity of V4pool cells.

orientation). The information from the target template held in PFmem cells additionally enhances the locations that result in a match between target and encoded feature. This allows the biasing of specific locations by the joint probability that the searched features are encoded at a certain location. The firing rate of FEF visuomovement cells represents the saliency and task-relevance of a location. The effect of the FEF visuomovement cells on the FEF movement cells (FEFm) is a feedforward excitation and surround inhibition. Thus, by increasing their activity slowly over time FEF movement cells compete for the selection of the strongest location. If a FEF movement cell exceeds a threshold, an eye movement is indicated towards the location of the center of gravity of movement cell activity.

3. Methods

In the present study we are specifically interested in the receptive field dynamics of V4 pool cells due to the feedback

from the FEF movement cells to V4gain cells. We used a similar experimental procedure as in Tolias et al. (2001). A red probe stimulus was presented to the model for 250 ms followed by the yellow saccade target stimulus (Fig. 2A). The probe was selected from a grid of possible probe positions and remained visible for the whole trial (Fig. 2B). The task of the model was to plan a saccade to the target stimulus.

We used colored stimuli to avoid an interference of the saccade target with the probe. The saccade target evoked no response in the *RG* channel of the model. This allows us to study the effect of the planned eye movement towards the saccade target on the activity evoked by the probe stimulus. Tolias et al. (2001) reported that the saccade target used evoked only weak responses even when placed within the classical receptive field of the observed cell. The receptive field structure was analyzed using the response in the *RG* channel of V4pool cells.

The receptive field of the cells was mapped by running 440 trials with variable probe positions. As a result we obtain a

21×21 matrix for each cell encoding the response of the cell with respect to the probe location. Since it is not possible to present a probe at the saccade target, the activity for this position was obtained by averaging the activity from the surrounding probes. For better illustration purposes we interpolated the response matrix to the resolution of the original image.

The classical (pre-saccadic) and peri-saccadic receptive field borders were determined using a half-maximum response threshold, which is a common method to analyze physiological data (Tolias et al., 2001). The pre-saccadic receptive field was mapped at $t = 80$ ms relative to probe onset where the model cells reach their maximum and the peri-saccadic receptive field at $t = 0$ ms relative to saccade onset. The receptive field of a given cell is then defined as the area in visual space in which the activity exceeds half of the maximal activity of this cell.

In order to withhold an eye movement plan towards the probe, the FEF fixation cell was activated until 330 ms after probe onset. The model was set to search for a yellow stimulus by activating appropriate cells in PFMem throughout a trial. This top-down activation was sufficient to bias the model dynamics such that it plans an eye movement towards the target stimulus.

4. Results

We simulated the model's response on a persistent probe until the onset of a saccade. The saccade target appeared 250 ms after probe onset and the model was programmed to plan an eye movement towards the saccade target. Fig. 3 shows a single trial. The stimulation by the probe stimulus activates the V4gain cells of the *RG* channel which in turn drive the FEFv cells. Since the probe should not elicit an eye movement the FEF fixation cells are activated. Thus, only little activity is found in the FEFm cells. When the saccade target is presented after 250 ms strong activity is built up in the V4gain cells of the *BY* channel, especially since the cells encoding the yellow color of the saccade target receive a top-down signal from the PFMem cells via IT. As a result of the enhanced processing of the saccade target a gradual shift of the response profile in the FEF visuomovement cells can be observed. Since the external, cognitive related, input into the fixation cell is removed, the FEF movement cells start to compete for the saccade target location. This increase in activity of the movement cells is fed back to all channels in V4gain such that the population response in the *RG* channel towards the probe gets distorted in the direction of the planned saccade target. At 378 ms the model indicates the onset of the eye movement. Due to the influence of the probe on the saccade plan the planned saccade endpoint is slightly shifted upwards.

In order to determine the spatiotemporal sensitivity of V4pool cells we presented the probe at different locations. Fig. 4 shows the spatiotemporal sensitivity profile of a single V4 pool cell of the *RG* channel which is sensitive to the probe stimulus. The cell fires to the appearance of the probe inside its classical receptive field. Although the profile changes in time

from a strong onset transient to a lower sustained response, the receptive field center is roughly at the same position long before saccade onset. At -80 ms the spatiotemporal sensitivity starts to expand towards the saccade target location and from -60 to -40 ms on, the peak sensitivity is shifted in the direction of the saccade target. In addition, the sensitivity profile is narrowly tuned around its center and its peak is enhanced again. The spatiotemporal sensitivity from this model cell is very similar to the spatiotemporal sensitivity observed in macaque V4 cells (Tolias et al., 2001). Fig. 5 shows five different examples of pre- and peri-saccadic receptive fields. In all cases the peri-saccadic receptive field shrinks and shifts towards the saccade target.

The present model does not take the detailed neuroanatomical structure of the simulated brain areas into account. Factors like changes in cortical magnification, i.e., the amount of cortical tissue which processes a certain part of the visual space (which is highest at the fovea) and the increase in receptive field size with increasing eccentricity are not implemented yet. Each part of the 'visual space' (the input image) is processed by the same amount of cells within a model area using a constant receptive field size. Therefore, the exact size of both the model receptive fields and the part of the simulated visual/cortical space is arbitrary. For the purpose of the present study we decided to simulate a rather small part of both the visual and the cortical space to obtain a high resolution in terms of the absolute number of cells which are involved in the reported receptive field dynamics to minimize simulation artifacts.

For example, let us assume that the saccade length in our simulation is 20° , then the receptive fields around the saccade target would be up to 15° in diameter (the square root of the receptive field area) as measured in monkey V4 (Boussaoud, Desimone, & Ungerleider, 1991). If we set the measured receptive field size equal to the size of the pre-saccadic model receptive fields shown in Fig. 5, the input image would cover a visual space of about $20^\circ \times 20^\circ$ (the original fixation would be outside the area simulated) and the peri-saccadic receptive field of the cell in the middle panel would be about 7° in diameter. The magnitude of the shift (distance between pre- and peri-saccadic receptive field center) of the cell shown in the leftmost panel would be roughly 8° .

We have further analyzed the pre- and peri-saccadic activity depending on the position of the probe in the classical (pre-saccadic) receptive field of a V4pool cell (Fig. 6). If the probe is presented in the center of the classical receptive field we observe the strongest onset response. The positions of the probe stimuli at the receptive field borders were chosen to both elicit a similar onset response. However, the peri-saccadic response clearly depends on the alignment of the probe to the endpoint of the planned saccade target. Only when the probe is aligned to the planned saccade endpoint the model predicts a sustained enhancement of the response. This model prediction is similar to observations in V4 after FEF microstimulation (Armstrong, Fitzgerald, & Moore, 2006).

