UvA-DARE (Digital Academic Repository)

Figure-ground segregation in a recurrent network architecture

Link to publication

Figure–Ground Segregation in a Recurrent Network Architecture

Pieter R. Roelfsema[1], Victor A. F. Lamme[1,2], Henk Spekreijse[1], and Holger Bosch[3]

Abstract

■ Here we propose a model of how the visual brain segregates textured scenes into figures and background. During texture segregation, locations where the properties of texture elements change abruptly are assigned to boundaries, whereas image regions that are relatively homogeneous are grouped together. Boundary detection and grouping of image regions require different connection schemes, which are accommodated in a single network architecture by implementing them in different layers. As a result, all units carry signals related to boundary detection as well as grouping of image regions, in accordance with cortical physiology. Boundaries yield an early enhancement of network responses, but at a later point, an entire figural region is grouped together, because units that respond to it are labeled with enhanced activity. The model predicts which image regions are preferentially perceived as figure or as background and reproduces the spatio-temporal profile of neuronal activity in the visual cortex during texture segregation in intact animals, as well as in animals with cortical lesions. ■

INTRODUCTION

The visual brain segregates scenes into objects and background, as a first processing step on the way to perception. A wide range of cues is exploited to detect boundaries between objects and the visual background. Generally, locations at which the image changes abruptly are assigned to object boundaries, whereas image regions that are relatively homogeneous are grouped together. Texture segregation provides a well-known example. Consider the image of Figure 1A, which consists of line elements with a homogeneous orientation, except in a square region where orientation is orthogonal. This square is perceived as coherent figural region superimposed on a background that may even appear to continue behind it. It is our aim to propose a neural network architecture that combines two elementary processes that underlie texture segregation. First, the network detects boundaries between figure and background. Second, it groups together image elements of the figure, to obtain a coherent representation of the figural region. We will discuss that these two processes impose conflicting constraints on the architecture of network connections, a problem that will be referred to as the "grouping–segmentation paradox."

The first process that is essential for texture segregation is boundary detection. Boundaries are defined by abrupt changes in the properties of the texture elements. Many studies on boundary detection have focused on the detection of singletons, individual texture elements that pop out because they differ from the surrounding elements in one of a number of elementary features (Nothdurft, 1992; Theeuwes, 1992; Moraglia, 1989). An orientation singleton, for example, is a figure that consists of a single line element superimposed on a background with a different orientation. Neurophysiological correlates of singleton detection have been uncovered in various visual areas, for image elements that differ from the background in their orientation (Kastner, Nothdurft, & Pigarev, 1997; Zipser, Lamme, & Schiller, 1996; Lamme, 1995; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995; Knierim & Van Essen, 1992; Allman, Miezin, & McGuinness, 1985), color (Schein & Desimone, 1990; Zeki, 1980), or direction of motion (Zipser et al., 1996; Lamme, 1995; Born & Tootell, 1992; Lagae, Gulyas, Raiguel, & Orban, 1989; Tanaka et al., 1986; Allman et al., 1985). These studies compared neuronal responses to (1) image elements in isolation, (2) image elements surrounded by similar neighbors, and (3) image elements with dissimilar neighbors that pop out. Responses to image elements presented in isolation are generally strongest. These are followed by responses to elements that pop out, which are in turn stronger than responses to elements with similar neighbors.

These results inspired models of boundary detection that assume that neurons with adjacent receptive fields are interconnected with inhibitory connections, that are strongest among neurons with a similar feature

[1]University of Amsterdam, [2]The Netherlands Ophthalmic Research Institute, [3]University of Geneva

