

## **A computational model of visual stability and change detection during eye movements in real-world scenes**

Fred H. Hamker

*Allgemeine Psychologie, Psychologisches Institut II, Westfälische  
Wilhelms-Universität, Münster, Germany*

Visual cognition and change detection critically depend on attention and gaze. A central question is the exact involvement of attention in change detection. This paper proposes a detailed computational model to illuminate the nature of visual stability and change detection during eye movements. Consistent with experimental findings the model predicts that change detection at the saccade endpoint is relatively easy. According to our model the planning of an eye movement provides an enhanced sensory processing of an object at the saccade target. This facilitates the memorization of that object. A comparison of the memorized target template with the representation after saccade allows for change detection. We hypothesize that the spatial reentry signal that emerges from planning an eye movement could be a crucial neural basis of visual stability.

Change blindness describes the difficulty of humans to notice changes in scenes. Most studies agree that focused attention overcomes this blindness to changes (for an overview see Rensink, 2002). Henderson and Hollingworth investigated change detection to saccade targets and found a high detection rate for saccade target deletions. According to the saccade target theory of visual stability the object at the saccade target is special because its location and features have been selected and memorized prior to the saccade (Currie, McConkie, Carlson-Radvansky, & Irwin, 2000; McConkie & Currie, 1996). The saccade target theory has been contrasted to extraretinal based theories which assume that an extraretinal eye position signal is used to remap all retinal locations. In extension to the saccade target theory, Henderson and Hollingworth suggested a visual memory theory to explain scene memory across saccades. Regarding change detection they suggested the following processes: Attention is allocated to the saccade target. Therefore, sensory

---

Please address all correspondence to: F. H. Hamker, Allgemeine Psychologie, Psychologisches Institut II, Westfälische Wilhelms-Universität, Fliehdnerstrasse 21, 48149 Münster, Germany. Email: fhamker@uni-muenster.de

This research was supported by DFG HA2630/2-1 and in part by the ERC Program of the NSF (EEC-9402726). I thank M. Lappe for helpful comments.

processing of the saccade target is gated and an internal representation of the stimulus in short-term memory is facilitated. During the saccade the sensory activity declines. Once the eyes reach the target vision starts with the process of comparing the stored representation with the new incoming signals. If no match between both occurs an error signal is generated.

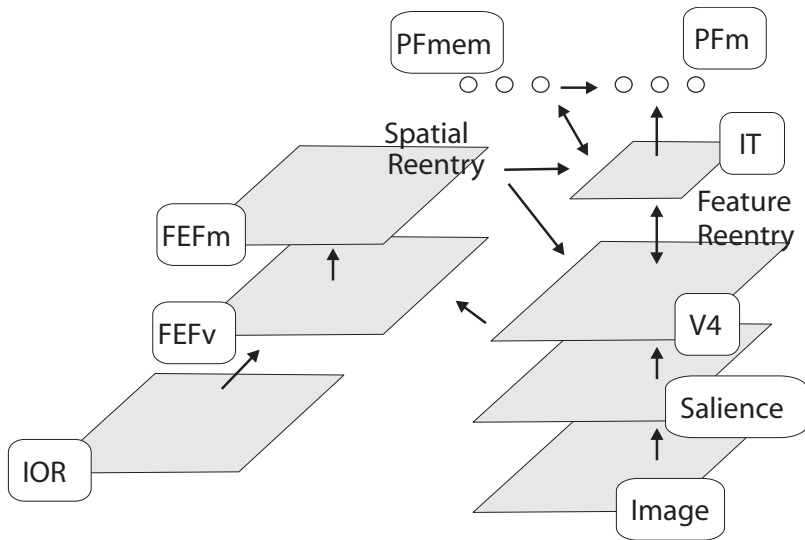
In this paper we investigate visual stability and change detection by means of a computational approach. In the above-mentioned theories the term “attention” is not clearly defined and the specific role of attention in change detection has not been described in detail. For example, attention in the saccade target theory is important to identify the saccade target as special. It is well known that the eye movement is associated with attention (Deubel & Schneider, 1996; Kowler, Anderson, Doshier, & Blaser, 1995; Rizzolatti, Riggio, & Sheliga, 1994) but it is not clear if attention is a separate but linked process or if attention emerges due to an extraretinal eye position signal. If the latter is the case the saccade target theory would be less different to the extraretinal based theories.

Several computational models have emphasized different aspects of eye movement control and some of them have been demonstrated in real-world scenes (Backer, Mertsching, & Bollmann, 2001; Findlay & Walker, 1999; Itti & Koch, 2000; Krieger, Rentschler, Hauske, Schill, & Zetsche, 2000; Rao, Zelinsky, Hayhoe, & Ballard, 2002). In all of these models the process that defines gaze is separated from the process that analyses a scene near gaze point.

Our model (see Appendix) is unique with respect to how perception is influenced by the planning of an eye movement (Hamker, 2003). It has been successfully shown to be consistent with electrophysiological and behavioural findings in attention tasks (Hamker, 2003; Hamker, 2004a) and recently extended to natural scene perception (Hamker, 2005; Hamker & Worcester, 2002). According to our model oculomotor areas responsible for planning an eye movement, such as the frontal eye field, influence perception prior to the eye movement. The activity reflecting the planning of an eye movement reenters the ventral pathway and sensitizes all cells within the movement field. We use this model to simulate a change detection task that is comparable to the one used by Henderson and Hollingworth (2003). Consistent with the saccade target theory the model predicts that a global remapping is not necessary for visual stability. However, since the spatial reentry signal, which is crucial for the emergence of attention in our model, can be defined as an extraretinal eye position signal, our model suggests that the saccade target theory is in part related to the extraretinal based theories.