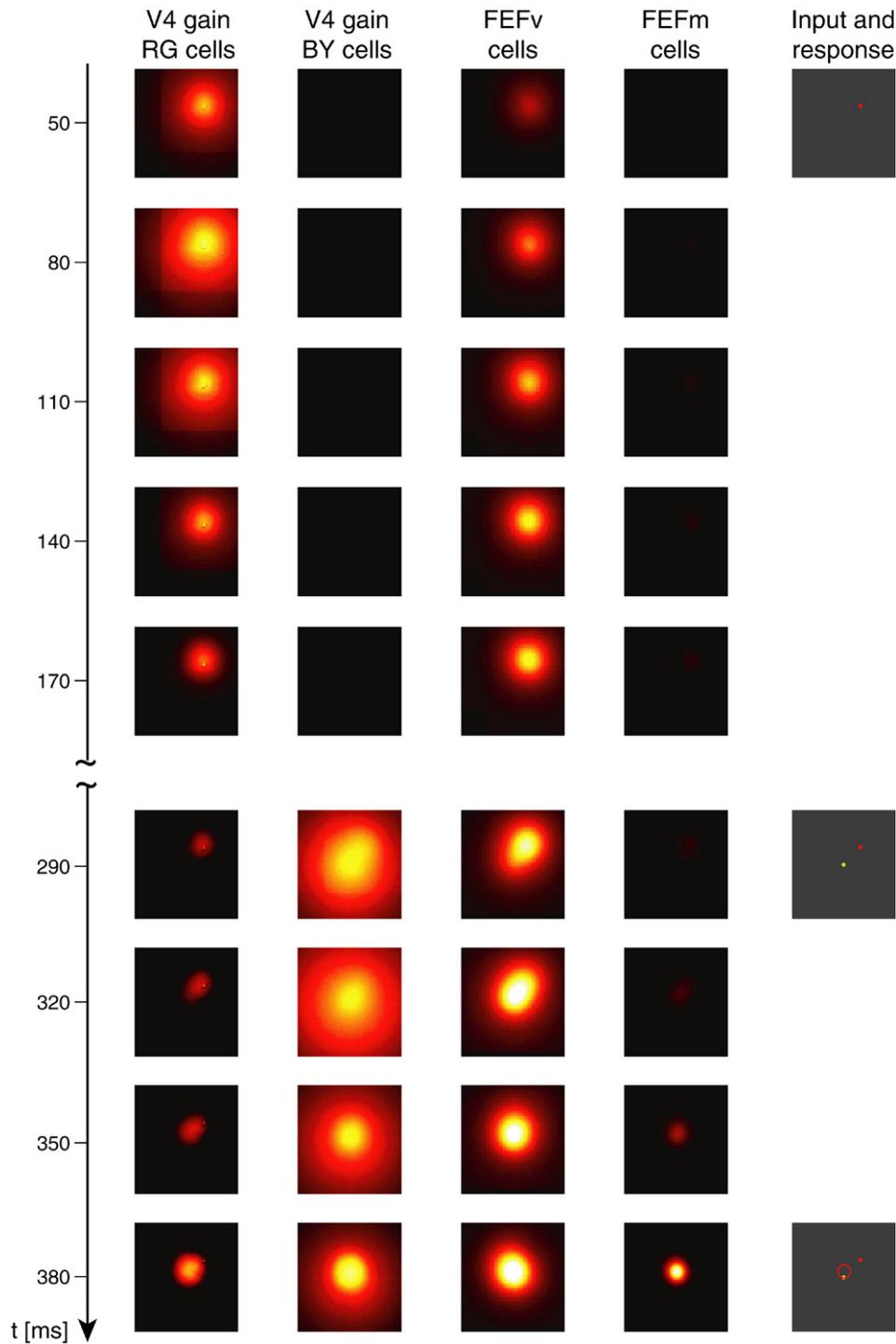


Fig. 3. Time course of the essential activity in the model. On the right side the changes in the input and the final location of the eye movement (red circle) are illustrated. The V4gain cells in the *RG* channel encode the probe stimulus and the ones in the *BY* channel encode the saccade target. The FEFv cells indicate salient or task relevant locations. The FEFm cells compete for the saccade target location. Details of the model's response are given in the text. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Discussion

Our model allows to simulate comparable experiments as being performed with humans or monkeys, even with respect to the temporal presentation of stimuli. The type of stimuli, however, has to be relatively simple. In addition, the number of possible experiments is limited since the model does not

account for the variable resolution and receptive field size across eccentricity nor does it simulate the eye movement itself. Despite these limitations, the model allows to compare its internal activity with cell recordings in different areas of the monkey brain, possibly with fMRI in humans and the model's outcome can be linked to behavior, such as the target selection for eye movements. In addition, our model is constrained

