BRAIN RESEARCH 1157 (2007) 167–176

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Interactions between higher and lower visual areas improve shape selectivity of higher level neurons—Explaining crowding phenomena

Janneke F.M. Jehee[a,*], *Pieter R. Roelfsema*[b,c], *Gustavo Deco*[d], *Jaap M.J. Murre*[a], *Victor A.F. Lamme*[a,b]

[a]*University of Amsterdam, Department of Psychology, Roetersstraat 15, 1018 WB, Amsterdam, The Netherlands*
[b]*The Netherlands Institute for Neuroscience, Department of Vision and Cognition, Meibergdreef 47, 1105 BA Amsterdam, The Netherlands*
[c]*Department of Experimental Neurophysiology, Center for Neurogenomics and Cognitive Research, Vrije Universiteit, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands*
[d]*Universitat Pompeu Fabra, Department of Technology, Passeig de Circumval·lació 8, 08003 Barcelona, Spain*

ARTICLE INFO

Article history:
Accepted 17 March 2007
Available online 12 April 2007

Keywords:
Visual perception
Computational neural network model
Shape selectivity
Crowding

ABSTRACT

Recent theories of visual perception propose that feedforward cortical processing enables rapid and automatic object categorizations, yet incorporates a limited amount of detail. Subsequent feedback processing highlights high-resolution representations in early visual areas and provides spatial detail. To verify this hypothesis, we separate the contributions of feedforward and feedback signals to the selectivity of cortical neurons in a neural network simulation that is modeled after the hierarchical feedforward–feedback organization of cortical areas. We find that in such a network the responses of high-level neurons can initially distinguish between low-resolution aspects of objects but are 'blind' to differences in detail. After several feedback–feedforward cycles of processing, however, they can also distinguish between objects that differ in detail. Moreover, we find that our model captures recent paradoxical results of crowding phenomena, showing that spatial detail that is lost in visual crowding is nevertheless able to evoke specific adaptation effects. Our results thus provide an existence proof of the feasibility of novel theoretical models and provide a mechanism to explain various psychophysical and physiological results.

1. Introduction

Within 40 ms after the light of an image hits the retina, cells in the primary visual cortex (V1) start to fire. The very first spikes already express orientation and spatial frequency selectivity. The same applies to cells in extrastriate areas that, only 10 ms later, instantaneously code for color, motion, stereo depth, etc. Even the highest levels of selectivity, such as face versus non-face in inferotemporal cortex, appear to be already expressed some 80–120 ms after the image is presented (Oram and Perret, 1992). A great part of object recognition is thus already established during the initial fast feedforward sweep of information transfer.

---

[*] *Corresponding author*. Center for Visual Science, University of Rochester, 242 Meliora Hall, Rochester, NY 14627-0270, USA. Fax: +1 585 271 3043.
E-mail address: jjehee@cvs.rochester.edu (J.F.M. Jehee).

0006-8993/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2007.03.090

From anatomical studies it is clear, however, that cortico-cortical connections not only run from lower areas to higher ones, but also in the reverse direction (Salin and Bullier, 1995). These fibers provide feedback signals that may play a role in object recognition as well. For example, it has been hypothesized that these feedback signals are necessary to process spatially detailed information (Hochstein and Ahissar, 2002). In this framework, feedforward processing involves rapid and automatic processes that provide basic object categorizations, yet incorporating a limited amount of detail. For a detailed and complete representation, higher areas would need to reach back to the low-level areas by means of cortical feedback mechanisms. Low-level areas contain neurons with smaller receptive fields than neurons in higher areas and are in that respect more suitable for the signaling of spatial detail (Hochstein and Ahissar, 2002; Lee et al., 1998; Roelfsema et al., 2000).

Recordings from face selective cells in the inferotemporal cortex are consistent with such an idea: these neurons convey two different modes of information in their firing patterns, starting at different latencies: global information ("is it a face or a non-face?") is transmitted in the initial part of the response and more detailed information ("who's face is it?") is transmitted later, beginning on average about 50 ms after global information (Sugase et al., 1999). Such dynamic changes in the tuning properties of IT neurons seem compatible with feedback interactions with lower visual areas, but whether this is the case has not yet been firmly established. It could equally well be that the two types of information are carried by parallel feedforward streams (such as the magno- and parvocellular inputs) having different latencies, one of which carries low-resolution information, while the other carries high-resolution information.

To verify which of these hypotheses is correct, it would be necessary to separate the contributions of feedforward and feedback signals to the selectivity of cortical neurons, which is notoriously difficult in in-vivo preparations. Some attempts have been made (Hupé et al., 1998; Lamme, 1995; Roelfsema et al., 1998). An example, in the context of scene segmentation, is shown in Fig. 1. Scene segmentation may benefit from interactions between higher and lower visual areas when segregation of a figure from the background is necessary at high spatial resolution (Roelfsema et al., 2000); effects of these interactions would, among others, manifest themselves in lower visual areas. Indeed, presented with the texture-defined square, neurons in V1 respond stronger when their receptive fields lie within the square than when they are stimulated by identical background elements, even when the square is much larger than the classical receptive field size (Lamme, 1995). This contextual modulation typically occurs at some delay with respect to the visual response itself and is abolished when V1 is isolated from feedback from higher areas (Lamme et al., 1998a). Other studies confirm the idea that feedback from extrastriate areas is necessary for V1 cells to signal figure-ground differences (Hupé et al., 1998).

To isolate the potential contribution of feedforward and feedback signals to the receptive field tuning properties of cortical neurons, computational neural network models provide a useful tool. Here, we study these contributions in a simulation that is modeled after the hierarchical feedforward–feedback organization of cortical areas with increasing receptive field size, as is found in the ventral stream of primates. We ask the following questions:

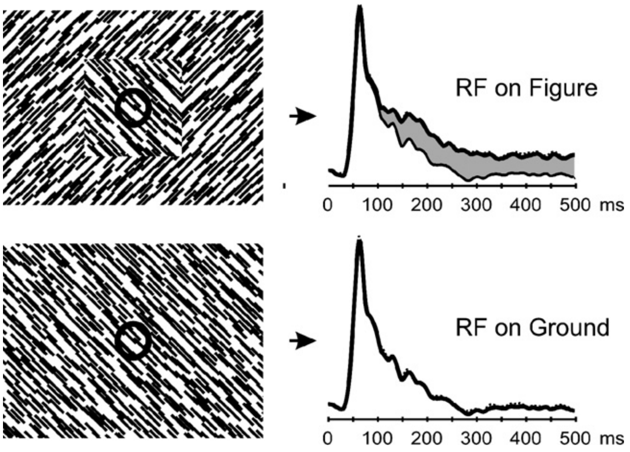

Fig. 1 – Contextual modulation. A V1 neuron responds stronger when its classical receptive field lies within the square (*top*) than when the neuron is stimulated by identical background elements (*bottom*). Black circles denote the classical receptive field. Gray area (*top right*) is the response difference between figure and ground. Modified from Lamme (1995).