Figure 1. Correlates of figure–ground segregation in area V1 of the macaque monkey. (A) Left: Square figure that segregates from the background on the basis of a difference in orientation. The circle indicates a receptive field of a neuron in area V1. Right: Stimulus that only contains background. The texture elements inside the receptive field are identical for the two stimuli. (B) Responses of a group of V1 neurons to the figure (thick line) and background (thin line). Responses to the figure are strongest, although the segments in the receptive field are the same. (C) Subtracting the response to the background from the response to the figure isolates the response enhancement. Note that the figural response enhancement occurs after the peak response (modified from Lamme et al., 1999).
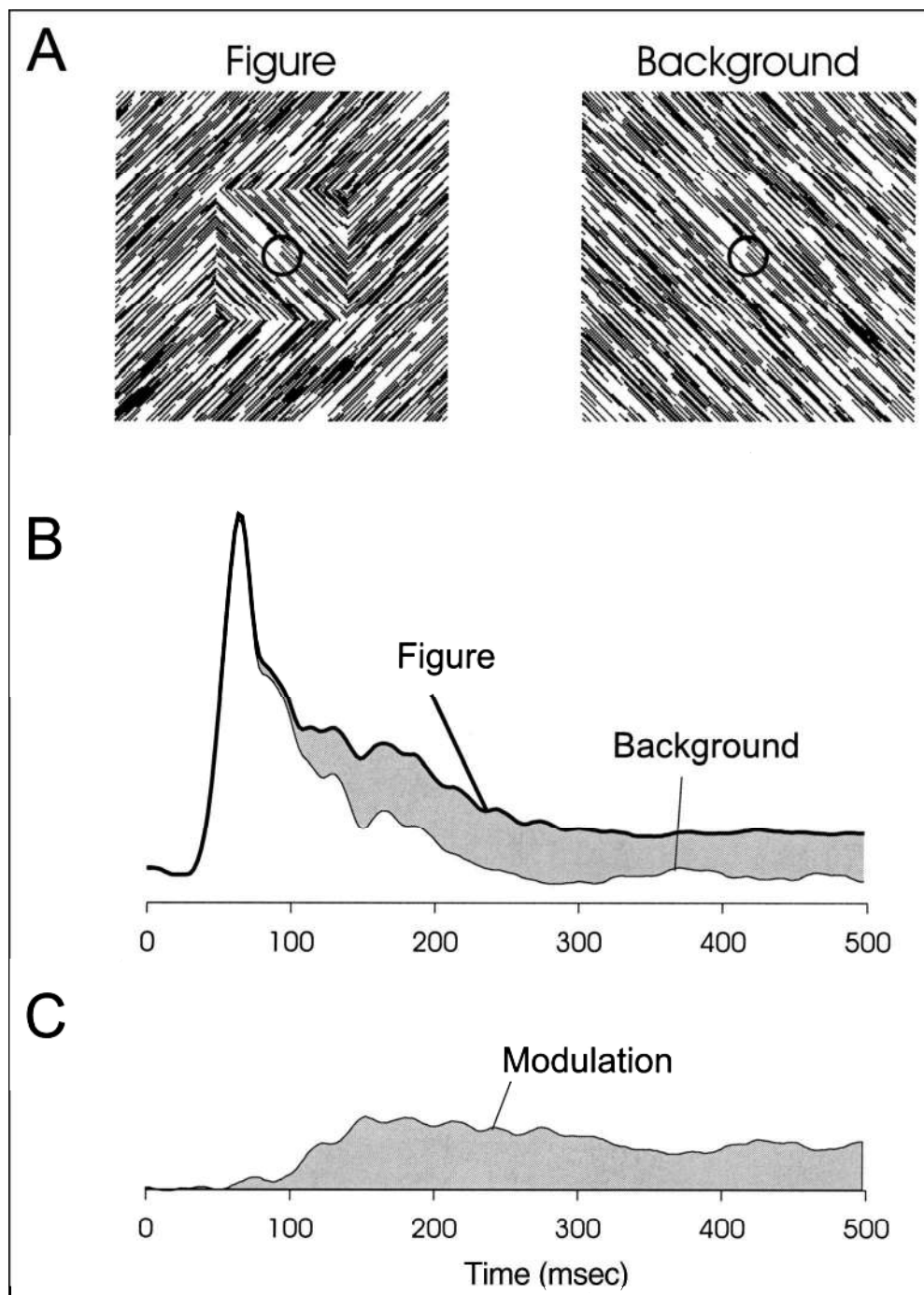
preference (Li, 1999; Stemmler, Usher, & Neibur, 1995; Malik & Perona, 1990; Grossberg & Mingolla, 1985). In such a connection scheme, neuronal responses to elements in the middle of a homogeneous region are inhibited maximally, and inhibition is weaker for responses to boundaries and singletons. Note that this scheme predicts that responses to the middle region of a figure (as in Figure 1A) are also maximally suppressed, because here the neighboring elements have a similar orientation, and provide strong inhibition. This prediction seems to be in conflict with our percept of the figure, which is entirely in the foreground. More importantly, it is also contradicted by cortical physiology. Responses to the middle of a textured square are enhanced, even in the primary visual cortex (Zipser et al., 1996; Lamme, 1995) (Figure 1B,C). Firing rates are enhanced homogeneously, for the entire figural region. Thus, the entire figure is grouped together by a process that demarcates neuronal responses to the figural elements by labeling them with an enhanced firing rate. How can such a labeling operation be implemented in the visual cortex?

Neurons that respond to a single coherent image region typically have a similar feature preference. Thus, the labeling operation could be implemented by spreading the rate enhancement among neurons with a similar feature preference (Roelfsema, Lamme, & Spekreijse, 2000), a process sometimes referred to as region-filling (Tanimoto, 1985) or coloring (Ullman, 1984). Such a process would eventually label the figural region in its entirety, and need not cross its boundaries, because here the feature preference of the activated neurons changes abruptly. Note, however, that this requires an excitatory interaction between neurons tuned to similar features.

We refer to these apparently conflicting constraints on the architecture of network connections is as the grouping–segmentation paradox. On the one hand, neurons that respond to similar image elements should support each other, to promote each other's coselection. This is important for the grouping of similar image elements into coherent regions. On the other hand, neurons that respond to similar image elements should inhibit each other, to allow boundary detection and pop out. It is our aim to propose a hierarchical neural network model that combines both connection schemes to resolve the paradox.

The model is composed of several areas, resembling the hierarchical organization of the visual cortex (Figure 2). Each model area consists of two layers, a feedforward and a feedback layer. Neurons in the feedforward layer inhibit each other if they have a similar feature preference. Thereby, boundaries and singletons are detected in an initial feedforward sweep of activity though the network. Neurons in the feedback layers rather excite each other if they have a similar feature preference. Feedback groups entire figural regions together by labeling them with the rate enhancement. The model

accounts for the timing of boundary detection and region filling that is observed in cortical physiology. Moreover, the model also explains the effects of lesions in higher visual cortical areas on texture segregation.

RESULTS

A Recurrent Model for Texture Segregation

The model does not attempt to simulate the details of receptive field properties in the various visual cortical areas. Such details would not only make the model unnecessarily complicated, but might even distract from the essence of the connection scheme. The model should rather be regarded as a computational recipe that combines boundary detection and grouping of figural regions in a single network architecture. Therefore, only two features are used, which will be referred to as "left oblique" and "right oblique." These features could be replaced by others, and more features could be added without changing the overall behavior of the network.

The overall layout of the proposed model resembles the organization of the visual cortex. It consists of several hierarchical levels corresponding loosely to areas V1, V2, V4, TEO, and TE (e.g., Olshausen, Anderson, & Van Essen, 1993; Burt & Adelson, 1983). The size of receptive fields increases in higher areas and the number of neurons decreases. Thus, higher areas represent the image at a coarser resolution. Each area is subdivided into two layers, a feedforward and a feedback layer (Figure 2A). Feedforward layers propagate activity to higher areas through feedforward connections (Figure 2, connection type 1) and feedback layers propagate activity in the opposite direction (connection types 2 and 3). Within each of the areas, the two layers can