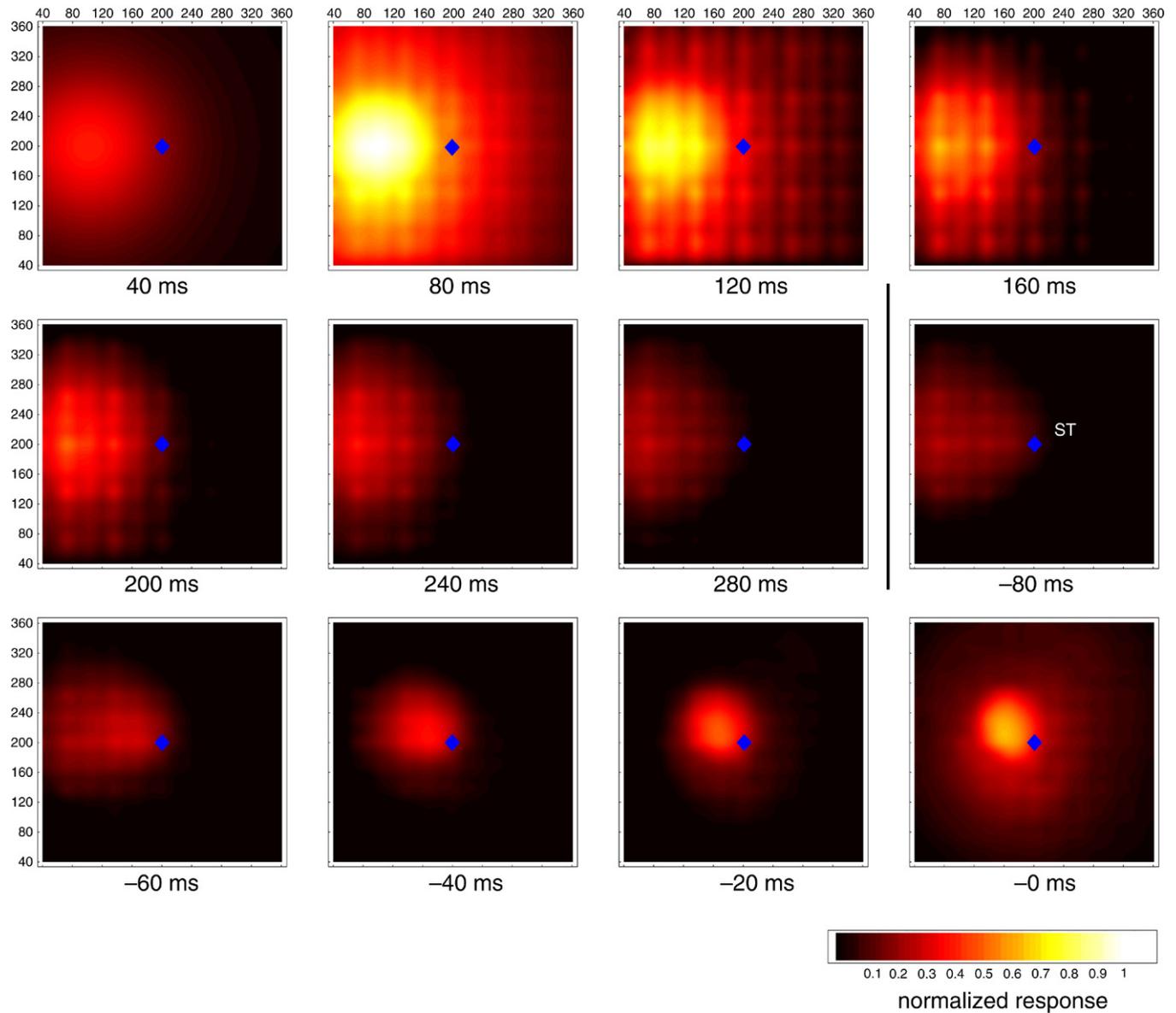


Fig. 4. : Spatiotemporal responses of a V4pool cell in relation to the saccade target (ST) as color coded-activity maps within the range of 40 to 360 pixels where the probe stimuli were presented. The first seven figures show the cell’s activity from 40 to 280 ms after probe onset at the different probe positions. The remaining five figures show the cell’s activity prior to saccade onset at 0 ms. The spatiotemporal response reflects a strong transient at about 80 ms relative to probe onset. The cell’s responsive area starts to expand towards the saccade target at -80 ms and then gradually shrinks and shifts towards the saccade target.

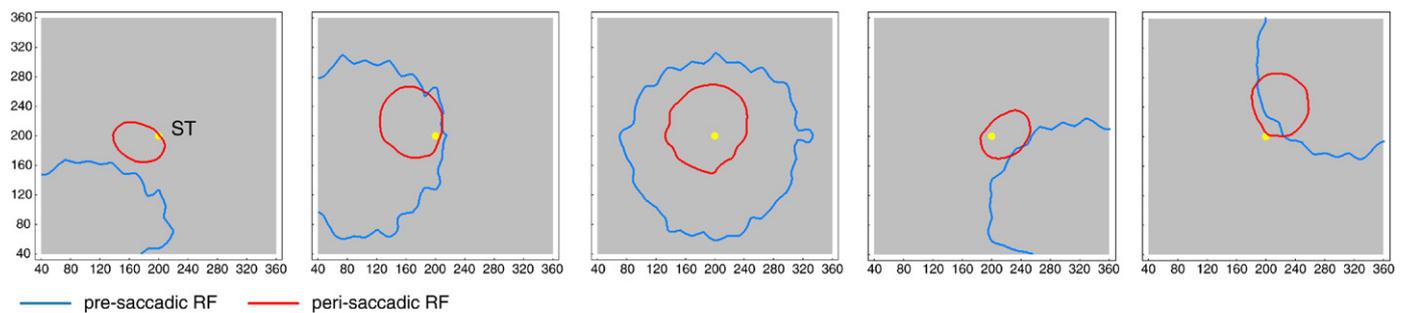


Fig. 5. Five different examples of pre-saccadic and peri-saccadic receptive field borders as measured by a half-maximum rule. Typically the peri-saccadic receptive fields shrink around the saccade target. Those which do not comprise the saccade target do shrink and shift towards the saccade target. Due to the lateral suppression in the V4gain and V4pool cells, the peri-saccadic receptive field can even lie outside the classical receptive field as observed in macaque V4 (Tolias et al., 2001).

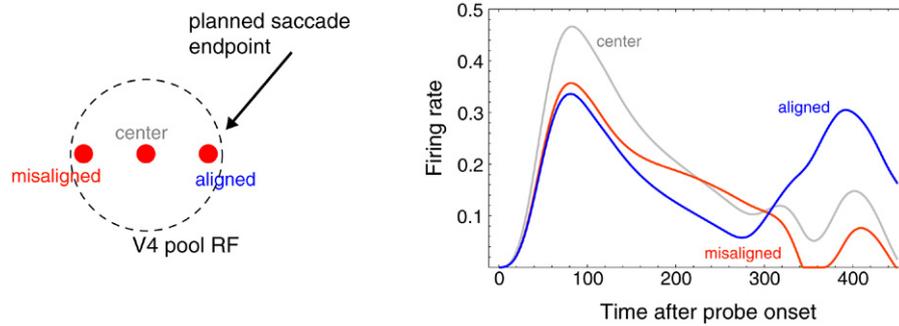


Fig. 6. Response of a V4pool cell to three different probe positions within its classical receptive field. The planned saccade endpoint is located outside the receptive field but close to its border as indicated by the arrow. When the aligned probe is presented the activity is reactivated prior to the eye movement.

from the functional point of view: Although the model is still relatively simple, it captures several essentials to survive in a real world context.

As far as the source of attention is concerned, we have presented further evidence for our approach to modeling vision in a distributed architecture using feedback (and top-down) to drive the system into a specific ‘attentional’ state. An attention system whose parts are solely devoted to the control of visual ‘information’ flow is probably not necessary. The source of attention has been often suggested as being localized in the parietal cortex. Our model suggests that the parietal cortex is probably not a direct source of attention, at least for the experiments discussed here. The parietal cortex is strongly linked to space perception and we suggest that the parietal cortex is just another piece of the whole network of visual perception which provides more spatially related biases into the system.