1. In the recurrent models, the higher areas 'reach back' to lower areas to find high-resolution information that was not provided by the feedforward sweep. Given the absence of direct connections between low-level areas such as V1 and motor areas (Felleman and Van Essen, 1991), this implies that the only way to report about these details would be via the higher areas. So, these will have to express that high-resolution information in some way by differentially firing to patterns that differ in their details. How do these high-level neurons in the end obtain their tuning to spatial detail? Is this possible for cells with large receptive fields (but see DiCarlo and Maunsell, 2003)? Does this not disturb their original tuning properties?
2. Detailed spatial features of an individual item can be more difficult to identify when other shapes are near it, a phenomenon known as crowding (Bouma, 1970; Toet and Levi, 1992). Global features of the stimulus set as a whole, such as the average orientation of a group of tilted patches, can nonetheless be reported under crowding conditions (Parkes et al., 2001). If top–down feedback interactions process spatial detail, this suggests that they do not come about for individual items when other shapes are in the vicinity. Paradoxically, however, spatial detail that is lost in crowding nevertheless is able to evoke specific adaptation effects (He et al., 1996). Can we explain both the loss of spatial detail in crowding and the paradoxical adaptation to these details in the same framework?

Here, we study these questions in the context of texture segregation, for which top–down feedback processes strongly related to a conscious report have been observed (Lamme et al., 1998b, 2002; Supèr et al., 2001). We use a recurrent neural network model that faithfully simulates these texture-segregation-related processes in V1 in a normal situation, and when the area is isolated from feedback, and compare the firing rate of the model's temporal cortex neurons in these two conditions. We simplify the object recognition aspect of the temporal cortex neurons to the extent that we consider a neuron's response as indicative of object discrimination whenever its responses reflect differences between objects. This is not the strong type of object selectivity that is traditionally observed in temporal cortex neurons, but we argue that the selectivity of our neurons could easily be amplified to obtain strong selectivity if further processing stages were added to the model. We furthermore consider two very basic aspects of shape selectivity, namely selectivity for the overall length of the contour that encloses an object, and the surface area that is covered by an object. These are abstractions of the tuning properties usually found for cells in TE, though it is not unlikely that responses of TE neurons do in fact reflect such object parameters. For example, Sary et al. (1993) studied cue invariant shape tuning of TE neurons using stimuli like squares, crosses, stars etc., which differed along various dimensions. It is not clear what feature dimensions were critical in mediating the strong selectivity of TE cells found in that study. One of these dimensions, however, could have been the contour–surface ratio. Hence, we use three stimuli (a bar, a square and a cross), two of which (the bar versus the other two) differ in the first aspect (contour length) and two of which (the square and the cross) only differ in the last aspect (figure area). We study to what extent selectivity to contour length and figure area depend on feedforward versus re-entrant processing.

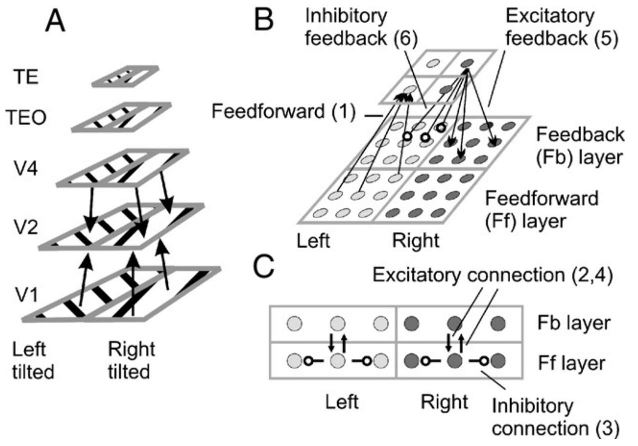

Fig. 2 – Architecture of the connections of the model. (A) The model is composed of five areas. Each area uses neurons with two feature selectivities, 'left-tilted' and 'right-tilted.' At each higher level in the model, the number of units decreases, and the size of the receptive fields increases. (B) Inter-areal connections. Each area is subdivided into feedforward and feedback layers. Units of the feedforward layer excite units with similar feature selectivity in the next higher level (connection type 1). Feedback connections excite units with similar feature selectivity (connection type 5) and inhibit units with the opposite feature selectivity (connection type 6) in the next lower area. (C) Intra-areal connections. In the feedforward pathway, neighboring units with similar feature selectivity inhibit each other (connection type 3). Units in the feedforward pathway and feedback pathway are interconnected via excitatory connections (connection type 2 and 4). Arrow: excitatory connection; open circle: inhibitory connection.

2. Results

2.1. *A re-entrant model for texture segregation*

We use as starting point the model of Roelfsema et al. (2002) and its main principles for texture segregation. The model was first motivated to resolve the 'grouping–segmentation paradox' (Roelfsema et al., 2002), which refers to conflicting constraints posed on neural architecture by grouping (similar image elements should support each other to allow grouping of coherent image regions) and segmentation (similar edge elements should inhibit one another to allow boundary detection and pop out). The model uses a hierarchical architecture to combine the apparently conflicting connection schemes. Within an area, neurons with similar feature preference inhibit each other to ensure detection of boundaries and singletons. These connections are inspired by models of boundary detection and pop out that use inhibitory connections between neurons of similar feature preference coding for nearby locations in the visual field (Grossberg and Mingolla, 1985; Li, 1999; Malik and Perona, 1990; Stemmler et al., 1995). Between areas, neurons with similar feature preference support each other via the model's feedback connections to allow grouping of figural regions (Roelfsema et al., 2002). The model (Fig. 2) is composed of five areas, corresponding to areas V1, V2 and the ventral stream areas V4, TEO and TE. Each area is subdivided into feedforward and feedback layers that interconnect via laminar connections, similar to the laminar division of feedforward and feedback synapses in cortex (Felleman and Van Essen, 1991). Each area uses only two receptive field selectivities, which will be referred to as 'left-tilted' and 'right-tilted', but these could be replaced by others. More features could be added without changing the overall network behavior. The image is represented at a coarser resolution in each successive area of the hierarchy.