Figure 2. Architecture of the connections in the model. (A) The model is composed of five areas, which are named after the areas of the temporal stream of the monkey brain. Each area contains units tuned to the left diagonal and right diagonal orientations, which are segregated into a feedforward (FF) and a feedback layer (FB). (B) Inter-areal connections. Units of the feedforward layer excite units with the same feature specificity in higher areas (connection type 1). Feedback connections excite units with a similar feature selectivity (connection type 3), and inhibit units with the opposite feature selectivity (connection type 2). (C) Intra-areal connections. In the feedforward pathway, neighboring units inhibit each other if they have a similar feature preference (connection type 4). This lateral inhibition results in an enhanced response to boundaries and singletons. Units in the feedforward and feedback layers with overlapping receptive fields are reciprocally interconnected with excitatory connections (connection types 5 and 6).
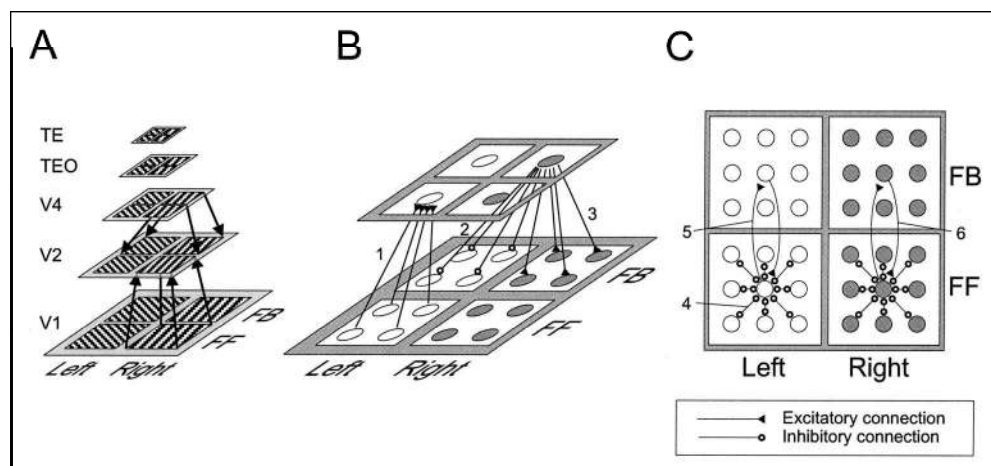
interact through interlaminar connections (connection types 5 and 6).

The Feedforward Pathway

At each level of the feedforward pathway, boundaries and singletons are detected. Neurons inhibit other cells with adjacent receptive fields that have a similar feature preference (Li, 1999; Stemmler et al., 1995; Malik & Perona, 1990; Grossberg & Mingolla, 1985) (connection type 4). To indicate the implications of this connection scheme it will first be shown how the network behaves if it consists of only the feedforward pathway (connection types 1 and 4 in Figure 2). Figure 3A illustrates the response of such a feedforward network to a textured square on a background with an orthogonal orientation. The image activates V1 neurons that respond to the figure with a left oblique orientation, as well as neurons that respond to the orthogonal surround. These two groups of neurons are shown in separate maps in Figure 3A, but would in reality be intermingled in a single retinotopic map. Neurons that have their receptive field on the boundary between figure and background have the strongest responses, because they receive less inhibition from their neighbors. This boundary enhancement occurs for neurons tuned to both orientations and it can therefore also be seen if the activity in the two maps is summed together (right panel in Figure 3A). Note, however, that the summed activity in area V1 evoked by the interior of the figure is as strong as the summed activity evoked by the background.

Figure 3. The activity profile of the feedforward layers, in the absence of feedback. (A) Left: Figure with a left diagonal orientation superimposed on a background with an orthogonal orientation. Middle: Units of model area V1 that are tuned to the left and right oblique respond to the figure and background, respectively. Gray levels indicate different response strengths; light shades correspond to regions of high activity, and dark shades to regions of low activity. Units tuned to either orientation respond most strongly to the edges. Right: The enhancement of responses to the edges is clear if activity is summed across the two features. (B) Response in the feedforward layers of the various model areas, summed across both features. Higher areas represent the image at a lower spatial resolution, and the edges are not represented individually. The response enhancement is strongest in area V4, where the figure fills only a few receptive fields. (C) Response to a larger figure. Now the maximal response enhancement occurs at a higher hierarchical level.
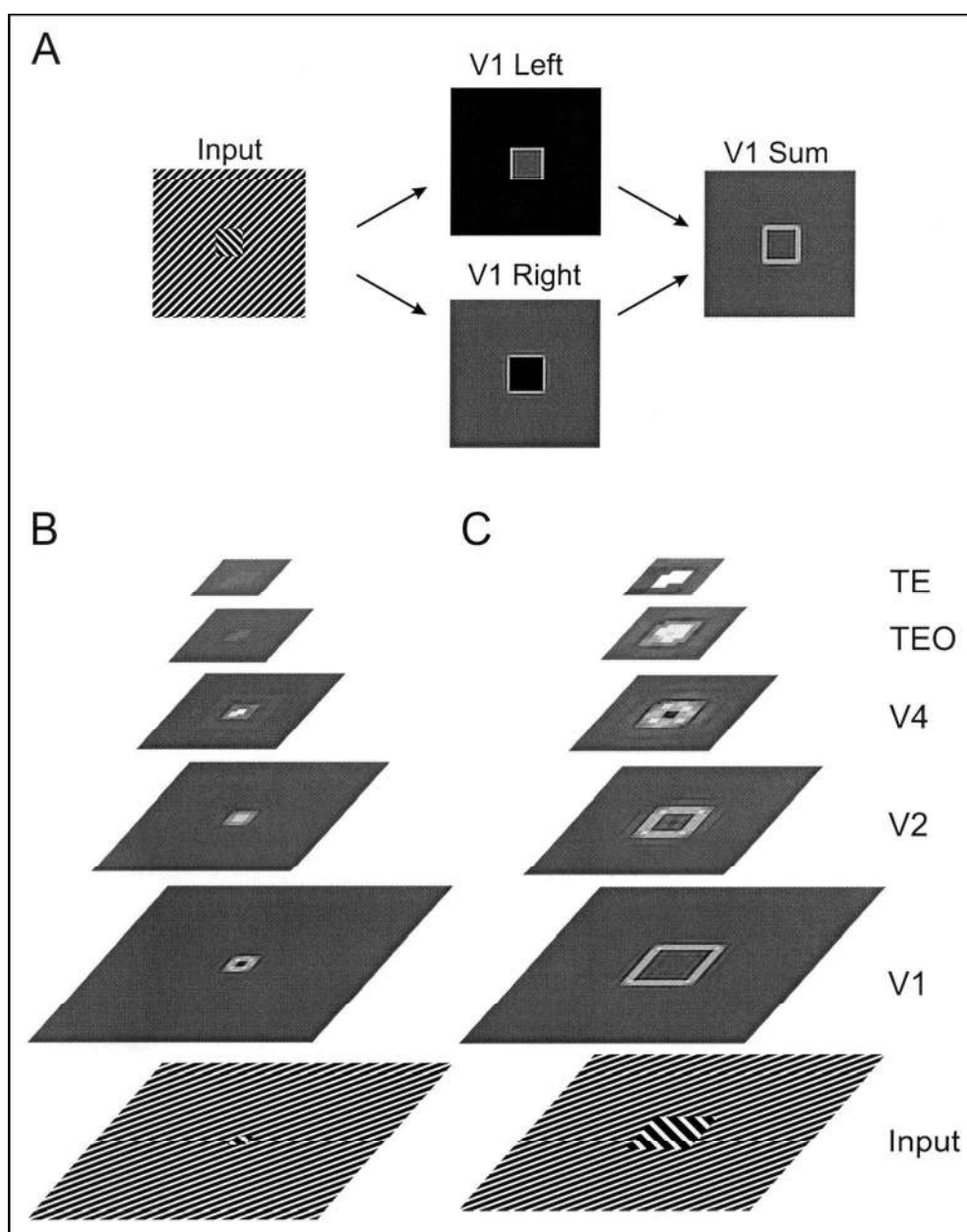
The situation is different for higher areas, which contain a representation of the same image, but at a coarser resolution. If the hierarchy is ascended, the interior of the figure is gradually filled in (Figure 3B,C) until an area is reached where the entire figure fills just one or a few receptive fields. Here, neurons that are activated by the figure respond strongly, because they have very few activated neighbors with the same feature preference and they therefore receive little lateral inhibition. In other words, at this level the figure is represented as if it is a singleton. Thus, responses evoked by the figure are strongest at a level where its size is comparable to the size of the receptive fields and large figures evoke maximal responses at higher hierarchical levels than small figures (Figure 3B,C).