## THE MODEL

The model (Figure 1) 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



**Figure 1.** Model for visual perception of natural scenes. First, information about the content and its low level stimulus-driven saliency is extracted. This information is sent further upwards to V4 and to IT cells that are broadly tuned to location. The target template is encoded in PF memory (PFmem) cells. PF match (PFm) cells indicate by comparison of PFmem with IT whether the target is actively encoded in IT. Feedback from PFmem to IT increases the strength of all features in IT matching the template. Feedback from IT to V4 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 increasing activity of a FEF movement (FEFm) cell is sent into extrastriate visual areas and facilitates the processing within its movement field. The IOR map memorizes recently visited locations and inhibits the FEF visuomovement cells.

(Baddeley, 1986). In addition, we simulate a map for the inhibition of return (IOR) which might be located in posterior parietal areas.

We investigate the putative interplay of the saccade system onto visual perception and memory in a change detection task. Change detection and transsaccadic scene perception requires the storage of visual information in visual short-term memory (VSTM; Irwin, 1992) and the use of the memorized pattern to indicate the change.

If we present a visual scene to the model, features such as colour, intensity, and orientation are computed in different V4 maps. The fact that features that are unique in their environment “pop-out” is to a first degree achieved by computing center-surround differences. V4 cells compete for representing their encoded stimuli. This competition is biased by feature-specific reentry from IT cells and spatial reentry from the frontal eye field movement cells. The growing receptive field size along the processing hierarchy requires that a number of V4

cells encoding identical stimuli at different locations project to a single IT cell. For simplicity the type of features in V4 and IT is identical. IT receives feature-specific reentry from the prefrontal memory cells and location-specific reentry from the frontal eye field movement cells (Hamker, 2003).

Initially, the prefrontal memory is empty. The model is set to memorize a pattern defined by the cells with the strongest activity in IT just prior to each saccade. At this time the dynamic representation in IT is strongly influenced by the oculomotor system, since the movement cell activity provides a delayed reentry signal to extrastriate areas such as V4 and IT. The assumption of such a reentry signal is supported by the finding that FEF microstimulation modulates V4 activity (Moore & Armstrong, 2003).

The planning of the eye movement is implemented as follows. The FEF visuomovement neurons receive convergent afferents from V4 and IT and add up the activity across all dimensions (colour, intensity, orientation). The information from the target template hold in PF memory 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 represent the saliency and task relevance of locations. The effect of the FEF visuomovement cells on the FEF movement cells 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 centre of gravity of movement cell activity.

There is currently no clear indication where cells that implement a dynamic memory for an inhibition of return are located. We assume that each eye movement triggers a charge of IOR cells at the location of saccade target. This causes a suppression of the recent saccade target in the visuomovement cells. IOR cells slowly decay.

The model “prefrontal cortex” serves for two major functions, memorizing a pattern in PF memory cells and indicating a match of the incoming pattern with the memorized pattern in PF match cells. Cells that memorize patterns are known to exist in prefrontal areas. In studies using a delayed match-to-sample task elevated firing rates within specific populations have been reported during the delay (Miller, Erickson, & Desimone, 1996). The activity of such populations in the prefrontal cortex is robust to interfering stimuli. Other prefrontal cells indicate a match detection in a delayed match-to-category task (Freedman, Riesenhuber, Poggio, & Miller, 2002), which is similar to our PF match cells. The capacity of transsaccadic memory has been shown to comprise about 3–4 items and its content is slow-decaying (Irwin, 1992). An item seems to be an integrated object representation, which comprises identity, colour, and location (Irwin, 1992). The present model capacity is only a single item. This simplification does not allow us to make predictions about the circumstances

that may define the size of memory. However, it does allow us to simulate perception of an object at the saccade target.

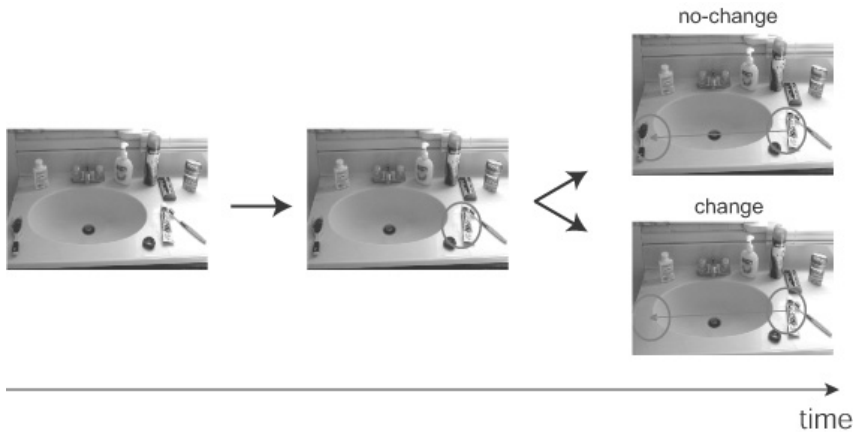
This model of visual perception has been developed to be consistent with a range of electrophysiological findings. The model “V4 area” (Hamker, 2004b) has been demonstrated to quantitatively account for receptive field competition in V4 (Reynolds, Chelazzi, & Desimone, 1999). A simplified version of the proposed systems model (Hamker, 2003) has been shown to match the time course of IT and V4 activity in visual search (Chelazzi, Duncan, Miller, & Desimone, 1998). The model is also consistent with findings in the FEF (Schall, 2002) and psychophysical data (Hamker & vanRullen, 2002).