2.2. *Feedforward activation detects texture-defined objects, yet lacks spatial detail*

To illustrate the model's general behavior, we will first discuss its response to a texture-defined square when only feedforward and horizontal connections are available to the model (Fig. 3). The image activates pools of neurons that are tuned to the left-tilted and right-tilted orientations. As soon as the first activity evoked by the image reaches V1, interactions between neighboring neurons start. Neurons that have their receptive field on the boundary receive less inhibition from their neighbors and have the strongest response. Boundary detection occurs for neurons with either feature selectivity and can therefore also be seen when the

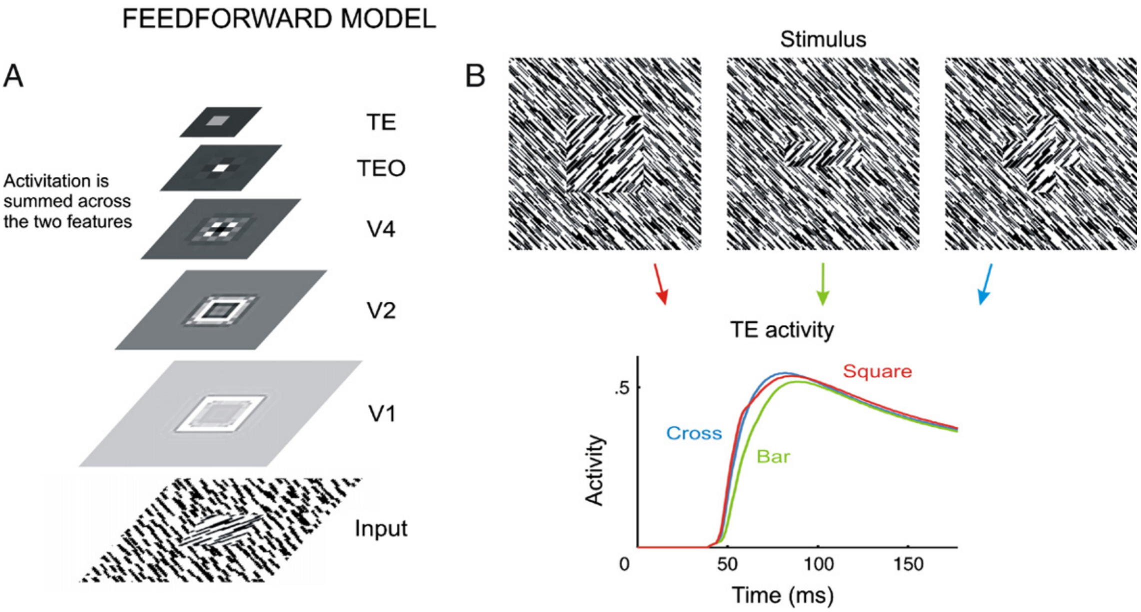

Fig. 3 – Activity in the feedforward model. (A) Activity profile of the five layers 65 ms after presenting the stimulus. To compare between figure and background, activity is summed across the two feature selectivities. Gray levels indicate different response strengths: light shades correspond to regions of high activity, dark regions correspond to regions of low activity. Higher areas represent the image at a coarser resolution. (B, *upper*) The three texture-defined stimuli used in this study: a square, a bar and a cross. (B, *lower*) Temporal activity profile of a TE unit that responds to the three shapes. The neuron discriminates between a bar and the other two objects, but not between a square and a cross.

activity in the two maps is summed. Note that, in lower areas, the summed activity evoked by the interior of the figure is as strong as the summed activity evoked by the background. This pattern of activity resembles that of monkey V1 neurons, which detect boundaries but no figural regions in animals with extrastriate lesions (Lamme et al., 1998a; see also Roelfsema et al., 2002). Meanwhile, in higher areas, more and more neurons become active and start to interact. These neurons represent the same image, though with reducing spatial resolution for each successive layer in the hierarchy. Therefore, if the hierarchy is ascended, the boundaries 'merge' into blobs of elevated activity, until an area is reached where the image evokes activity in a single or very few neuronal pools. Note that blobs of elevated activity appear in the summed activity profile because few higher-level neurons are activated by the image regions, resulting in little inhibition, whereas model neurons tuned to the background are strongly inhibited by activated neighbors of similar feature preference.

The limited spatial resolution of the high-level representation (in TE) is illustrated by the fact that these neurons' responses do not distinguish between all possible shapes. Fig. 3B shows the responses of TE neurons to the texture-defined bar, cross and square. The temporal activity profile of TE neurons that respond to the shape discriminates between a texture-defined bar and the other two stimuli, but not between the texture-defined square and cross. This was a general finding for all TE neurons tuned to both texture orientations. In other words, on the basis of the feedforward and horizontal interactions, the TE neurons can distinguish between shapes with different contour length, but not between shapes that have an equal contour length but differ in figure area (the square and the cross).[1] This makes sense as these network connections function as texture contour detectors.

2.3. *Re-entrant processing adds selectivity for spatial detail to TE neurons*

Let us now consider the model's behavior when feedback connections are added and re-entrant cycles of feedback–feedforward interactions are allowed. The spatially blurred blob of activity in area TE that results from stimulation with a square is fed back to lower areas. This results in a gradual enhancement of responses to the interior of the figure in lower areas ('filling-in'), such that neurons responding to the figure region are now activated more strongly than neurons responding to the background (Fig. 4, *left*), mimicking the neurophysiological results obtained in monkey V1 with texture-defined squares (Lamme, 1995; Lamme et al., 1999). In neurophysiology, responses to texture boundaries are enhanced first and responses to the figural area are enhanced later in time, and model V1 responses follow the precise time course of these responses in primary visual cortex (see Roelfsema et al., 2002 for a detailed comparison of V1 responses). Even though the feedback comes from a spatially

---

[1] In fact, all three stimuli differ in figure area. However, the difference in area between the cross and the square is much larger than the difference between the cross and the bar. Thus, if area were (also) driving the TE response, the cross and square would have evoked different responses in TE. As this is not the case, we feel it is safe to conclude that in the feedforward model contour length alone is driving the responses of the TE neurons.