The Feedback Pathway

The feedback pathway ensures that responses to the entire figural region are enhanced relative to responses to the background. Each figure yields a maximal response at one of the hierarchical levels of the feedforward pathway. Activity in the feedforward pathway reaches the feedback pathway through the interlaminar connections (connection type 5 in Figure 2). The enhanced response to the figure is subsequently fed back to neurons at lower hierarchical levels that also respond to the figure. Feedback neurons excite neurons at the next lower layer with a similar feature preference, but inhibit neurons with the opposite feature preference (connection types 2 and 3, respectively; see Hahnloser, Douglas, Mahowald, & Hepp, 1999; Chey, Grossberg, & Mingolla, 1997; Finkel & Edelman, 1989, for similar connection schemes).

Higher areas represent the image at a courser resolution. This implies that the spatial resolution of the feedback signal in any area is coarser than the resolution

of the feedforward signal. Physiological data on the topography of cortico-cortical connections indicate that feedback projections are indeed somewhat spatially diffuse (Salin & Bullier, 1995). Thus, a mechanism is required in the feedback pathway to prevent the enhanced activity evoked by the figure from spilling over to the background. In the model, this is accomplished by a multiplicative interaction between feedforward and feedback (e.g., Grossberg, 1999; Fukushima, 1988). This ensures that only neurons that receive bottom-up support from the feedforward pathway are influenced by feedback. In other words, feedback is gated by feedforward activation (see Methods), as is illustrated in Figure 4. Suppose that a left oblique figure on an orthogonal background is shown and that the maximal response enhancement in the feedforward pathway occurs at hierarchical level N. This enhanced activity enters into the feedback pathway through interlaminar connections. Neurons of the feedback layer at level N that are tuned to the left oblique propagate the enhanced activity to similarly tuned neurons at level $N - 1$. In the feedback layer of level $N - 1$, only those neurons that have their receptive field on the figural region can be influenced by this enhanced feedback, because they receive bottom-up support from neurons in the feedforward layer that have overlapping receptive field (through connections of type 5 in Figure 2). The enhanced feedback cannot spill over to neurons that are tuned to this orientation and that have a receptive field on the background (small cells in Figure 4A), since these cells do not receive bottom-up support. Physiological evidence supports such gating of feedback by feedforward activation. Most neurons can only be influenced by contextual stimuli that are positioned outside their receptive field, if there is another stimulus inside their receptive field that drives the cell (Zipser et al., 1996; Kapadia, Ito, Gilbert, & Westheimer, 1995; Sillito et al.,

Figure 4. Interaction between feedforward and feedback. In the feedback layer of level N, only a single unit is activated by the left diagonal orientation (A). This unit responds more strongly than units at the same level that are tuned to the orthogonal orientation (B), due to the lateral inhibition in the feedforward pathway. The response enhancement is propagated to neurons of the feedback layer of level $N - 1$ that are also tuned to the left oblique (A, thick connections). The feedback projection is somewhat diffuse. However, only neurons at level $N - 1$ that receive bottom up support from the feedforward layer can benefit from the enhanced feedback. Feedback connections to cells that do not receive bottom up support have no effect (stippled connections). The feedback signal for the right diagonal orientation is gated similarly (B), but this feedback signal is weaker (thinner connections).
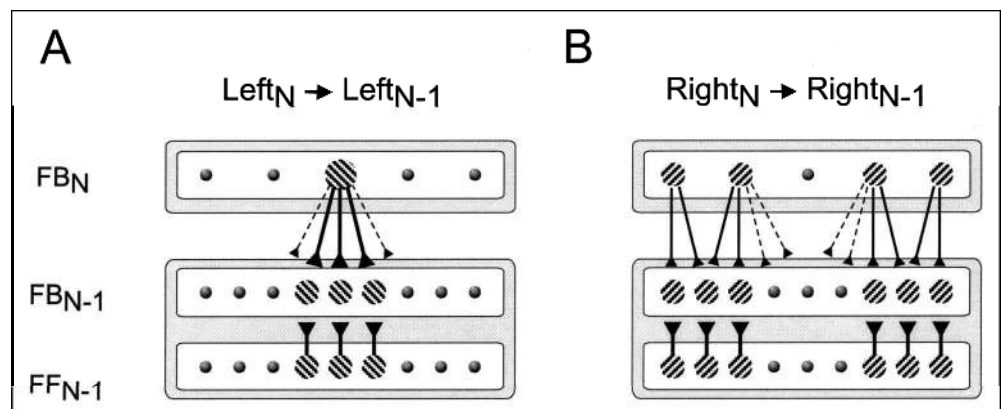
1995; Knierim & Van Essen, 1992; Tanaka et al., 1986; Allman et al., 1985). In principle, such contextual effects can be mediated through lateral connections as well as through feedback connections. Recent studies suggest that at least some of these contextual effects indeed depend on feedback connections (Hupé et al., 1998;