In addition to these more general, architectural issues, we have been interested in a number of specific questions regarding the source and goal of attention. We have suggested earlier that the FEF movement cells provide a useful feedback signal that is consistent with many observations (Hamker, 2005a). The model predictions of the peri-saccadic shift and shrinkage of receptive fields (Fig. 4) and the selective enhancement of the aligned probe (Fig. 6) further support our claim. This does not mean that we want to rule out more visually related feedback signals as suggested by some studies (Juan, Shorter-Jacobi, & Schall, 2004; Thompson, Biscoe, & Sato, 2005), we only suggest that a part of the overall attention effect can originate in the movement cells of the frontal eye field or the burst cells of the superior colliculus, since both have a similar time course, are highly interconnected and project directly/indirectly to visual areas (Wurtz, Sommer, & Cavanaugh, 2005).

The other issue raised is the effect of a feedback signal on visual processing. The model’s feedback signal from the movement cells provides a natural spatial structure as determined by the competitive interactions. In earlier simulations we observed that the population varies its size and shape with respect to the task (Hamker, 2005c) due to the interactions of spatial and feature-specific feedback. We also observed a split of spatial attention due to the presence of multiple activity hills (Hamker, 2005c). It is yet not fully clear how the width of the feedback signal affects recognition

performance. However, it is probably not necessary that the feedback signal must selectively cover the to be recognized object as suggested by spotlight models of visual attention which only gate the content within the focus of attention into higher areas for recognition. Our model suggests that the feedback signal tunes the receptive field structure. A shift of the receptive fields towards the planned endpoint of the saccade, in addition to the overall gain enhancement, increases the processing recourses around the location of the next fixation. Furthermore, a shrinkage of the receptive fields reduces the influence of clutter on the population response (Moran & Desimone, 1985). Thus, a feedback signal that alters the receptive field structure by increasing the gain could improve object recognition in a number of ways that are less error prone than a spotlight model which only gates visual processing.

Our model inherently predicts that the number of receptive fields increases peri-saccadically around the planned saccade target. The exact receptive field dynamics, however, depend on a number of parameters. A strong shrinkage requires that the width of the feedback signal is smaller than the receptive field of the observed cell and the receptive field size of the V4pool cells is sufficiently larger than that of the V4gain cells. A shift of the receptive field is generally observed with a variety of parameter settings. When the distance between the center of the feedback signal and the receptive field is large, an expansion of the receptive field can also be observed. However, with longer distances between the location of the feedback signal and receptive field the effect diminishes.

The model predicts that the observed receptive field changes in V4 can emerge by a gain increase due to feedback and simple competitive interactions. We especially obtained a detailed temporal description of receptive field changes due to the simulation of a systems-level network. However, we have only developed a general mechanism that can qualitatively account for the observed receptive field changes in V4. We suggest that this mechanism might also operate in other brain areas such as MT. Future, more quantitative analyses have to take into account several additional factors such as the exact size and location of the pre-saccadic receptive field with reference to the saccade target and the representation of visual space due to cortical magnification. Furthermore, the cortical location of gain modulation plays a role (Compte & Wang, 2006). Such more detailed models would also require more specific experimental data at the single cell level.

Acknowledgement

This work has been supported by the Deutsche Forschungsgesellschaft (DFG-Project HA2630/4-1).

Appendix A. Model equations

Attention in the model has two different sources, one is stimulus-driven and the other is task-driven (Fig. 1). In order to compute the stimulus-driven source we (i) create multi-resolution feature maps, (ii) compute multi-resolution contrast maps using center-surround operations and (iii) combine both in feature conspicuity maps. For computing the first two steps we largely follow Itti, Koch, and Niebur (1998) but see Hamker (2005c) for differences in these early processing stages.

The initial conspicuity is then continuously updated to reflect the task-relevance. The relevance of each feature is determined by the search template (target). Feedback enhances the gain of feedforward processing and the network ultimately settles onto a final response and a specific attentional state. Thus, attention emerges by the dynamics of vision.

A.1. Feature maps

We currently use color, intensity and orientation as basic features. To construct the color channels R , G , B , Y , the color values r , g , and b from the image are normalized by $I = (r + g + b)/3$ in order to decouple hue from intensity. For each pixel in the pyramid we generate the color channels $R = r - (g + b)/2$ for red, $G = g - (r + b)/2$ for green, $B = b - (r + g)/2$ for blue, and $Y = (r + g)/2 - |r - g|/2 - b$ for yellow (negative values are set to zero). The brain represents colors within an opponency system $RG = R - G$ and $BY = B - Y$. The values can be negative and positive, but for an easy visualization we shift and rescale the values to 0–255. All features are represented within a Gaussian pyramid, which is constructed by progressively low-pass filtering and sub-sampling the input images of the channels (Burt & Adelson, 1983).

A.2. Contrast maps

Contrast maps represent the conspicuity of each feature. In analogy to the known influence of lateral excitation and surround inhibition, center-surround operations ‘ \ominus ’ calculate the difference of maps with a fine scale σ and a coarse scale $s = \sigma + \delta$. This operation across spatial scales is done by interpolation to the fine scale and then point-by-point subtraction. The variation of the distance δ between resolutions results in a multi-scale feature extraction (Itti et al., 1998). For each pixel of the resolution σ we create intensity contrast maps $\mathcal{I}(c, s)$ by subtracting the map with the coarse scale s from the one with center scale c .

$$\mathcal{I}(c, s) = |I(c) \ominus I(s)| \quad \begin{array}{l} c \in \{2, 3\} \\ \delta \in \{3, 4\}. \end{array} \quad (1)$$

Similarly, we create the color double opponent values by

$$\begin{aligned} \mathcal{RG}(c, s) &= |RG(c) \ominus RG(s)| \\ \mathcal{BY}(c, s) &= |BY(c) \ominus BY(s)|. \end{aligned} \quad (2)$$

Since the contrast in the color channels is small for natural images we stretch the scale by a non-linear function s :

$$\begin{aligned} \widehat{\mathcal{RG}}(c, s) &= s(\mathcal{RG}(c, s)) \\ \widehat{\mathcal{BY}}(c, s) &= s(\mathcal{BY}(c, s)) \end{aligned} \quad (3)$$

with

$$s(x) = \begin{cases} k_C \cdot x & \text{if } k_C \cdot x \leq 255 \\ 255 & \text{if } k_C \cdot x > 255 \end{cases} \quad \text{and } k_C = 3. \quad (4)$$

This specific scaling function s can equal very high contrast values up to 255, but since the hue is not very high we typically do not run into this situation.