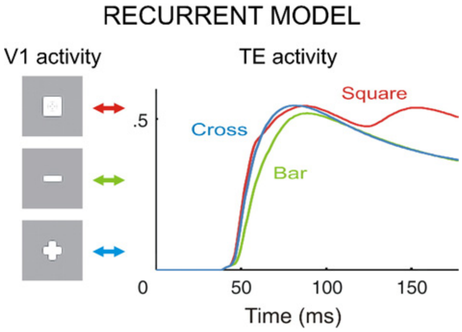

Fig. 4 – Activity in the recurrent model. (*Left*) Summed V1 response to each of the three texture-defined shapes of Fig. 3 after 165 ms. The activity profile is presented in the same format as in Fig. 3. The responses of neurons that cover the interior of the shapes are enhanced relative to the background, which can be viewed as a marker of top–down feedback processes. (*Right*) Temporal activity profile of TE neurons to the three shapes. The neurons initially only discriminate between the texture-defined bar and the other two shapes, but at longer latencies they also discriminate between the texture-defined cross and square.

diffuse higher-level representation, the filling-in activity perfectly respects the boundaries of all three objects.

The filling-in signal at lower levels provides an additional feedforward input, so that now the surface area is also influencing high-level responses. This is reflected in the temporal activity pattern of the neurons in TE (Fig. 4, *right*). Initially, the neurons only discriminate between the texture-defined bar and the other two shapes, but at longer latencies they also discriminate between the texture-defined cross and square, which have equal boundary length, but differ in figure area. Although we have used different stimuli, the difference between the peak of activity elicited by boundary length of the square and the peak elicited by its area is within the range of differences between peaks of global and fine information as reported by Sugase et al. (1999) for face-selective neurons in inferotemporal cortex (66 ms in the model vs. 51±39 ms in inferotemporal cortex).

2.4. *The re-entrant model explains crowding phenomena*

We use the same parameters as above and test our re-entrant model on crowding. Our model aims to capture three findings related to crowding. The first is that crowding effects are negligible (or stimuli are easily identified) when the spatial distance between stimuli is large, they become more pronounced when this distance is reduced, until at some distance, identification of individual stimuli is hindered severely (Parkes et al., 2001). The same is observed in our model with respect to the figures that are only identified after recurrent processing, i.e., the squares and crosses. Squares and crosses that are widely separated evoke clearly different responses in our TE neurons (Fig. 5A). When these stimuli come in close proximity, however, the TE neurons lose their selectivity for individual objects: their firing rates are identical irrespective of the spatial locations of individual crosses or squares (Fig. 5B). Because stimuli are presented at different spatial locations with respect to the TE receptive field center in this case (cf. Fig. 4), we also presented the model the same stimuli in isolation at these locations. Presented in isolation, squares and crosses evoke differential responses in the TE neurons (data not shown), ruling out the possible confound of spatial location.

A second observation that has been made is that features of objects that are lost in crowding may nevertheless induce

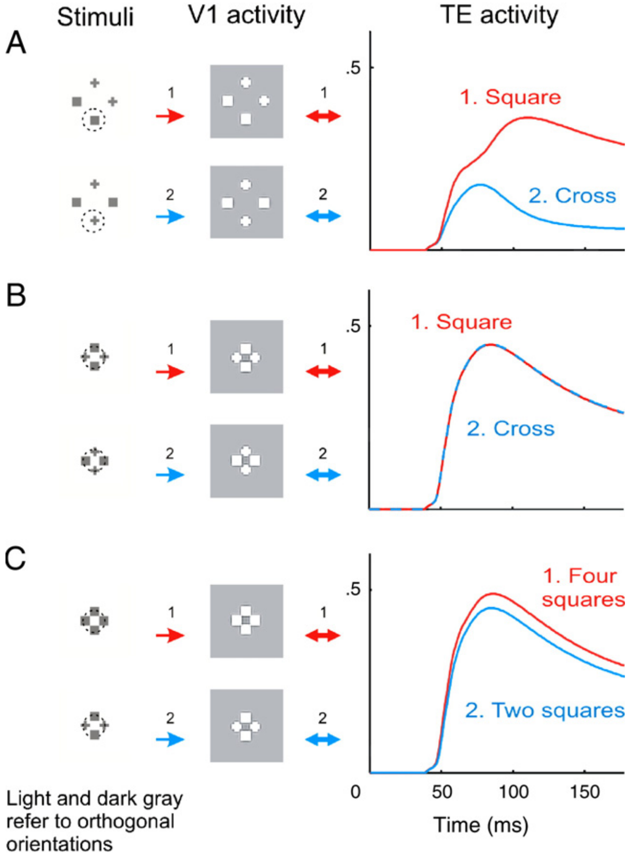

Fig. 5 – Crowding in the recurrent model. (A) Stimuli are widely separated. (*Left*) Schematic display of stimuli. The circle represents the receptive field region where activity of the recorded TE unit exceeds half its maximum rate. (*Middle*) The summed V1 response to the stimuli after 165 ms. The activity profile is presented in the same format as in Fig. 3. (*Right*) The temporal activity profile of TE neurons to the crosses and squares. Red: activity to the four stimuli; blue: activity to the same four stimuli, but a cross and a square have swapped locations. Widely separated stimuli evoke clearly different responses in our TE neurons. (B) Stimuli in close proximity. The firing rate of TE neurons is identical irrespective of the spatial locations of individual crosses or squares, but filling in of the individual shapes at the level of V1 is still spatially selective. (C) Different compositions of the stimulus arrays. Blue is the TE activity to two squares and two crosses, red is the activity to four squares. The TE neurons respond differently to different compositions of the object arrays, even though individual selectivity has been lost. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

selective adaptation effects, presumably in lower-level areas (He et al., 1996). Consistent with this empirical observation, the filling-in of the individual shapes at the level of model V1 is still spatially selective and remains confined to the exact stimulus shape, even in a condition of crowding at higher levels (Fig. 5B).