Lamme, Supèr, & Spekreijse, 1998). Moreover, a similar gating by feedforward activation occurs for response modulations that are caused by visual attention. Attention has strongest effects on neurons that have an appropriate stimulus in their receptive field, but hardly influences cells that do not receive bottom up activation

Figure 5. Spatio-temporal pattern of response enhancement in area V1 of the model and the macaque monkey. (A) The input on the left was presented. After 65 msec, the model exhibits a response enhancement that is confined to the boundary between figure and background. After 190 msec, responses to the entire figural region are enhanced due to feedback from higher areas. (B) Thick curves show the total response (left + right orientation) of model V1 units. Units had a receptive field on the boundary between figure and background (left) or in the interior of the figure (right). Thin curves show the response to the background. (C) Thick curves show population responses in macaque area V1 to the figure boundary (left), or to the interior of a figure (right). Thin curves, response to the background. Arrows indicate the latency at which the response enhancement became significant (modified from Lamme et al., 1999).
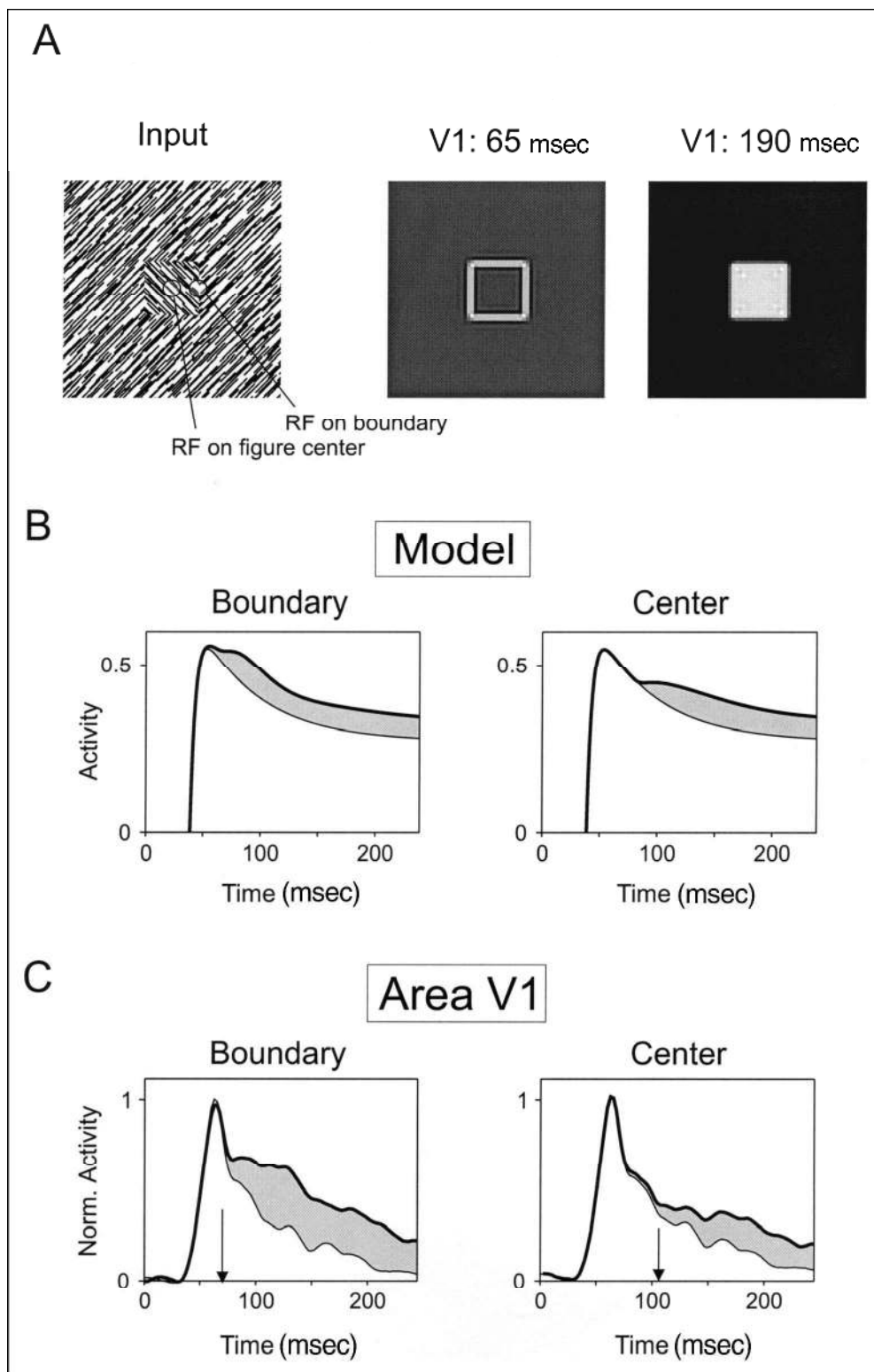
(McAdams & Maunsell, 1999; Treue & Martínez Trujillo, 1999). These attentive effects have to be attributed to feedback connections.

The enhancement of the response to the figure at level $N - 1$ is subsequently fed back to the next lower level and eventually also reaches area V1. In each of the areas, feedback initially influences neurons in the feedback layers, but these effects are also propagated to the feedforward layers, through the interlaminar connections (connection type 6 in Figure 2). The interlaminar connections from the feedback layer to the feedforward layer are indirectly excitatory, because they reduce the impact of lateral inhibition (see Methods), a connection scheme was chosen to avoid strong excitatory loops, which may result in an uncontrolled amplification of activity (Crick & Koch, 1998).