We average the maps obtained by a different course scale $s = \sigma + \delta$ to receive one contrast value per channel and center scale:

$$\begin{aligned} \mathcal{I}(c) &= \frac{1}{\#s} \bigoplus_s \mathcal{I}(c, s) \\ \mathcal{RG}(c) &= \frac{1}{\#s} \bigoplus_s \widehat{\mathcal{RG}}(c, s) \\ \mathcal{BY}(c) &= \frac{1}{\#s} \bigoplus_s \widehat{\mathcal{BY}}(c, s). \end{aligned} \quad (5)$$

Orientation contrast maps are computed for each orientation $\theta \in \{0, 45, 90, 135\}$ using the center-surround operation with a fine scale c and a course scale $s = \sigma + \delta$.

$$\mathcal{O}(c, s, \theta) = \begin{cases} O(c, \theta) \ominus O(s, \theta) & \text{if } O(c, \theta) > O(s, \theta) \\ 0 & \text{else.} \end{cases} \quad (6)$$

A.3. Feature conspicuity maps (Salience)

The feature conspicuity maps combine the feature information \mathbf{V} , like orientation θ or intensity I , with its gain P , obtained from the conspicuity such as \mathcal{O} or \mathcal{I} , into a population code. We construct a space, whose axes are defined by the represented features and by the conspicuity. The feature information is encoded by the location of the cell $i \in N = \{1, 2, 3, 4, 5, 6\}$ in feature space and the conspicuity value ($P \in \{\mathcal{O}, \mathcal{I}, \mathcal{RG}, \mathcal{BY}\}$) determines the firing rate r_i :

$$r_i = P \cdot g(\mathbf{u}_i - \mathbf{V}). \quad (7)$$

Specifically we use a Gaussian tuning curve to model how each unit i is tuned around its preferred value \mathbf{u}_i with the selectivity parameter σ_g :

$$g(\mathbf{u}_i - \mathbf{V}) = \exp\left(-\frac{\|\mathbf{u}_i - \mathbf{V}\|^2}{\sigma_g^2}\right). \quad (8)$$

To apply the same range of selectivity parameters $\sigma_g^2 \in \{0.05 \dots 0.2\}$ for all channels we normalize the feature values \mathbf{V} of each channel between 0 and 1 and get \tilde{I} , \tilde{RG} , \tilde{BY} , $\tilde{\theta}$ and $\tilde{\sigma}$. The initial conspicuity value should typically lie within the range of 0 and 1. Thus, we also normalize contrast values to $\tilde{\mathcal{I}}$,

$\widetilde{\mathcal{R}}\mathcal{G}$, $\widetilde{\mathcal{B}}\mathcal{Y}$, $\widetilde{\mathcal{O}}$. We finally obtain the populations for each channel with scale c at each location \mathbf{x} :

$$\begin{aligned} r_{I,i}(c, \mathbf{x}) &= \widetilde{\mathcal{I}}(c, \mathbf{x}) \cdot g(u_i - I(c, \mathbf{x})) \\ r_{RG,i}(c, \mathbf{x}) &= \widetilde{\mathcal{R}}\mathcal{G}(c, \mathbf{x}) \cdot g(u_i - RG(c, \mathbf{x})) \\ r_{BY,i}(c, \mathbf{x}) &= \widetilde{\mathcal{B}}\mathcal{Y}(c, \mathbf{x}) \cdot g(u_i - BY(c, \mathbf{x})). \end{aligned} \quad (9)$$

The orientation information is transferred into two channels, one for scale or spatial frequency σ and one for orientation θ . The orientation channel reads:

$$r_{\theta,i}(c, \theta, \mathbf{x}) = \widetilde{\mathcal{O}}(c, \theta, \mathbf{x}) \cdot g(u_i - \theta). \quad (10)$$

Since it is not feasible to represent orientations in different maps within a population code, we combine the maps across orientations:

$$r_{\theta,i}(c, \mathbf{x}) = \max_{\theta} (r_{\theta,i}(c, \theta, \mathbf{x})). \quad (11)$$

In order to further reduce the information we ignore the different center scales using only the maximal contribution across the center scale:

$$\begin{aligned} r_{\theta,i}(\mathbf{x}) &= \max_{c, \mathbf{x}' \in RF(\mathbf{x})} r_{\theta,i}(c, \mathbf{x}) \\ r_{I,i}(\mathbf{x}) &= \max_{c, \mathbf{x}' \in RF(\mathbf{x})} r_{I,i}(c, \mathbf{x}) \\ r_{RG,i}(\mathbf{x}) &= \max_{c, \mathbf{x}' \in RF(\mathbf{x})} r_{RG,i}(c, \mathbf{x}) \\ r_{BY,i}(\mathbf{x}) &= \max_{c, \mathbf{x}' \in RF(\mathbf{x})} r_{BY,i}(c, \mathbf{x}). \end{aligned} \quad (12)$$

The 5th conspicuity map is gained from the spatial resolution of the steerable filters. Thus, the orientation information is also transferred into features encoding spatial frequency σ :

$$r_{\sigma,i}(c, \theta, \mathbf{x}) = \mathcal{O}(c, \theta, \mathbf{x}) \cdot g(u_i - \sigma). \quad (13)$$

As for orientation, we combine the maps across spatial frequencies:

$$r_{\sigma,i}(\theta, \mathbf{x}) = \max_c (r_{\sigma,i}(c, \theta, \mathbf{x})) \quad (14)$$

and repeat the same process across the orientation:

$$r_{\sigma,i}(\mathbf{x}) = \max_{\theta} (r_{\sigma,i}(\theta, \mathbf{x})). \quad (15)$$

A.4. V4in cells

For each channel $d \in \{\theta, I, RG, BY, \sigma\}$ we use a one-dimensional space \mathbb{R} to encode the features with $i \in N$ units at each location \mathbf{x} . V4in units receive input from 5 channels (d): $r_{\theta,i,\mathbf{x}}$ for orientation, $r_{I,i,\mathbf{x}}$ for intensity, $r_{RG,i,\mathbf{x}}$ for red–green opponency, $r_{BY,i,\mathbf{x}}$ for blue–yellow opponency and $r_{\sigma,i,\mathbf{x}}$ for spatial frequency.

$$\begin{aligned} \tau^{V4in} \frac{d}{dt} r_{d,i,\mathbf{x}}^{V4in} &= w \cdot r_{d,i,\mathbf{x}} + \max_{x'} w_{x,x'} \cdot r_{d,i,x'}^{V4in} \\ &- (r_{d,i,\mathbf{x}}^{V4in} + a \cdot z_{d,i,\mathbf{x}}^{\text{sat}}); \quad \tau^{V4in} = 0.01 \text{ s} \end{aligned} \quad (16)$$