A third observation is that in conditions of crowding the *average* stimulus content is still accessible for a subject's report (Parkes et al., 2001). In our case, that would imply that, at the level of TE, it would matter whether the array consists of only squares or only of crosses or of a mixture of both. We indeed observe that the model TE neurons respond differently to different compositions of the object arrays, even though the individual selectivity is lost (Fig. 5C). In summary, our model explains three important aspects of crowding phenomena and solves apparent contradictions within these phenomena.

3. Discussion

3.1. *Dynamical tuning properties of TE neurons*

We have shown that single neurons in model TE reveal two different modes of information in their firing patterns, starting at different latencies, one of which results from feedback interactions with lower-level areas. With the initial feedforward sweep, the TE neurons can detect a segregating texture-defined stimulus. However, only after additional feedback–feedforward passes, the neurons also distinguish between stimuli of different shape. More precisely, the feature that is detected during the feedforward sweep of our model is the texture contour length. What is added by the feedforward–feedback interactions is further selectivity for the texture surface area enclosed by the contour. Hence the bar and square/cross are distinguished by the feedforward sweep, while the cross and square (which have equal circumference lengths, but different surface areas) are only distinguished after feedforward–feedback interactions. These interactions allow the lowest levels of the model to represent spatial detail that was not detected in a feedforward manner and to subsequently transmit this information to higher levels in the model. Our results thus provide an existence proof of the feasibility of recent theoretical models (Hochstein and Ahissar, 2002; Lee et al., 1998; Roelfsema et al., 2000) by showing that further selectivity can emerge by means of feedforward–feedback interactions.

We hypothesize that similar mechanisms are at play between lower and higher visual cortical areas during texture segregation, improving selectivity of cortical higher-level neurons to texture-defined shapes. In our model, TE neurons are not as selective for texture-defined shapes as has been found by Sary et al. (1993). They found that TE neurons responded in an almost all-or-nothing fashion to shapes like stars, crosses, circles, etc. We would probably need to introduce further selectivity of the feedforward and feedback connections to achieve a similar result. Although it is not known whether this selectivity to texture-defined shapes arises early or late in the responses of TE neurons (Sary et al., 1993; R. Vogels, personal communication), we believe that the feedforward–feedback interactions described here play a principal role in shaping the selectivity of high-level neurons. In face-selective neurons, for example, dynamic changes in tuning properties have been observed that resemble those reported here. Selectivity evolved from 'face versus non-face' towards selectivity for the faces of different individuals (Sugase et al., 1999), i.e., neurons conveyed coarse information in the initial part of their response and more detailed information later. We conjecture that these changes may also result from top–down interactions with early visual areas for the processing of spatial detail. Our results provide formal proof that such a mechanism is a theoretical possibility.

We used only relatively simple stimuli in our simulations and did not include units tuned to more complex stimulus features in the model. While complex stimulus representations (e.g., faces in inferotemporal cortex) and more basic representations in primary visual cortex are well studied phenomena, cortical feature representations of intermediate complexity and the input they provide to higher-level neurons are less well understood. Thus, to robustly handle more complex tuning, the model would have to be adjusted by adding connections and units that are relatively unconstrained because little is currently known about their selectivity, and little is known about the input they provide to higher-level neurons. We deliberately did not include these connections that would add many free parameters and that would distract from the simple idea proposed here: that representations of object contour and object area appear in lower-level visual areas due to neuronal interactions and that these representations are conveyed to higher-level areas using feedforward–feedback interactions. The model therefore uses simple and transparent mechanisms to compute contours and figural regions and uses the same set of parameters repeated across multiple levels to keep their number at a minimum. Despite a low number of parameters, the model is able to capture data from widely different sources: model responses reflect object contour and subsequently object area as in primary visual cortex (Lamme, 1995; Lamme et al., 1999, see also Roelfsema et al., 2002) and capture dynamic changes in tuning properties as in inferotemporal cortex (Sugase et al., 1999), while still able to capture the main psychophysical observations on perceptual crowding (He et al., 1996; Parkes et al., 2001).

It is not our intention to argue that spatial detail is never detected in a feedforward manner; indeed, purely feedforward computational models of visual processing (e.g., Riesenhuber and Poggio, 1999; Wallis and Rolls, 1997) are able to distinguish between learned stimuli that differ in small aspects, and results from neurophysiology and psychophysics similarly suggest that some stimulus aspects can be detected by higher-level areas in a feedforward fashion (e.g., Oram and Perret, 1992; Thorpe et al., 1996). We rather suggest recurrent interactions to be beneficial whenever detail is not being detected by higher level areas during the initial feedforward sweep. This could happen, for instance, when novel stimuli are presented to which higher-level receptive fields are not yet tuned, or when stimuli are presented in a visually cluttered environment and need to be segregated from the background.

In the recurrent model, the TE neuron differentiates between the bar and the square/cross initially and then between the square and the cross/bar, so that the temporal activity profile as a whole differentiates between the objects. One way in which the system could discriminate between the three objects at any given moment in time might be integrating the responses over time (Fig. S1). Neural integrators have been found in parietal and prefrontal cortex in a number of studies (e.g., Britten et al., 1992; Roitman and Shadlen, 2002). An alternative might be that information is conveyed in the temporal pattern of activity (e.g., Singer, 1999; VanRullen et al., 2005).

Adaptive Resonance Theory (Carpenter and Grossberg, 1987) also uses feedback to enhance low-level activity consistent with the global percept. However, ART, the interactive activation model (McClelland and Rumelhart, 1981) and related models (e.g., Fukushima, 1988) propose a matching process between higher and lower areas that verifies if expected information is really there. In contrast, we suggest that feedback interactions between higher and lower visual areas signal spatial detail that was not provided in the fast feedforward sweep. Tsotsos et al. (1995) propose a model using mechanisms similar to the ones presented here, i.e., initial detection of the globally most salient stimulus representation via feedforward connections and subsequent localization of these representations in lower levels using feedback. This model proposes mechanisms for attentional selection, whereas our model focuses on cortical mechanisms to transmit information that was not conveyed to higher levels initially.

3.2. *Crowding*

If feedback interactions process spatial detail, what happens under crowding conditions where spatial detail is lost? We sought to answer this question by testing our recurrent model on crowding. The model captures the main psychophysical observations on crowding. In actual experiments, observers are unable to report about individual items under crowding conditions (Bouma, 1970; Parkes et al., 2001; Toet and Levi, 1992). Similarly, detailed information about a texture-defined shape is unavailable in the simulated temporal response profile of higher-level neurons when other shapes are presented near it. In addition, our results parallel findings that demonstrate that even though individual stimuli in a crowded area are blocked from a report, global features of the segregated stimulus set as a whole do get through to higher areas (Parkes et al., 2001). Furthermore, the V1 signal is left intact, which is consistent with a demonstration of orientation-specific adaptation to a crowded stimulus (He et al., 1996).