We do not explicitly model the properties that emerge from the physical eye movement. During this period vision is impaired. V4 cells would receive input from a smeared retinal image and after the eyes land the new retinal image would mask the most recent representation. In an overt visual search task with a homogeneous background V4 cells have been shown to persist activity after saccade onset but the firing rate declines within 100 ms (Chelazzi, Miller, Duncan, & Desimone, 2001). Due to the large receptive fields and late position in the hierarchy, IT cells maintain activity during the movement and increase their rate after the eyes reach the target (Chelazzi et al., 1998). It seems that perception during a saccade is not necessarily interrupted. Although once the eye started sweeping across the scene the image in early visual areas is smeared and masked with the postsaccadic image, late processing areas could maintain the representation that is present just prior to the saccade in order to support a continuous percept.

The high foveal resolution of the object at the target location after the saccade allows a preferred processing (Rolls & Tovee, 1995). Any form of spatial attention that might be present initially after the eyes land cannot be dissociated from this fovea effect. Since we do not model the retinal resolution and cortical magnification we use the decline phase of the reentry signal to simulate the condition after eye movement. The decreasing reentry signal enhances sensory processing of the object at the target location for a brief period until it vanishes.

## EXPERIMENT

Henderson and Hollingworth (2003) instructed participants to observe complex scenes to prepare for a later memory test. Participants had been told that they were shown two scenes: One original and a slightly modified version. They would have to indicate whether the original and the modified scene differ. More important is the additional instruction according to which they have to immediately report object changes while studying the original scene. Here we are only interested in the deletion condition. Objects that were deleted had been placed in an uncluttered region of the scene. If a subject's gaze crossed a boundary around such an object it was removed by displaying a changed scene.

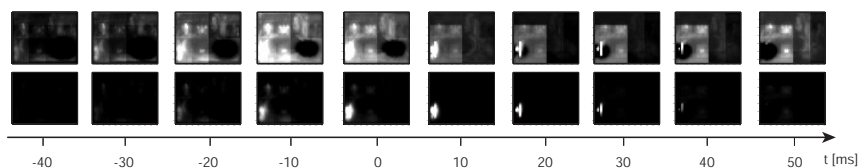


**Figure 2.** Experimental paradigm to simulate change detection of objects at the saccade target. A natural scene is presented to the model. Initially, the selection of the saccade target location is only guided by stimulus-driven saliency. The most salient location is at the toothpaste. In the present simulations we use the second saccade, which is directed towards the hairbrush to investigate change detection. At the time the FEF movement cells indicate the onset of a saccade the scene either remains identical (no-change condition) or the hairbrush is removed (change condition). The activity within the model is not reset during the change of the scenes.

This experiment is comparable to a free viewing task, in which the model's gaze is only driven by the stimulus saliency. The model scans the scene by visiting one object after another (Hamker & Worcester, 2002). We set up the model to memorize the strongest IT pattern in PF memory cells just prior to each saccade. We use this model's behaviour for change detection (Figure 2). Once a location is selected for an eye movement we either do not change the scene or we remove the object at the saccade target. Thus, we receive simulation data for no-change and change trials. The PF match cells indicate if the model finds the memorized pattern in the scene.

## RESULTS

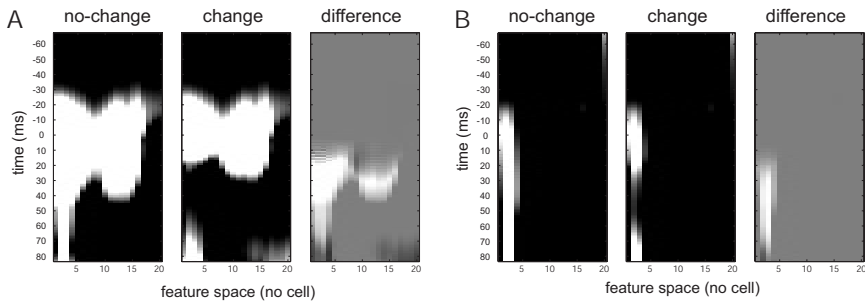
We first demonstrate how the model operates in a free viewing task and then demonstrate change detection. Figure 3 shows the FEF visuomovement and movement cell activity around the time the model selects the hairbrush. FEF visuomovement activity can be interpreted as a measure of the strength of processing at each location since visuomovement cells receive their main input from V4 and IT and average across all dimensions. Prior to each saccade all locations with objects are processed in parallel and represent potential targets. The movement cells read out the information from the visuomovement cells and select the target location for the eye movement. Thirty ms prior to saccade, the



**Figure 3.** Activity of the FEF visuomovement (top) and movement (bottom) cells around the time of saccade begin (0 ms). Each image is a snapshot of the activity at the time indicated on the axis. The space within each image refers to the scene shown in Figure 2. The planning of the saccade is a dynamic process. Visuomovement cells reflect areas of high processing in V4 and IT and thus potential target locations for saccadic eye movements. The area around the location of the most recent eye movement is inhibited by cells from the IOR map. Movement cells determine the saccade target and provide a reentry signal to enhance sensory processing in V4 and IT. Spatial reentry emerges earlier than 30 ms prior to the saccade and is most prominent at the beginning of the eye movement.

model shows a preferred processing around the planned target location. This effect emerges from the reentry of movement cell activity into V4 and IT. After target selection the IOR cells get charged by the movement cells and inhibit visuomovement cells with receptive fields around the target location. The movement cell activity gets suppressed by FEF fixation cells, that are activated after the eye movement.