The temporal profile of the response enhancement in the feedforward layer of area V1 of the full network (that includes the feedback layers) is illustrated in Figure 5. Initially, the activity profile across V1 resembles the pattern that is obtained in the absence of feedback. Lateral inhibition within the feedforward pathway enhances responses to the boundaries, but responses to the interior of the figure are not yet enhanced (Figure 5A,B). It takes additional time before the effects of feedback are expressed, because the feedforward pathway first has to propagate the activity to higher areas. Enhanced feedback reaches neurons in the feedforward layers of area V1 that respond to the interior of the figure at a latency of approximately 100 msec and is maintained thereafter. Responses to the entire figure are now homogeneously enhanced. Note that the very same units that would exhibit an enhanced responses to a boundary also enhance their responses to the interior of a figure, albeit at a later point in time.

For comparison, Figure 5C reproduces data of Lamme, Rodriguez-Rodriguez, & Spekreijse (1999), which were obtained in area V1 of monkeys involved in a texture segregation task. Also in macaque area V1, responses to the boundaries between figure and background are enhanced first, at a latency of approximately 70 msec. Thereafter, responses to the interior of the figure are enhanced, at a latency of about 110 msec. Thus, an architecture for texture segregation in which boundary detection is performed by the feedforward sweep and feedback groups image regions together reproduces the main features of physiological data in area V1.

Lesions in Extrastriate Areas

Lesions in various visual areas of cats (De Weerd, Sprague, Vandenbussche, & Orban, 1994) and monkeys (Huxlin, Saunders, Marchionini, Pham, & Merigan, 2000; De Weerd, Desimone, & Ungerleider, 1996; Merigan, 1996; Merigan, Nealey, & Maunsell, 1993) have been shown to interfere with the perception of the shape of

Figure 6. The effect of lesions in higher visual areas on activity in area V1. (A) Magnitude of response enhancement in macaque area V1 to an image containing a square figure that is defined by an orientation difference. Responses at different positions along a line (left panel) are shown next to each other. Middle: Profile of response enhancement (figure response minus background response) in control animals. Right: Profile of response enhancement in an animal with a lesion that included (parts of) areas V3, V3a, V4, V4t, MT, MST, FST, PM, DP, and 7a. The lesion reduced the response enhancement to the interior of the figure, while the response enhancement to boundaries was maintained (modified from Lamme et al., 1998). (B) Response profile of the model to a similar stimulus configuration. Middle: profile of response enhancement in area V1 of the complete model. Right: V1 response profile of a model in which all areas above area V1 were removed.
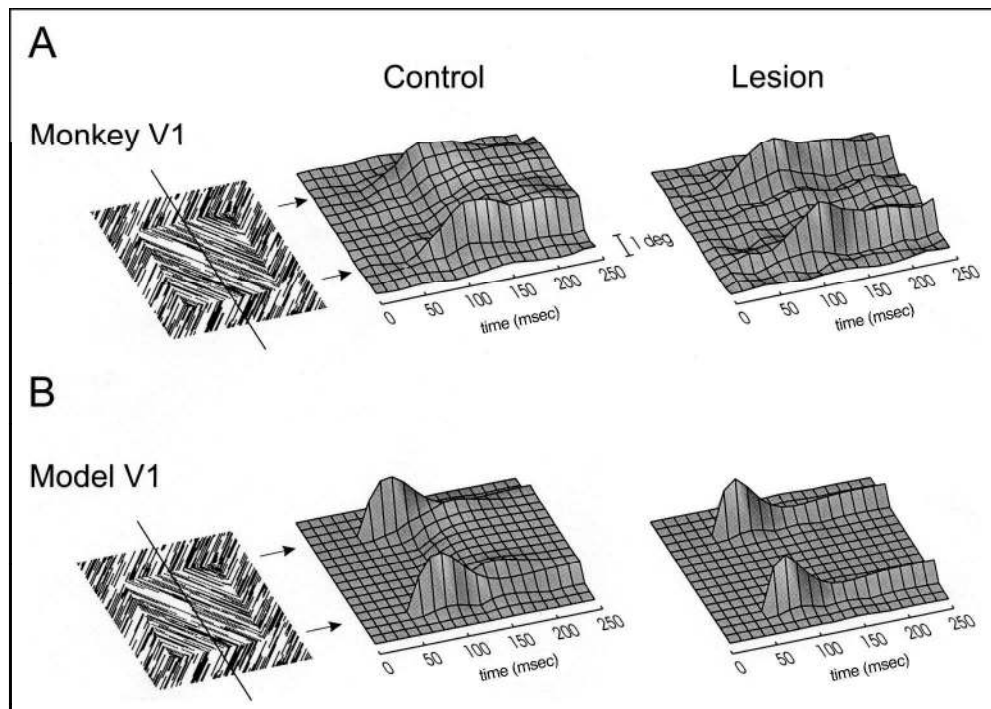
texture defined figures. Remarkably, such lesions need not impair the detection of singletons (Merigan et al., 1993). These psychophysical results are supported by a recent physiological study, which showed that singletons can still be detected by neurons in area V1 when area V2 is silenced pharmacologically (Hupé, James, Girard, & Bullier, 2001). These findings are in accordance with the proposed connection scheme, since a disruption of feedback to area V1 should impair grouping of image regions, but spare the detection of boundaries and singletons.

There is also physiological data that supports the involvement of feedback connections in the grouping of coherent image regions. A recent study (Lamme et al., 1998) investigated the effect of a relatively large extrastriate lesion on neuronal correlates of texture segregation in macaque area V1. Such a lesion does not change feedforward processing in area V1, but removes many of the neurons that provide feedback. Figure 6A compares the spatio-temporal profile of the response enhancement in area V1 of monkeys with an extrastriate lesion to that of control monkeys. The neurons that were tested had their receptive fields at various locations along an imaginary line across the figure, and background responses were subtracted from figure responses in order to isolate the response enhancement (gray area in Figure 5C). Boundary detection occurs both in animals with and without the lesion, at a relatively early latency (Figure 6A). In control animals, the interior of the figure is subsequently filled in with the rate enhancement. Filling of the figure interior does not occur in animals with an extensive extrastriate lesion. The model readily reproduces these results if all visual areas are removed, except area V1. As expected, boundary detection occurs in the complete as well as in the lesioned model, but the response enhancement to the interior of the figure is absent in the model with a lesion (Figure 6B).