Surprisingly, the model V1 surface representations are enhanced in crowding, implying that feedback interactions are not hampered under crowding conditions. Instead, the representations all fall within the top–down feedback ‘window’ of a single unit in higher areas and can therefore not be selected (enhanced) individually. Thus, although the spatial detail of the individual objects is available at the level of V1, this enhanced activity cannot be conveyed for each object individually to the TE neurons providing the feedback, leaving the latter blind to this detail (while the signal would come through for isolated objects, see e.g., Fig. 4). Rather, the V1 representations within the feedback window are grouped, making the global properties of the segregated stimulus set as a whole available in the response of the higher-level neurons. This explanation is reminiscent of the one proposed by He et al. (1996), who suggested that the limited spatial resolution of attention would be responsible for the crowding phenomenon, i.e., the attentional window would not be small enough to select just one target. Our simulations support this explanation, when attention is viewed as relying on feedback mechanisms.

The stimuli used in the He et al. (1996) and Parkes et al. (2001) experiments differ from the ones used here. In the He et al. (1996) and Parkes et al. (2001) experiments, the orientation of the crowded stimulus differed slightly in orientation from surrounding stimuli, while our model is selective to two orientations of orthogonal orientation. To fully account for their results, the model would have to be modified by including neurons tuned to more orientations. We prefer to use a simpler model with a smaller number of parameters. A small numbers of parameter values are relatively easy to constrain using experimental data, while the number of free parameters associated with more complex models might arise the suspicion of tinkering until higher-level responses showed the wished-for dynamics and the model fit the crowding data. In the model, perceptual crowding results when stimulus representations all fall within the top–down feedback window of a single higher-level unit, resulting in an inability to select representations individually. We hypothesize that similar mechanisms are at play in cortical visual neurons when presented with a crowded display of shapes. Cortical higher-level neurons have broader orientation tuning, and individual V1 representations of slightly different orientation could all fall within the feedback window of a higher-level neuron. Indeed, results from neurophysiology show that the structure of feedback connections matches the structure of the higher-level receptive field (Murphy et al., 1999; Wang et al., 2006). Our model shows that such feedback mechanisms are a possible explanation for crowding phenomena.

In conclusion, recurrent interactions between higher and lower visual areas allow high-level neurons in the present model to express different modes of information in their firing patterns. Given the extensive neural architecture for recurrency (Salin and Bullier, 1995), it is likely that recurrent interactions serve many computational functions, of which processing of spatial detail is but an example.

4. Experimental procedure

4.1. *Model architecture*

The model (Fig. 2) is composed of five areas, each of which is subdivided into feedforward and feedback layers. The first area (model V1) contains 64 × 64 units tuned to a left-tilted orientation and the same number of units tuned to a right-tilted orientation. Units are selective to orientation, but the neuronal mechanisms leading to orientation selectivity are not modeled explicitly (see Olshausen and Field, 1996; Rao and Ballard, 1999; Somers et al., 1995 for models of orientation selectivity). At each higher level in the model, the number of units decreases (by a factor four, Burt and Adelson, 1983), and the size of the receptive fields increases. The image is thus represented at a coarser resolution in each successive area of the hierarchy. The receptive fields of neighboring neurons overlap by 30% in all areas higher than V1.

4.2. Connections and weights

Each unit in layer L of the feedforward pathway receives input from 9 units in the $L-1$ layer that have the same feature selectivity (connection type 1 in Fig. 2 with weight w_1). The strength of the feedforward connections depends on the distance between the receptive field centers $d(i, j)$ according to an approximate Gaussian distribution and is normalized with the sum of the 9 values ($\sigma_U = 0.85$ interunit distance; chosen such that the total output weight is equal for each unit in layer $L-1$). In model area V1, however, each neuron is stimulated by a single input unit of similar feature preference A with activity a_i, $\mathrm{InpFF}^A_{V1,j} = a_i$. Input units use two feature preferences, left-tilted and right-tilted. The activity of the left-tilted input units is obtained by setting all input units overlaying left-tilted image elements to 1 and all other input units to 0, and similar for the right-tilted input units. A further source of input comes from the corresponding unit in the feedback layer (connection type 2 with weight w_2), which modulates feedforward activity (Fukushima, 1988; Grossberg, 1999; Roelfsema et al., 2002). Each unit receives inhibition from 24 neighboring units with similar feature selectivity and via 1 inhibitory self-connection (connection type 3 with weight w_3) (Grossberg and Mingolla, 1985; Li, 1999; Malik and Perona, 1990; Stemmler et al., 1995). The strength of the inhibitory connections depends on the distance between the receptive field centers $d(i, k)$ according to a Gaussian distribution and is normalized with the sum of the 25 values ($\sigma_V = 0.9$ interunit distance).

Each unit in the feedback pathway has a corresponding unit in the feedforward pathway from which it receives excitatory input (connection type 4 with weight w_4). Furthermore, it receives excitatory feedback input from units with similar feature selectivity in the next higher area (connection type 5 with weight w_5) and inhibitory feedback input from units with the opposite feature selectivity in the next higher area (connection type 6 with weight w_6) (Chey et al., 1997; Finkel and Edelman, 1989; Hahnloser et al., 1999).

Connection weights w_i for connection types i were as follows: $w_1 = 2$, $w_2 = 3$, $w_3 = 3$, $w_4 = 1$, $w_5 = 3$, $w_6 = 8$. We performed a systematic exploration of the model's weight parameters that enabled us to draw a bifurcation diagram (Fig. S2; Szabo et al., 2004). The connection weights above come from a regime of operation of the network that is characterized by a correct behavior of model V1 cells (i.e., as real V1 neurons: early enhanced response to boundaries and subsequent enhancement of the figural region, Lamme, 1995; Lee et al., 1998). Parameter values were not very critical for the results and were kept constant throughout all simulations. Note that dynamic tuning of higher-level responses and crowding phenomena always follow when contour and figural areas are detected in model V1 (Fig. S2).