We have explained the mechanisms of overt and covert attention in the model and demonstrated the model's performance. Regarding change detection the model has to recognize whether the object at the target location is removed or not. As we have explained in the model section the PF memory cells memorize the strongest pattern active in IT just prior to saccade selection. Objects at the target location are most likely to be memorized since the reentry signal enhances their sensory processing (Hamker, 2004a). Once a pattern is in short-term memory it biases competition among IT cells by a feature-specific reentry signal. This form of feature-based attention represents a search process for a specific pattern in the scene. PF match cells indicate by their activity if the search template held in PF memory cells is currently encoded in the scene (Figure 1). We expect that the removal of an object results in a decline of activity in IT cells which encode properties of this object. Figure 4A shows this for the intensity channel in the model. The dark areas of the hairbrush located in the lower left part of the scene (Figure 2) are represented by cells on the left side of the feature space (cells 2–3 in Figure 4A). Although the receptive fields of IT cells cover the full scene the model dynamics at the time of the eye movement ensures that the removal of the hairbrush results in a decline of representative IT cells, even if the scene contains other dark areas elsewhere. The higher firing of IT cells (Figure 4A) in the no-change condition relative to the change condition is reflected by light areas in the difference signal. Thus, the visual persistence during the eye movement encoded by IT cells reflects the recently attended item.



**Figure 4.** Comparison of IT and PF match cell activity in the no-change and change condition. (A) IT cell activity in the intensity channel. The model encodes intensity information by a population of 20 cells. Cells with lower numbers are selective for dark patches and cells with high numbers for light areas. We simulate nine populations of IT cells with different receptive fields (RF). In order the present IT activity across the full scene we only plot the maximal activity regardless of RF location. 50 ms prior to saccade IT activity is very low since inhibitory cell populations with a larger time constant suppressed IT activity after a strong firing due to the most recent eye movement. IT cells increase again their activity and initially encode the scene in parallel. Due to the spatial reentry signal they focus their processing at the saccade target by suppressing IT cells with RFs that do not contain the saccade target (not shown). The removal of the object results in a decline of the activity of cells encoding the objects properties. By taking a difference between the no-change and change conditions we see that the no-change condition is reflected by a stronger activity of several IT cells (grey indicates no difference). (B) PF match cell activity in the intensity channel. Despite the obvious difference in IT activity, change detection is performed by searching for a target template hold in PF memory cells. In this case the model selects the dark patches of the hairbrush as a search template since the IT cells encoding this feature fired most prominent prior to the eye movement. After the eye movement the model continues to search for this pattern. PF match cells indicate a successful search by their activity since they only fire by the joint activity of a PF memory and respective IT cell.

The detection of a change, however, has to be an active process, which requires the match of IT activity with the memorized target template hold in PF memory cells. In the intensity channel the model has memorized a pattern that represents the dark spots of the hairbrush. Thus, in the no-change condition the PF match cells (Figure 4B) continue to fire around the time of the eye movement but in the change condition their activity declines.

## PREDICTIONS

First of all the model predicts that an object change at the saccade target location can be well detected. The reentry signal that necessarily occurs at the saccade target location enhances sensory processing of objects in its vicinity and increases their probability to be memorized. The model also predicts that object properties can be reported when the scene disappears before the eyes land at the target location. Irwin & Zelinsky (2002) found a high accuracy of report even



when the object had never been foveated. Thus, object identification begins before the eyes actually fixate on the object.

Visual stability might be achieved through continuity. Before the eye movement spatial reentry generates a representation that prepares for the one that can be expected after the eye movement. Processing capacity is restricted to the same issues just prior, during and just after the eye movement. Thus, parallel changes elsewhere in the scene are likely to go unnoticed.

The model also predicts that change detection is an active process that requires to actively maintain a search template while the eyes sweep across the scene. Thus, eye movements recruit resources that could interfere with other tasks to be done in parallel. For example, it has been shown that mental rotation (Irwin & Brockmole, 2000) and shifts between global and local judgements (Brockmole, Carlson, & Irwin, 2002) are impaired during saccadic eye movements.

For simplicity the model searches for the same pattern it has memorized prior to the saccade. One can only speculate about the exact templates used for identification. It is likely that the brain generates a prediction instead of a copy. This makes it difficult to predict the search template used in experiments. Henderson and Hollingworth (2003) investigated the effect of deletion as well as type and token changes. The detection rates for type and token changes decrease with the length of the saccade. Since the spatial resolution is very poor in the periphery, and predictions are difficult to make based on very little information, we would expect a slower decline if changes do not affect low frequency components and the sign of contrast.