V1 Responses to Complex Shapes

We next investigated the network's response to more complex shapes. Candidate objects are extracted in the feedforward pathway through lateral inhibition. The relative magnitude of responses to figure and background therefore depends on the local shape of the object, as is indicated schematically in Figure 7E.

Figure 7. Response to complex objects in model area V1. (A–C) Sustained response profile (at 230 msec) across model area V1 to a U shape (A), a plus shape (B), and a figure with a hole (C). (D) Spatio-temporal response profile in model area V1 along an imaginary line in (A). Note the transient response enhancement to the center of the U shape. (E) The relative amount of lateral inhibition that is generated in the feedforward pathway can be approximated by counting the number of active neighbors that share a unit's feature preference. At straight boundaries, inhibitory input to units that respond to figure and background is balanced (five active neighbors with a similar tuning). Inhibition is weaker at convex corners (three neighbors), and stronger at concave corners (six or seven neighbors).
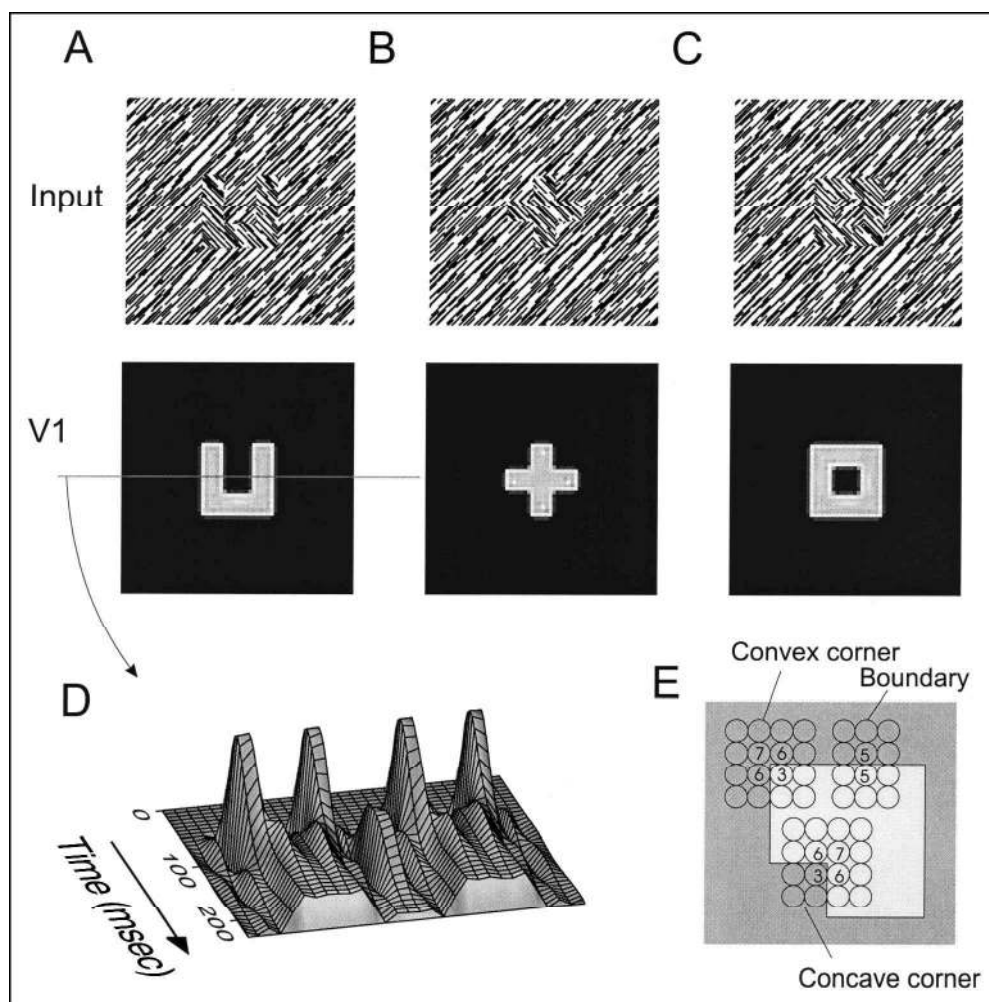
Neurons that have their receptive field on a homogeneous region are maximally inhibited, whereas neurons with a receptive field close to a boundary receive less inhibition, as was discussed above. This holds for neurons that respond to the figure, as well as for neurons that respond to the background. The response enhancement on the two sides of a straight boundary is therefore balanced. At corners, however, the strength of lateral inhibition differs between figure and background. At convex corners, neurons that respond to the figure receive less inhibition than those that respond to the background (Figure 7E). Thus, if a convex figure such as a square is presented, responses to the figure are at least as strong as responses to the background, at each image location and at each hierarchical network level. At concavities, the opposite is true and the neurons that respond to the figure receive the strongest inhibition. This results in a relative enhancement of responses to the background. However, such an inappropriate enhancement of background responses only occurs in lower visual areas. In higher areas, the figure is eventually represented by the activity of a few neurons and local concavities are lost due to the reduced spatial resolution. Therefore, responses to the figure are always strongest at these higher hierarchical levels. In the network, feedback from these higher areas eventually also overcomes the inappropriate enhancement of background responses in lower areas, as is illustrated in Figure 7A–C. It is essential that feedback is gated by feedforward activation during the propagation of feedback to lower visual areas, because details of the figure can be filled in successively at lower hierarchical network levels, until the figure's shape, including its concavities, is represented faithfully at the lowest level. The small receptive fields of area V1 allow the network to represent the figural region at a maximal spatial resolution.