4.3. Activity and updating

The activity of the network units is described by continuous variables that would represent the mean activity of a group or pool of functionally similar neurons in physiology. Activity in the feedforward pathway is updated according to the following equations:

$$\tau_1 \frac{d}{dt}\mathrm{FF}^A_{L,i} = -\mathrm{FF}^A_{L,i} + f^{0.1}_{15}\Bigg(\sum_{j\in U}\left\{\frac{g(d(i,j),\sigma_U)}{\sum_{j\in U} g(d(i,j),\sigma_U)}\mathrm{FF}^A_{L-1,j}\right\} \times \left(w_1 + w_2\mathrm{FB}^A_{L,i}\right) - w_3\sum_{k\in V}\left\{\frac{g(d(i,k),\sigma_V)}{\sum_{k\in V} g(d(i,k),\sigma_V)}\mathrm{FF}^A_{L,k}\right\}\Bigg) - 0.75\mathrm{AF}^A_{L,i} \tag{1}$$

$$\tau_2 \frac{d}{dt}\mathrm{AF}^A_{L,i} = -\mathrm{AF}^A_{L,i} + \mathrm{FF}^A_{L,i} \tag{2}$$

Where $\mathrm{FF}^A_{L,i}$ stands for the activity of pool i with feature selectivity A in the feedforward layer of area L, U is a neighborhood of 9 pools in the $L-1$ layer that have the same feature selectivity, V is a neighborhood of 25 pools in layer L that have the same feature selectivity, $\mathrm{FB}^A_{L,i}$ stands for the activity of pool i with the same feature selectivity in the feedback layer of area L, $\mathrm{AF}^A_{L,i}$ stands for a local inhibitory process of pool i with feature selectivity A in the feedforward layer of area L, and f is a squashing function:

$$f^\theta_s(x) = 0.5(1 + \tanh)(s(x-\theta)) \tag{3}$$

The slope of the initial response transient is determined by τ_1, a time constant that was set to 10. After this transient response, the activity of the neurons is reduced by a local inhibitory process, $\mathrm{AF}^A_{L,i}$, with time constant $\tau_2 = 100$. This process was included to model the transient responses of visual cortical neurons (e.g., Maunsell and Gibson, 1992). To model the transient input coming from neurons in the LGN and retina, the input I to V1 neurons is reduced by such a local inhibitory process AI^A_i as well:

$$\tau_3 \frac{d}{dt}I^A_i = -I^A_i - 0.75\mathrm{AI}^A_i \tag{4}$$

$$\tau_2 \frac{d}{dt}\mathrm{AI}^A_i = -\mathrm{AI}^A_i + I^A_i \tag{5}$$

where $\tau_3 = 200$, chosen such that the model's input layer simulates the decaying responses of LGN and retina, but the model's behavior is not very critical to its value. Activity in the feedback pathway is updated according to the following equations:

$$\tau_4 \frac{d}{dt}\mathrm{FB}^A_{L,i} = -\mathrm{FB}^A_{L,i} + f^{0.65}_{35} \times \Bigg(w_4\mathrm{FF}^A_{L,i} + w_5\sum_{l\in W_i}\left\{\frac{g(d(i,l),\sigma_W)}{\sum_{l\in W_i} g(d(i,l),\sigma_W)}\mathrm{FB}^A_{L+1,l}\right\} - w_6\sum_{m\in Y_i}\Bigg\{\frac{g(d(i,m),\sigma_Y)}{\sum_{m\in Y_i} g(d(i,m),\sigma_Y)} \times \sum_{n\in Z_m}\left[\frac{g(d(m,n),\sigma_Z)}{\sum_{n\in Z_m} g(d(m,n),\sigma_Z)}\mathrm{FB}^{A'}_{L+1,n}\right]\Bigg\}\Bigg) - 0.75\mathrm{AB}^A_{L,i} \tag{6}$$

$$\tau_2 \frac{d}{dt}\mathrm{AB}^A_{L,i} = -\mathrm{AB}^A_{L,i} + \mathrm{FB}^A_{L,i} \tag{7}$$

where W_i is a neighborhood of units in the $L+1$ layer that receive input from unit i in the feedforward layer of area L (hence, feedforward and feedback connections are reciprocal), Y_i is a neighborhood in layer $L+1$ that is as large as W_i but contains units with the opposite feature selectivity A', and Z_m is a neighborhood in layer $L+1$ that contains 25 units with the opposite feature selectivity A'. The time constant of the feedback pathway ($\tau_4=50$) is larger than that of the feedforward pathway, $\sigma_W=\sigma_Y=0.85$, and $\sigma_Z=0.9$ interunit distance. Also, a local inhibitory process $AB^{A}_{L,i}$ was included to model the transient responses of visual cortical neurons.

The model was updated synchronously. The equations were solved using Euler's method. An adequate fit to the experimental V1 data was obtained if each time step of the model was set to 1.25 ms. To avoid boundary effects, all values outside the bounds of the layers were computed by assuming that the layer is periodic. In the figures, 40 ms was added to the data points to account for the delays before area V1 (Nowak et al., 1995) as the retina and the LGN were not included in the model.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.brainres.2007.03.090.

REFERENCES

Bouma, H., 1970. Interaction effects in parafoveal letter recognition. Nature 226, 177–178.

Britten, K.H., Shadlen, M.N., Newsome, W.T., Movshon, J.A., 1992. The analysis of visual motion: a comparison of neuronal and psychophysical performance. J. Neurosci. 12, 4745–4765.

Burt, P.J., Adelson, E.H., 1983. The laplacian pyramid as a compact image code. IEEE Trans. Commun. 31, 532–540.

Carpenter, G.A., Grossberg, S., 1987. ART2, self-organization of stable category recognition codes for analog input patterns. Appl. Opt. 26, 4919–4930.

Chey, J., Grossberg, S., Mingolla, E., 1997. Neural dynamics of motion grouping, from aperture ambiguity to object speed and direction. J. Opt. Soc. Am. A, Opt. Image Sci. Vis. 14, 2570–2594.

DiCarlo, J.J., Maunsell, J.H., 2003. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. J. Neurophysiol. 89, 3264–3278.

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, 1–47.

Finkel, L.H., Edelman, G.M., 1989. Integration of distributed cortical systems by reentry: a computer simulation of interactive functionally segregated visual areas. J. Neurosci. 9, 3188–3208.