## DISCUSSION

The perception of a visually stable environment during eye movements is an ongoing debate. Although perception during eye movements is strongly impaired we do not perceive the world in snapshots. Von Helmholtz (1896) argued that image motion during the eye movement could be sensed to compensate for the shift in image position. Extraretinal theories postulate an extraretinal signal that remaps the full scene during each saccade in order to achieve a space-constant representation. It was suggested that the motor-related signal of saccades can be sent back as a “corollary discharge” (Sperry, 1950) or “efference copy” (von Holst & Mittelstaedt, 1950). However, the origin and type of such an extraretinal signal has remained unclear. Perisaccadic mislocalization experiments in total darkness lead to a hypothetical eye-position signal that starts to rise 200 ms before the saccade begins and ends after the eyes have landed (Schlag & Schlag-Rey, 2002). Remapping of receptive fields has been found in LIP (Kusunoki & Goldberg, 2003), the superior colliculus (Walker, Fitzgibbon, & Goldberg, 1995), and the frontal eye field (Umeno & Goldberg, 1997). However, it is not clear if this predictive shift of receptive

fields contributes to visual stability or primarily updates the oculomotor system.

Alternatively to the visual remapping hypothesis a link between the representation before and after the saccade could be established by the saccadic target object or nearby landmarks (Currie et al., 2000; Deubel, Schneider, & Bridgeman, 2002; McConkie & Currie, 1996). Locating this object at the onset of the new fixation would be sufficient for relating the contents of the two representations with each other (Bridgeman, van der Heijden, & Velichkovsky, 1994).

Our model is consistent with the saccade target theory of visual stability. However, since the spatial reentry signal from the movement cells can be defined as an extraretinal eye position signal our model indicates a possible relation of the saccade target theory to extraretinal theories. We predict a modulatory signal that is directed to the saccade target. This signal could be responsible for the reported mislocalization of briefly flashed stimuli towards the endpoint of the saccade in illuminated rooms (Lappe, Awater, & Krekelberg, 2000; Ross, Morrone, & Burr, 1997).

In extension to a mere theory the presented model helps to reveal the underlying processes and illuminates those effects often referred to as attention. Early perception is primarily feedforward driven and meets high-level expectations while late perception emphasizes the features of interest and implements focal processing (Hamker, 2004a). We suggest that competition and cooperation implements a dynamic filter that allows the brain to connect planning processes with the physical world. The neural correlate of visual stability across eye movements could have its origin in spatial reentry. Perception just prior to a saccade is focused onto the saccade target and its vicinity and allows to make a hypothesis about the situation after saccade. The testing of this hypothesis links cognition before and after the saccade with each other which might be crucial for perceiving stability. This definition of vision as an active process does not require a stable space-based representation around the eye movement.

Regarding change detection the saccade target theory refers to attention for mediating the processes of object selection and memorization. Our model provides a testable explanation about the underlying neural processes. Spatial reentry (e.g., from the FEF movement cells) provides an enhanced sensory processing of the item at the saccade target. It becomes the object of interest, it is more likely to be memorized and its representation can be used to make an estimate of the situation after the eye movement. After the saccade the generated search template allows a comparison with a high-level representation such as in area IT. Similar as suggested by the coherence theory of Rensink (2000) our model demonstrates that an all-purpose representation is not necessary. The representation needed for the task at hand is dynamically created.

This idea receives further support from IT cell recordings during search for familiar objects in natural scenes. Sheinberg and Logothetis (2001) compared trials in which a familiar object was present but the eye movement was directed elsewhere with trials in which the monkey directed his eye towards the familiar

object. They found significantly different target-selective responses 95 ms before the eyes move. This provides strong evidence for our assumption that object analysis occurs even before the object is fixated. Present familiar objects that are not the goal of the next fixation show no perisaccadic activity. Our model explains this effect primarily based on the spatial reentry signal, which is under normal conditions directed to the saccade target location.

### Limitations of the model

We do not model the properties that emerge through the physical movement of the eyes and the decrease of spatial resolution from the fovea to the periphery. Thus, we are unable to make predictions about the period of time from the beginning of the eye movement to the new fixation and ignore this period in the simulation. We think that our assumptions about the conditions after saccade are plausible. The objects which appear in the fovea will receive an enhanced sensory processing such that an hypothesis testing is possible without engaging in a detailed search to locate the object of interest.

Our search template contains simple object features. Thus, change detection is only possible if these features are unique in the near environment of the saccade. The grouping of complex parts into some object description is not implemented in the present version. However, we have demonstrated that the dynamics of the model suppress the same features elsewhere in the scene. The limitation due to the simple feature spaces used does not seem to affect the line of arguments.

Studies suggest that approximately four visual objects or six spatial locations can be stored in VSTM (Luck & Vogel, 1997; Pashler, 1988). The present model makes very little assumptions about memory. We have implemented a memory with a single item capacity. However, even sophisticated models of spatial working memory are limited to a single entry (Compte, Brunel, Goldman-Rakic, & Wang, 2000). This limitation is inherent to simple continuous attractor neural networks that implement a population-based winner-take-all process (Amari & Arbib, 1977; Deneve, Latham, & Pouget, 1999). The assumption of a fixed number of storage places with a single item capacity is probably too simple, since the capacity varies due to experimental conditions. Alternatively short-term memory could store patterns with nested oscillations in which memory patterns repeat on each low-frequency (5–12 Hz) oscillation (Lisman & Idiart, 1995). If we would extend the model to have a capacity of four items it would still predict that the saccade target is special. However, it would allow us to investigate more subtle aspects of visual salience and task relevance of objects at other locations. Another factor that limits present neural models of working memory is that we did not understand sufficiently well how the brain defines an object as compared to a collection of features. Working memory seems to encode integrated objects (Luck & Vogel, 1997). Although our model of working memory is relatively simple it is sufficient to study the interaction of

perception and memory. In future versions it would be interesting to extend the model in order to provide predictions for the controversy about the amount of information that can be maintained during consecutive eye movements (Henderson & Hollingworth, 2003; Irwin & Zelinsky, 2002). Despite these limitations the presented model is a novel and very comprehensive computational account for explaining the role of covert and overt attention in change detection. Earlier computational models have typically focused on either recognition and attention or memory alone. Our model combines these three aspects of perception. It helps to explore the underlying neural mechanisms on the systems level that cannot be addressed at the level of more abstract theories.