$w_{x,x'} \cdot r_{d,i,x'}^{V4in}$ spatially pools the input response to approximate a specific receptive field size of the cells in V4in. This step is necessary since the input image is downsampled to the lowest center scale c used. We used a Gaussian ($\sigma = 6$) to determine

the weight $w_{x,x'}$ depending on the distance between x and x' in pixels. $z_{d,i,\mathbf{x}}^{\text{sat}}$ represents the neural saturation to simulate the phasic response after stimulus onset:

$$\tau^{\text{sat}} \frac{d}{dt} z_{d,i,\mathbf{x}}^{\text{sat}} = r_{d,i,\mathbf{x}}^{V4in} - z_{d,i,\mathbf{x}}^{\text{sat}}; \quad \tau^{\text{sat}} = 0.5 \text{ s}. \quad (17)$$

A.5. V4gain cells

A feature-specific top-down projection to V4gain cells originates in IT and a location-specific projection is received from the FEF movement cells.

$$\begin{aligned} \tau^{V4gain} \frac{d}{dt} r_{d,i,\mathbf{x}}^{V4gain} &= w \cdot r_{d,i,\mathbf{x}}^{V4in} + \Gamma(A - \max_i (r_{d,i,\mathbf{x}}^{V4gain})) \\ &\times \left(w^{\text{FEFm},V4} r_{d,i,\mathbf{x}}^{V4in} \cdot r_{\mathbf{x}}^{\text{FEFm}} + w^{\text{IT},V4} \max_{\mathbf{x}^{\text{IT}}} (r_{d,i,\mathbf{x}}^{V4in} \cdot r_{d,i,\mathbf{x}^{\text{IT}}}^{\text{IT}}) \right) \\ &- w_{\text{inh}} \left(r_{d,i,\mathbf{x}}^{V4gain} + 1 \right) \cdot \sum_j r_{d,j,\mathbf{x}}^{V4gain} - w_{\text{inh}}^{\text{map}} r_d^{\text{map}} \end{aligned} \quad (18)$$

r_d^{map} is an inhibitory unit which receives its input from all cells in the map.

Parameters used for the above equations are : $\tau^{V4gain} = 0.01 \text{ s}$; $w = 0.1$; $w^{\text{FEFm},V4} = 50$; $w^{\text{IT},V4} = 4$; $w_{\text{inh}} = \frac{1}{\#\mathbf{i}}$; $w_{\text{inh}}^{\text{map}} = \frac{1.5}{\#\mathbf{x}}$.

A.6. V4pool cells

An efficient processing within a population code requires that the conspicuity of each feature is adequately combined across hierarchy levels. Due to the increase in RF size, the conspicuity of features from different locations converges onto a single location. A weighted sum of the conspicuity of identical features across space shows a multiplicity effect: increasing the number of identical stimuli within a RF enhances the conspicuity (Hamker, 2005c). We suggested to separate between the content and relevance of a stimulus independent of the number of stimuli using a maximum pooling. We have shown that such a max-pooling allows to reproduce the data of Reynolds et al. (1999) showing the influence of attention on the competition within a receptive field (Hamker, 2004).

$$\begin{aligned} \tau^{V4pool} \frac{d}{dt} r_{d,i,\mathbf{x}}^{V4pool} &= \max_{x'} w_{x,x'} \cdot r_{d,i,x'}^{V4gain} - w_{\text{inh}} \left(r_{d,i,\mathbf{x}}^{V4pool} + 1 \right) \\ &\cdot \sum_j r_{d,j,\mathbf{x}}^{V4pool} - w_{\text{inh}}^{\text{map}} r_d^{\text{map}} \end{aligned} \quad (19)$$

r_d^{map} is an inhibitory unit which receives its input from all cells in the map. $w_{x,x'}$ is determined by a Gaussian ($\sigma = 10$) depending on the distance between x and x' in pixels.

Parameters used for the above equations are : $\tau^{V4pool} = 0.01 \text{ s}$; $w = 0.4$; $w_{\text{inh}} = \frac{1}{\#\mathbf{i}}$; $w_{\text{inh}}^{\text{map}} = \frac{1.5}{\#\mathbf{x}}$.

A.7. IT cells

The features with their respective conspicuity and location in V4pool project to IT, but only within the same dimension d .

The complexity of the encoded features does not increase. The size of IT is fixed to a 3×3 map of cells with overlapping receptive fields.

$$\begin{aligned} \tau^{\text{IT}} \frac{d}{dt} r_{d,i,x}^{\text{IT}} = & \max_{i,x' \in \text{RF}(x)} \left(F(w \cdot r_{d,i,x'}^{\text{V4pool}}) \right) \\ & + \Gamma(A - \max_i(r_{d,i,x}^{\text{IT}})) \\ & \cdot \left(w^{\text{FEFm,IT}} \max_{i,x' \in \text{RF}(x)} \left(F(r_{d,i,x'}^{\text{V4pool}}) \cdot r_{x'}^{\text{FEFm}} \right) \right. \\ & + w^{\text{PFmem,IT}} \max_{i,x' \in \text{RF}(x)} \left(F(r_{d,i,x'}^{\text{V4pool}}) \cdot r_{d,i}^{\text{PFmem}} \right) \\ & \left. - r_{d,i,x}^{\text{IT}} \cdot \left(w_{\text{inh}} \sum_j r_{d,j,x}^{\text{IT}}(t) + w_{\text{inh}}^{\text{map}} r_d^{\text{map}}(t) \right) \right) \end{aligned} \quad (20)$$

$F(r_{d,i,x'}^{\text{V4pool}}) = r_{d,i,x'}^{\text{V4pool}} \cdot g(\|\mathbf{u}_i^{\text{IT}} - \mathbf{u}_i^{\text{V4pool}}\|)$ is a function that weights the argument according to the feature similarity, where g is a Gaussian.

$$\tau^{\text{IT}} = 0.01 \text{ s}; w_{\text{inh}} = \frac{3}{\#i}; w_{\text{inh}}^{\text{map}} = \frac{3}{\#\mathbf{x}}; w = 0.4; w^{\text{PFmem,IT}} = 5; w^{\text{FEFm,IT}} = 10.$$

A.8. FEF visuomovement cells

The visuomovement cells receive afferents from V4in, and V4pool at the same retinotopic location irrespective of the feature information and thus, they encode the conspicuity of locations or often referred to as saliency. Inhibition of return suppresses the saliency at locations that have been visited recently. We define an additional influence from the working memory to further bias those locations that match the target template in all channels.