Fukushima, K., 1988. A neural network for visual pattern recognition. IEEE Comput. 21, 65–75.

Grossberg, S., 1999. The link between brain learning, attention, and consciousness. Conscious. Cogn. 8, 1–44.

Grossberg, S., Mingolla, E., 1985. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. Psychol. Rev. 92, 173–211.

Hahnloser, R., Douglas, R., Mahowald, M., Hepp, K., 1999. Feedback interactions between neuronal pointers and maps for attentional processing. Nat. Neurosci. 2, 746–752.

He, S., Cavanagh, P., Intrilligator, J., 1996. Attentional resolution and the locus of visual awareness. Nature 383, 224–338.

Hochstein, S., Ahissar, M., 2002. View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36, 791–804.

Hupé, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P., Bullier, J., 1998. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394, 784–787.

Lamme, V.A.F., 1995. The neurophysiology of figure-ground segregation in primary visual cortex. J. Neurosci. 15, 1605–1615.

Lamme, V.A.F., Rodriguez-Rodriguez, V., Spekreijse, H., 1999. Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. Cereb. Cortex 9, 406–413.

Lamme, V.A.F., Supèr, H., Spekreijse, H., 1998a. Feedforward, horizontal, and feedback processing in the visual cortex. Curr. Opin. Neurobiol. 8, 529–535.

Lamme, V.A.F., Zipser, K., Spekreijse, H., 1998b. Figure-ground activity in primary visual cortex is suppressed by anesthesia. Proc. Natl. Acad. Sci. U. S. A. 95, 3263–3268.

Lamme, V.A.F., Zipser, K., Spekreijse, H., 2002. Masking interrupts figure-ground signals in V1. J. Cogn. Neurosci. 14, 1044–1053.

Lee, T.S., Mumford, D., Romero, R., Lamme, V.A.F., 1998. The role of the primary visual cortex in higher level vision. Vis. Res. 38, 2429–2545.

Li, Z., 1999. Contextual influences in V1 as a basis for pop out and asymmetry in visual search. Proc. Natl. Acad. Sci. U. S. A. 96, 10530–10535.

Malik, J., Perona, P., 1990. Preattentive texture discrimination with early vision mechanisms. J. Opt. Soc. Am. A, Opt. Image Sci. Vis. 7, 923–932.

Maunsell, J.H., Gibson, J.R., 1992. Visual response latencies in striate cortex of the macaque monkey. J. Neurophysiol. 68, 1332–1344.

McClelland, J.L., Rumelhart, D.L., 1981. An interactive model of context effects in letter perception. Part 1. An account of basic findings. Psychol. Rev. 88, 375–407.

Murphy, P.C., Duckett, S.G., Sillito, A.M., 1999. Feedback connections to the lateral geniculate nucleus and cortical response properties. Science 286, 1552–1554.

Nowak, L.G., Munk, M.H.J., Girard, P., Bullier, J., 1995. Visual latencies in areas V1 and V2 of the macaque monkey. Vis. Neurosci. 12, 271–384.

Olshausen, B.A., Field, D.J., 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607–609.

Oram, M.W., Perret, D.I., 1992. Time course of neural responses discriminating different views of the face and head. J. Neurophysiol. 68, 70–84.

Parkes, L., Lund, J., Angelucci, A., Solomon, J.A., Morgan, M., 2001. Compulsory averaging of crowded orientation signals in human vision. Nat. Neurosci. 4, 739–744.

Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neurosci. 2, 79–87.

Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nat. Neurosci. 2, 1019–1025.

Roelfsema, P.R., Lamme, V.A.F., Spekreijse, H., 1998. Object-based attention in the primary visual cortex of the macaque monkey. Nature 395, 376–381.

Roelfsema, P.R., Lamme, V.A.F., Spekreijse, H., 2000. The implementation of visual routines. Vis. Res. 40, 1385–1411.

Roelfsema, P.R., Lamme, V.A.F., Spekreijse, H., Bosch, H., 2002. Figure-ground segregation in a recurrent network architecture. J. Cogn. Neurosci. 14, 525–537.

Roitman, J.D., Shadlen, M.N., 2002. Response of neurons in posterior parietal cortex (area LIP) during a combined reaction-time direction-discrimination task. J. Neurosci. 22, 9475–9489.

Salin, P.A., Bullier, J., 1995. Corticocortical connections in the visual system: structure and function. Physiol. Rev. 75, 107–154.

Sary, G., Vogels, R., Orban, G.A., 1993. Cue-invariant shape selectivity of macaque inferior temporal neurons. Science 260, 995–997.

Singer, W., 1999. Time as coding space? Curr. Opin. Neurobiol. 9, 189–194.

Somers, D.C., Nelson, S.B., Sur, M., 1995. An emergent model of orientation selectivity in cat visual cortical simple cells. J. Neurosci. 15, 5448–5465.

Stemmler, M., Usher, M., Niebur, E., 1995. Lateral interactions in primary visual cortex: a model bridging physiology and psychophysics. Science 269, 1877–1880.

Sugase, Y., Yamane, S., Ueno, S., Kawano, K., 1999. Global and fine information coded by single neurons in the temporal visual cortex. Nature 400, 869–873.

Supèr, H., Spekreijse, H., Lamme, V.A.F., 2001. Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nat. Neurosci. 4, 304–310.

Szabo, M., Almeida, R., Deco, G., Stetter, M., 2004. Cooperation and biased competition model can explain attentional filtering in the prefrontal cortex. Eur. J. Neurosci. 19, 1969–1977.

Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381, 520–522.

Toet, A., Levi, D.M., 1992. The two-dimensional shape of spatial interaction zones in the parafovea. Vis. Res. 32, 1349–1357.

Tsotsos, J.K., Culhane, S.M., Wai, W.Y.K., Lai, Y., Davis, N., Nuflo, F., 1995. Modeling visual attention via selective tuning. Artif. Intell. 78, 507–545.

VanRullen, R., Guyonneau, R., Thorpe, S.J., 2005. Spike times make sense. Trends Neurosci. 28 (1).

Wallis, G., Rolls, E.T., 1997. Invariant face and object recognition in the visual system. Prog. Neurobiol. 51, 167–194.

Wang, W., Jones, H.E., Andolina, I.M., Salt, T.E., Sillito, A.M., 2006. Nat. Neurosci. 9, 1330–1336.