## REFERENCES

- Amari, S., & Arbib, M. A. (1977). Competition and cooperation in neural nets. In J. Metzler (Ed.), *Systems neuroscience* (pp. 119–165). San Diego, CA: Academic Press.
- Backer, G., Mertsching, B., & Bollmann, M. (2001). Data- and model-driven gaze control for an active-vision system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*, 1415–1429.
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press/Clarendon Press.
- Bridgeman, B., van der Heijden, A., & Velichkovsky, B. (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences*, *17*, 247–292.
- Brockmole, J. R., Carlson, L. A., & Irwin, D. E. (2002). Inhibition of attended processing during saccadic eye movements. *Perception and Psychophysics*, *64*, 867–881.
- Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*, 2918–2940.
- Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory guided visual search. *Cerebral Cortex*, *11*, 761–772.
- Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, *10*, 910–923.
- Currie, C. B., McConkie, G. W., Carlson-Radvansky, L. A., & Irwin, D. E. (2000). The role of the saccade target object in the perception of a visually stable world. *Perception and Psychophysics*, *62*, 673–683.
- Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience*, *2*, 740–745.
- 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.
- Deubel, H., Schneider, W. X., & Bridgeman, B. (2002). Transsaccadic memory of position and form. *Progress in Brain Research*, *140*, 165–180.
- Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*, 661–674.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2002). Visual categorization and the primate prefrontal cortex: Neurophysiology and behavior. *Journal of Neurophysiology*, *88*, 929–941.

- Hamker, F. H. (2003). The reentry hypothesis: Linking eye movements to visual perception. *Journal of Vision, 11*, 808–816.
- Hamker, F. H. (2004a). A dynamic model of how feature cues guide spatial attention. *Vision Research, 44*, 501–521.
- Hamker, F. H. (2004b). Predictions of a model of spatial attention using sum- and max-pooling functions. *Neurocomputing, 56C*, 329–343.
- Hamker, F. H. (2005). How the detection of objects in natural scenes constrains attention in time. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention*. Amsterdam: Elsevier.
- Hamker, F. H., & VanRullen, R. (2002). The time course of attentional selection among competing locations [Abstract]. *Journal of Vision, 2*(7), 7a.
- Hamker, F. H., & Worcester, J. (2002). Object detection in natural scenes by feedback. In H. H. Bülthoff, S.-W. Lee, T. A. Poggio, & C. Wallraven (Eds.), *Biologically motivated computer vision* (pp. 398–407). New York: Springer.
- Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception and Psychophysics, 65*, 58–71.
- Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 307–317.
- Irwin, D. E., & Brockmole, J. R. (2000). Mental rotation is suppressed during saccadic eye movements. *Psychonomic Bulletin and Review, 7*, 654–661.
- Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception and Psychophysics, 64*, 882–895.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.
- 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.
- Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*, 1897–1916.
- Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vision, 13*, 201–214.
- Kusunoki, M., & Goldberg, M. E. (2003). The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey. *Journal of Neurophysiology, 89*, 1519–1527.
- Lappe, M., Awater, H., & Kreckelberger, B. (2000). Postsaccadic visual references generate pre-saccadic compression of space. *Nature, 403*, 892–895.
- Lisman, J. E., & Idiart, M. A. (1995). Storage of 7 +/- 2 short-term memories in oscillatory subcycles. *Science, 267*, 1512–1515.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.
- McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 563–581.
- Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience, 16*, 5154–5167.
- Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature, 421*, 370–373.
- Pashler, H. (1988). Familiarity and the detection of change in visual displays. *Perception and Psychophysics, 44*, 369–378.
- Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7*(1/2/3), 17–42.
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53*, 245–277.
- Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research, 42*, 1447–1463.

- Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanism subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*, 1736–1753.
- Rizzolatti, G., Riggio, L., & Sheliga, B. M. (1994). Space and selective attention. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 231–265). Cambridge MA: MIT Press.
- Rolls, E. T., & Tovee, M. J. (1995). The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. *Experimental Brain Research*, *103*, 409–420.
- Ross, J., Morrone, M. C., & Burr, D. C. (1997). Compression of visual space before saccades. *Nature*, *386*, 598–601.
- 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*, *357*, 1073–1082.
- Schlag, J., & Schlag-Rey, M. (2002). Through the eye, slowly: Delays and localization errors in the visual system. *Nature Reviews: Neuroscience*, *3*, 191–215.
- Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *Journal of Neuroscience*, *21*, 1340–1350.
- Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal Comparative Physiology and Psychology*, *43*, 482–489.
- Unemo, M. M., Goldberg, M. E. (1997). Spatial processing in the monkey frontal eye field: I. Predictive visual responses. *Journal of Neurophysiology*, *78*, 1373–1383.
- Von Helmholtz, Hermann (1896). *Handbuch der Physiologischen Optik*. Hamburg-Leipzig, Germany: Leopold Voss.
- Von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenz-Prinzip. *Die Naturwissenschaft*, *37*, 464–476.
- Walker, M. F., Fitzgibbon, E. J., & Goldberg, M. E. (1995). Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. *Journal of Neurophysiology*, *73*, 1988–2003.