$$\begin{aligned} \tau^{\text{FEFv}} \frac{d}{dt} r_{d,i,x}^{\text{FEFv}} = & \sum_d \left(w^{\text{V4in,FEFv}} \max_i r_{d,i,x}^{\text{V4in}} \right. \\ & + w^{\text{V4pool,FEFv}} \max_{i,x' \in \text{RF}(x)} r_{d,i,x'}^{\text{V4pool}} \\ & + w^{\text{PFmem}} \prod_d \max_{i,x' \in \text{RF}(x)} r_{d,i}^{\text{PFmem}} \cdot r_{d,i,x'}^{\text{V4pool}} \\ & + \sum_{x'} w_{x,x'} r_{x'}^{\text{FEFv}} - w^{\text{IOR}} r_x^{\text{IOR}} \\ & \left. - r_x^{\text{FEFv}} (w_{\text{inh}} \max_x r_x^{\text{FEFv}} + w_{\text{inh}}^{\text{map}} z^{\text{map}}) \right) \end{aligned} \quad (21)$$

$$\tau^{\text{FEFv}} = 0.01 \text{ s}; w_{x,x'} = 0.25 \cdot \exp\left(\frac{(x-x')^2}{0.004}\right); w^{\text{V4in}} = 0.05; w^{\text{V4pool}} = 0.15; w^{\text{IOR}} = 1; w_{\text{inh}} = 2; w_{\text{inh}}^{\text{map}} = \frac{3}{\#\mathbf{x}}; w^{\text{PFmem}} = 50.$$

A.9. FEF movement cells

The movement map reads out the saliency and transfers it to a target location for the planned eye movement, which is also used as a feedback signal for gain enhancement in V4 and IT. This is consistent with several findings indicating a strong

overlap between spatial attention and eye movements (Deubel & Schneider, 1996; Kowler et al., 1995; Kustov & Robinson, 1996; Rizzolatti et al., 1987). Please refer to Hamker (2005a) for a discussion about the origin of spatial attention and eye movement planning. If the activation exceeds a threshold an eye movement is triggered. The movement cells are inhibited by fixation cells.

$$\begin{aligned} \tau^{\text{FEFm}} \frac{d}{dt} r_x^{\text{FEFm}} = & w^{\text{FEFv,FEFm}} r_x^{\text{FEFv}} - w_{\text{inh}}^{\text{v}} \sum_x r_x^{\text{FEFv}} \\ & + \sum_{x'} w_{x,x'} r_{x'}^{\text{FEFm}} - w^{\text{f}} r^{\text{f}} \\ & - r_x^{\text{FEFm}} w_{\text{inh}}^{\text{map}} \sum_x r_x^{\text{FEFm}} \end{aligned} \quad (22)$$

$$\tau^{\text{m}} = 0.02 \text{ s}; w_{x,x'} = 1.3 \cdot \exp\left(\frac{(x-x')^2}{0.002}\right); w^{\text{v}} = 0.3; w^{\text{f}} = 0.7; w_{\text{inh}}^{\text{map}} = 0.075; w_{\text{inh}}^{\text{v}} = \frac{0.4}{\#\mathbf{x}}.$$

If the expectation exceeds the threshold Γ_o^{m} at the time t_o , we calculate the center of gravity to indicate the location of an eye movement. Thus, in cases with a split of attention the overt shift differs from the covert shift.

$$\mathbf{x}_c = \frac{\sum_x r_x^{\text{FEFm}}(t_o) \cdot \mathbf{x}}{\sum_x r_x^{\text{FEFm}}(t_o)}. \quad (23)$$

A.10. Fixation unit

Some tasks demand an eye movement only when a target is in the scene, but not when the scene contains only distractors. Thus, we define a fixation unit, that is under control of a very simple cognitive process (r^{c}). In the simulations here, we set $r^{\text{c}} = 10$ for the first 330 ms after probe onset. The fixation unit also resets the movement units: after the expectation in the movement map exceeds the threshold Γ_o^{m} and thus an eye movement is initiated at the time t_o , the fixation unit gets activated for a brief period T^{SAC} .

$$\begin{aligned} \tau^{\text{f}} \frac{d}{dt} r^{\text{f}} = & w^{\text{m}} I^{\text{m}} + w^{\text{c}} r^{\text{c}} - r^{\text{f}} \\ I^{\text{m}} = & \begin{cases} 1 & \text{if } r^{\text{m}}(t_o) > \Gamma_o^{\text{m}} \& t < t_o + T^{\text{SAC}} \\ 0 & \text{else} \end{cases} \end{aligned} \quad (24)$$

$$\tau^{\text{f}} = 0.012 \text{ s}; T^{\text{SAC}} = 50 \text{ ms}; \Gamma_o^{\text{m}} = 0.8; w^{\text{m}} = 4; w^{\text{c}} = 0.6.$$

A.11. Inhibition of return (IOR)

The IOR map serves as a buffer to memorize recently visited locations. Recently visited locations are overt and covert shifts of spatial attention. We regard each location \mathbf{x} as inspected, dependent on the selection of an eye movement at time t_o and location \mathbf{x}_c or when the attended item at location \mathbf{x}_m does not sufficiently match the target template. The latter case is calculated in the control units and expressed by the variable I^{c} . In this case the IOR cells are charged at the location of the highest expectation in the movement map for a period of time

T^{IOR} . The IOR buffer slowly decays with a low weight w_{inh} .

$$\begin{aligned} \tau^{\text{IOR}} \frac{d}{dt} r^{\text{IOR}} &= (1 - r_x^{\text{IOR}})(I_x^{\text{SC}} + w^m I_x^m \cdot I^c) - w_{\text{inh}} r_x^{\text{IOR}} \\ I_x^{\text{SC}} &= \begin{cases} \exp\left(-\frac{(\mathbf{x} - \mathbf{x}_c)^2}{0.01}\right) & \text{if } t < t_o + T^{\text{IOR}} \\ 0 & \text{else} \end{cases} \\ I_x^m &= \exp\left(-\frac{(\mathbf{x} - \mathbf{x}_m)^2}{0.01}\right) \quad r_{\mathbf{x}_m}^m = \max_{\mathbf{x}}(r_{\mathbf{x}}^m) \quad (25) \\ \tau^{\text{IOR}} &= 0.01 \text{ s}; w^m = 1; w_{\text{inh}} = 0.02; \Gamma_o^m = 0.8; T^{\text{IOR}} = 50 \text{ ms}; \Gamma_c^m = 0.4. \end{aligned}$$