## APPENDIX

We now give a formal description of the model. The model can be divided into a static preprocessing part that computes stimulus-driven saliency and a dynamic part that simulates the temporal dynamics of perception.

*Preprocessing and saliency.* Starting from the colour image, we extract orientation maps  $O(\sigma, \theta)$  with varying resolution  $\sigma$  and orientation  $\theta$ , intensity  $I$ , red–green  $RG = R - G$ , and blue–yellow  $BY = B - Y$  maps (Itti, Koch, & Niebur, 1998).

The conspicuity of each feature is implemented by centre-surround operations “ $\ominus$ ”. We construct orientation contrast  $\mathcal{O}(c, s, \theta)$ , intensity contrast  $\mathcal{I}(c)$ , as well as red–green  $\mathcal{R}\mathcal{G}(c)$  and blue–yellow  $\mathcal{B}\mathcal{Y}(c)$  double opponency (Itti et al., 1998). For each feature, we combine the feature information into an attribute  $\mathbf{V}$  and its corresponding contrast value into a gain factor  $P$  of a population code. This dual coding principle is a very important characteristic. A feature is presented by the location of cell activity, and the conspicuity of this feature is represented by the strength of activity. At each location  $x_1, x_2$  we construct a space, whose axes are defined by the represented features and by one additional conspicuity axis. The population is then defined by a set of neurons  $i \in N$  sampling the feature space, with each neuron tuned around its preferred value  $\mathbf{u}_i$ . For each neuron  $y_i$  we obtain an activity value:

$$y_i = P \cdot g(\mathbf{u}_i - \mathbf{V}) \quad (1)$$

Specifically we use a Gaussian tuning curve with the selectivity parameter  $\sigma_g$ :

$$g(u_i - V) = \exp\left(-\frac{\|u_i - V\|^2}{\sigma_g^2}\right) \quad (2)$$

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  $V \in \{I, RG, BY, \theta, \sigma\}$  of each channel between zero and one. The cell activity within the population should typically lie within the range of zero and one. Thus, we also normalize the contrast values to  $\tilde{\mathcal{I}}, \tilde{\mathcal{R}\mathcal{G}}, \tilde{\mathcal{B}\mathcal{Y}}, \tilde{\mathcal{O}}$ . We finally receive the populations for each channel with scale  $c$  at each location  $x$ :

$$\begin{aligned} y_i^I(c, x) &= \tilde{\mathcal{I}}(c, x) \cdot g(u_i - I(c, x)) \\ y_i^{RG}(c, x) &= \tilde{\mathcal{R}\mathcal{G}}(c, x) \cdot g(u_i - RG(c, x)) \\ y_i^{BY}(c, x) &= \tilde{\mathcal{B}\mathcal{Y}}(c, x) \cdot g(u_i - BY(c, x)) \\ y_i^\theta(c, x) &= \max_{\theta} \left( \tilde{\mathcal{O}}(c, \theta, x) \cdot g(u_i - \theta) \right) \\ y_i^\sigma(c, x) &= \max_{\sigma} \left( \tilde{\mathcal{O}}(c, \theta, x) \cdot g(u_i - \sigma) \right) \end{aligned} \quad (3)$$

We now have  $\#c$  maps, where  $\#c$  is the number of center scales, with a population at each point  $x$  for a different centre scale  $c$ . To combine these maps across space into one map with the lowest resolution (highest  $c$ ) we use a maximum operation ( $\max_{c, x \in RF(x)}$ ).

*Temporal dynamics.* Each map in the model represents a functional area of the brain. It contains at each location  $x$  a population of  $i$  cells encoding feature values (Equation 4), with the exception of the maps in the frontal eye field and IOR that only encode space ( $i = 1$ ). In addition V4 and IT have separate maps for different dimensions  $d$  (RG, BY, etc.). The population of cells is driven by its input  $y_{d,i,x}^\dagger$ . Feedback implements an input gain control to enhance the representation of certain features and biases the competition (Desimone & Duncan, 1995) among active populations. Feature-specific reentry operates within the models ‘‘ventral pathway’’ and enhances cell populations whose input matches the feedback signal. Spatial reentry arrives from the frontal eye field and boosts features at a certain location, generally the target of the next saccade.  $I^{f\,inh}$  induces competition among cells and  $I^{inh}$  causes a normalization and saturation. Both terms have a strong short range and weak long range inhibitory effect.

$$\tau \frac{d}{dt} y_{d,i,x} = y_{d,i,x}^\dagger + I^L + I^G - y_{d,i,x} \cdot I_{d,x}^{inh} - I_{d,x}^{f\,inh} \quad (4)$$

The following maps use implementations of the general equation quoted above (Equation 4).

*V4.* Each V4 layer receives input from a different dimension ( $d$ ) in the feature conspicuity maps:  $y_{i,x}^\theta$  for orientation,  $y_{i,x}^I$  for intensity,  $y_{i,x}^{RG}$  for red–green opponency,  $y_{i,x}^{BY}$  for blue–yellow opponency and  $y_{i,x}^\sigma$  for spatial frequency. V4 cells receive feature specific feedback from IT cells ( $I^L = I^L(y^{IT})$ ) and spatial reentry from the frontal eye field ( $I^G = I^G(y^{FEFm})$ ).

*IT.* The populations from different locations in V4 project to IT, but only within the same dimension. We simulate a map containing nine populations with overlapping receptive fields. We do not increase the complexity of features from V4 to IT. Thus, our model IT populations represent the same feature space as our model V4 populations. The receptive field size, however, increases in our model, so that several populations in V4 converge onto one population in IT:  $y_{i,d,x}^\dagger = w^1 \max_{x' \in RF(x)} y_{i,d,x'}^{V4}$ . IT receives feature-specific feedback from the prefrontal memory ( $I^L = I^L(y^{PFmem})$ ) and location-specific feedback from the frontal eye field ( $I^G = I^G(y^{FEFm})$ ).

*PFmem.* The memorization of pattern is achieved through recurrent excitation. Whether a pattern should be memorized depends on the task. Thus, we define a variable  $I^{\text{store}}(t) \in \{0, 1\}$  that controls the inhibition and determines that a pattern is memorized prior to each saccade. The first and strongest pattern closes the gate for other patterns.

$$\begin{aligned}
 I_d^{\text{inh}}(t) &= w_{\text{inh}} \sum_j y_{d,j}(t) \\
 I_d^f{}^{\text{inh}}(t) &= (0.7 - 0.6 \cdot I^{\text{store}}) \sum_j y_{d,j}(t) \\
 y_{d,i}^\dagger(t) &= \sigma(\Gamma_{\text{mem}} - w^{\text{cue}} \max_j y_{d,j}(t)) \max_x \sigma(y_{d,i,x}^{\text{IT}}(t) - C); \\
 \sigma(a) &= \max(a, 0)
 \end{aligned} \tag{5}$$

*PFm.* To determine whether a pattern in the visual scene is similar to the pattern in memory we define PF match cells which compare the current pattern in IT cells with those in PF memory. Activity rises in PF match cells only if both populations match.

$$\begin{aligned}
 I_d^{\text{inh}}(t) &= w_{\text{inh}} \sum_j y_{d,j}(t) \\
 I_d^f{}^{\text{inh}}(t) &= w_{f \text{ inh}} \sum_j y_{d,j}(t) \\
 y_{d,i}^\dagger(t) &:= w^\dagger y_{d,i}^{\text{PFmem}}(t) \max_x y_{d,i,x}^{\text{IT}}(t)
 \end{aligned} \tag{6}$$

*FEFv.* The visuomovement neurons receive convergent afferents from V4 and IT  $y_x^{\text{I}a} = w^{\text{V4}} \sum_d \max_i y_{d,i,x}^{\text{V4}} + w^{\text{IT}} \sum_d \max_{i,x' \in \text{RF}(x)} y_{d,i,x'}^{\text{IT}} + w^{\text{FEFm}} y^{\text{FEFm}}$ . The information from the target template additionally enhances the locations that result in a match between target and encoded feature  $y_x^{\text{I}b} = w^{\text{PFmem}} \prod_d \max_i y_{d,i,x}^{\text{PFmem}} y_{d,i,x}^{\text{V4}}$  at all locations simultaneously. This allows the biasing of specific locations by the joint probability that the searched features are encoded at a certain location.

*FEFm.* The effect of the visuomovement cells on the movement cells is a feedforward excitation and surround inhibition:  $y_x^\dagger = w^{\text{FEFv}} y_x^{\text{FEFv}} - w^{\text{FEFf}} y^{\text{FEFf}} - w_{\text{inh}}^{\text{FEFv}} \sum_x y_x^{\text{FEFv}}$ . Thus, by increasing their activity slowly over time premotor cells compete for the selection of the strongest location.

*FEFf.* After the FEFm cells exceed a threshold  $\Gamma_o^{\text{FEF}}$  and initiate an eye movement at the time  $t_0$ , we assume that the FEFf cells receive input for a brief period  $T^{\text{SAC}}$  and then decay.

$$\begin{aligned}
 \tau \frac{d}{dt} y(t) &= w^{\text{FEF}} I^{\text{FEF}}(t) - y(t) \\
 I^{\text{FEF}}(t) &= \begin{cases} 1 & \text{if } y^{\text{FEFm}}(t_0) > \Gamma_o^{\text{FEF}} \\ 0 & \text{else} \end{cases} \quad t < t_0 + T^{\text{SAC}}
 \end{aligned} \tag{7}$$

*IOR.* There is currently no clear indication where cells that ensure an inhibition of return are located. We regard each location  $x$  as inspected, dependent on the selection of an eye movement at  $y_x^{\text{FEFm}}(t_e) > \Gamma_o^{\text{FEF}}$  or when a match in the PFm cells is lost. In this case the IOR cells are charged at the location of the strongest FEFm cell for a period of time  $T^{\text{IOR}}$ . This causes a suppression of the recently attended location in the FEFv map. IOR cells get slowly discharged by decay with a low weight  $w_{\text{inh}}$ .

$$\begin{aligned}
 \tau \frac{d}{dt} y_x^{\text{IOR}} &= (1 - y_x^{\text{IOR}})(w^{\text{FEFm}} I_x^{\text{FEFm}} - w_{\text{inh}} y_x^{\text{IOR}}) \\
 I_x^{\text{FEFm}} &= \begin{cases} \exp - \frac{(x - x_m)^2}{0.01} & \text{if } t < t_e + T^{\text{IOR}} \\ 0 & \text{else} \end{cases} ; y_{x_m}^{\text{FEFm}} = \max_x (y_x^{\text{FEFm}})
 \end{aligned} \tag{8}$$