A.12. PFMem

We model a simple recurrent local circuit for working memory to encode the target template. The memorization of a pattern is achieved through recurrent excitation. Whether a pattern should be memorized depends on the task. The variable $I^{\text{store}}(t) \in \{0, 1\}$ determines when a pattern should be memorized. It is set externally according to the task instruction. If a pattern is memorized ($r_{d,j}^{\text{T}}$ is high), the term $\Gamma(r_{\Gamma_{\text{mem}}} - w^{\text{cue}} \max_j r_{d,j}^{\text{T}})$ ensures that no other stimulus in level II can penetrate the memory. In the experiments simulated here we defined the template of the saccade target externally with I_d^{Target} . r^{mem} indicating by an activity of one that a pattern is in memory.

$$\begin{aligned} \tau^{\text{PFmem}} \frac{d}{dt} r^{\text{PFmem}} &= \Gamma(\Gamma_{\text{mem}} - \max_j r_{d,j}^{\text{PFmem}}) \max_{\mathbf{x}} \Gamma(r_{d,i,\mathbf{x}}^{\text{IT}} - C) \\ &+ I_{d,i}^{\text{Target}} + \sum_j w_{ij} r_{d,j}^{\text{PFmem}} - r_{d,i}^{\text{PFmem}} \cdot w_{\text{inh}} \sum_j r_{d,j}^{\text{PFmem}} \\ &+ (0.7 - 0.6 \cdot S_d \cdot I^{\text{store}}) z_d \quad (26) \end{aligned}$$

$$\tau^{\text{PFmem}} = 0.012 \text{ s.}$$

For controlling the memorization and deletion we define the following variables:

$$\begin{aligned} r_d^{\text{mem}}(t) &:= \begin{cases} 1 & \text{if } \max_i(r_{d,i}^{\text{PFmem}}) > \Gamma^{\text{mem}} \\ 0 & \text{else} \end{cases} \\ S_d &= 1 \quad \text{if } \max(r_{d,i}^{\text{PFmem}}) > 0.1 \\ S_d &= 0 \quad \text{if } I^{\text{mem}} = 0 \text{ \& } r_d^{\text{mem}} = 0. \quad (27) \end{aligned}$$

The lateral weights w_{ij} are computed from a Gaussian with

$$w_{ij} = 0.45 \cdot \exp\left(\frac{\min((i-j)^2, (I-i-j)^2)}{0.005}\right) \text{ for orientation}$$

$$w_{ij} = 0.45 \cdot \exp\left(\frac{(i-j)^2}{0.005}\right) \text{ for other dimensions}$$

$$w_{\text{inh}} = 0.25; C = 0.05; \Gamma_{\text{mem}} = 0.35.$$

References

Armstrong, K. M., Fitzgerald, J. K., & Moore, T. (2006). Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron*, *50*, 791–798.

- Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*, 529–534.
- Bisley, J. W., & Goldberg, M. E. (2006). Neural correlates of attention and distractability in the lateral intraparietal area. *Journal of Neurophysiology*, *95*, 1696–1717.
- Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1991). Visual topography of area TEO in the macaque. *Journal of Comparative Neurology*, *306*, 554–575.
- Burt, P. J., & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, *3*, 532–540.
- Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, *7*, 308–313.
- Compte, A., & Wang, X. J. (2006). Tuning curve shift by attention modulation in cortical neurons: A computational study of its mechanisms. *Cerebral Cortex*, *16*, 761–778.
- Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *Journal of Neuroscience*, *17*, 3201–3214.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*, 1827–1837.
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*, 193–202.
- Hamker, F. H. (2003). The reentry hypothesis: Linking eye movements to visual perception. *Journal of Vision*, *11*, 808–816.
- Hamker, F. H. (2004). Predictions of a model of spatial attention using sum- and max-pooling functions. *Neurocomputing*, *56*, 329–343.
- Hamker, F. H. (2005a). The reentry hypothesis: The putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebral Cortex*, *15*, 431–447.
- Hamker, F. H. (2005b). Modeling attention: From computational neuroscience to computer vision. In L. Paletta, et al., (Eds.), *LNCS: Vol. 3368. Attention and performance in computational vision, second international workshop on attention and performance in computer vision* (pp. 118–132). Berlin, Heidelberg: Springer-Verlag.
- Hamker, F. H. (2005c). The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision. *Journal for Computer Vision and Image Understanding*, *100*, 64–106.
- Hamker, F. H. (2005d). A computational model of visual stability and change detection during eye movements in real world scenes. *Visual Cognition*, *12*, 1161–1176.
- Hamker, F. H. Modeling feature-based attention as an active top-down inference process. *BioSystems* (in press).
- Ignashchenkova, A., Dicke, P. W., Haarmeier, T., & Thier, P. (2004). Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. *Nature Neuroscience*, *7*, 56–64.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 1254–1259.
- Juan, C. H., Shorter-Jacobi, S. M., & Schall, J. D. (2004). Dissociation of spatial attention and saccade preparation. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 15541–15544.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Psychology*, *4*, 219–227.
- Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*, 1897–1916.
- Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, *384*, 74–77.
- 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*, 9596–9601.
- Mazer, J. A., & Gallant, J. L. (2003). Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron*, *40*, 1241–1250.

- McAdams, C. J., & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience*, *19*, 431–441.
- Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, *421*, 370–373.
- Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*, 782–784.
- Muller, J. R., Philiastides, M. G., & Newsome, W. T. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 524–529.
- Ogawa, T., & Komatsu, H. (2004). Target selection in area V4 during a multidimensional visual search task. *Journal of Neuroscience*, *24*, 6371–6382.
- Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160–174.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.
- Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanism subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*, 1736–1753.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.
- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologica*, *25*, 31–40.
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway?. *Trends in Cognitive Sciences*, *8*, 363–370.
- Schall, J. D. (2002). The neural selection and control of saccades by the frontal eye field. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *357*, 1073–1082.
- Spratling, M. W. (2005). Learning viewpoint invariant perceptual representations from cluttered images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*, 753–761.
- Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., & Schiller, P. H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron*, *29*, 757–767.
- Thompson, K. G., & Schall, J. D. (2000). Antecedents and correlates of visual detection and awareness in macaque prefrontal cortex. *Vision Research*, *40*, 1523–1538.
- Thompson, K. G., Biscoe, K. L., & Sato, T. R. (2005). Neuronal basis of covert spatial attention in the frontal eye field. *Journal of Neuroscience*, *25*, 9479–9487.
- Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238.
- Wurtz, R. H., Sommer, M. A., & Cavanaugh, J. (2005). Drivers from the deep: The contribution of collicular input to thalamocortical processing. *Progress in Brain Research*, *149*, 207–225.
- Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*, 72–75.