The resolution of a conflict between higher and lower areas is illustrated in Figure 7D, which shows the spatio-temporal response profile of area V1 along a line in an image containing a U-shaped figure (Figure 7A). Initially, responses to the boundaries are enhanced in area V1, since these are extracted locally. Thereafter, the model attempts to fill in the regions between boundaries. Note that the response to the interior of the U is enhanced transiently. This response enhancement is caused by excitatory feedback from neurons in areas V2 and V4, where responses to the opening of the U are strongly enhanced. Indeed, this opening is initially extracted as a candidate object, since its size matches the size of the receptive fields in these areas. However, the enhancement of responses to the opening disappears eventually, because the respective V2 and V4 neurons are overruled by feedback from higher areas that contain the correct assignment of figure and background. When feedback of these higher areas is propagated down to area V1, the

response enhancement is restricted to the actual figure (Figure 7D).

DISCUSSION

The present model illustrates a general scheme for combining a grouping operation that binds image elements into coherent regions, with boundary detection and pop out, in a single network architecture. Only two features, right and left oblique, were used in the simulations. However, these features could be replaced by others and more features could be added, without changing the results. It is likely that the mechanisms that are at work during texture segregation on the basis of orientation contrast are similar those responsible for the segregation by other cues. Indeed, the spatio-temporal profile of the figural response enhancement is relatively independent of the feature that differentiates between foreground and background (Zipser et al., 1996). However, not all feature contrasts permit pop out, and the effortless segregation of image regions (Julesz, 1981; Treisman & Gelade, 1980). The differences between feature contrasts that do and do not permit effortless detection of boundaries have been addressed in previous modeling studies (Malik & Perona, 1990; Bergen & Adelson, 1988; Caelli, 1985).

The location and shape of the figural region is labeled at all hierarchical levels of the model by an enhancement of neuronal firing rates. This raises the question of how such a spatial profile of enhanced activity can be read out by other areas of the visual cortex. It is not immediately obvious how other areas could distinguish between figural image regions, and image regions that yield stronger neuronal responses for another reason, for example, because they have a higher contrast. We would like to suggest an attractive possibility for read out, which relies on an additional population of neurons whose firing rate does not depend on figure–ground relationships (nonmodulated cells; Zipser et al., 1996; Lamme, 1995). The shape and spatial extent of the figure can be read out by comparing the activity of these nonmodulated neurons to the activity of neurons that do exhibit the response enhancement. Thus, neurons in other areas can obtain a relatively pure signal reflecting figural shape if they receive excitatory input from the modulated cells and inhibitory input from the nonmodulated cells. This connection scheme compensates for changes in contrast, since contrast changes affect the responses of nonmodulated and modulated cells equally. Another advantage of this type of coding is that the figural region is represented at a high spatial resolution (e.g., in area V1; Figure 7A–C) if the activity of the modulated cells is compared to the activity of nonmodulated cells, but that the neuronal responses also remain available for the extraction of more local features, such as the location, orientation, and depth of individual line elements.

In a relatively elaborate modeling approach, Grossberg and Mingolla (1985, 1993), Grossberg (1994), and Cohen and Grossberg (1984) suggested that boundaries and surface properties are represented by separate neural systems, localized in different functional compartments of areas V1 and V2, which are distinguished on the basis of their cytochrome oxidase staining (Livingstone & Hubel, 1988). In the present model, the connection schemes that mediate boundary detection and grouping of image regions are rather implemented in different layers, and the signals that reflect both processes are carried by the very same neurons, in accordance with physiology. The same connection scheme is reiterated at every hierarchical network level, which keeps the number of parameters that control the network's behavior at a minimum. In spite of its minimal complexity, the model accurately reproduces the spatiotemporal profile of activity in cortical areas during texture segregation in intact animals as well as in animals with extrastriate lesions.

The network has to solve two problems, singleton detection and grouping of coherent image regions, at the same time. The constraints that these two problems impose on the architecture of the network connections are apparently conflicting, a conflict that we referred to as the grouping–segmentation paradox. The first problem is that the network has to make a choice: Which of the regions should be designated as foreground and which one as background? The Gestalt psychologists (e.g., Koffka, 1935) have described rules of perceptual organization that guide this choice in human observers. One rule states that if the image plane is subdivided into two regions that have different sizes, then the smaller region is more likely to be perceived as figure and the larger as background. Another, related rule states that convex image regions are more likely to be perceived as figural, and concave regions as background (Gibson, 1994; Kanizsa & Gerbino, 1976; Koffka, 1935). These Gestalt rules are implemented in the feedforward layers of the network, where neighboring neurons tuned to similar features inhibit each other (Li, 1999; Stemmler et al., 1995; Malik & Perona, 1990; Grossberg & Mingolla, 1985). Inhibition is weakest for neurons that respond to singletons, namely, image components surrounded by dissimilar neighbors. A singleton is therefore favored as the figure because it evokes the strongest activity. Physiological data indicate that inhibitory interactions among neurons tuned to similar features occur in multiple areas, including areas V1[1] (Kastner et al., 1997; Zipser et al., 1996; Lamme, 1995; Sillito et al., 1995; Knierim & Van Essen, 1992; Allman et al., 1985), V4 (Schein & Desimone, 1990; Zeki, 1980), MT (Born & Tootell, 1992; Lagae et al., 1989; Tanaka et al., 1986; Allman et al., 1985), and MST (Tanaka et al., 1986), areas that represent the visual field at various spatial resolutions. In the model, these inhibitory interactions also account for the detection of larger figural regions. The

figure pops out at a hierarchical level where its size best matches the size of the receptive fields.

Recent psychophysical research uncovered that the probability of perceiving a region as figural also increases if that region has a familiar shape (Peterson, Harvey, & Weidenbacher, 1991). To account for this dependence, the feedforward pathway of our model would have to be modified, by including shape-selective neurons, which are indeed abundant in higher visual areas such as area IT (Tanaka, 1995). Previous studies demonstrated that such a bias a favor of familiar shapes can be incorporated in a neural network, if shape selective neurons provide excitatory feedback to neurons in lower areas that have a similar tuning (Vecera & O'Reilly, 1998; Fukushima, 1988). We note that such a connection scheme complies with the proposed architecture, in which feedback connections are between neurons with a similar feature preference.

The second problem that is solved by the network is the identification of the entire figural region, which has to be grouped together. This appeals to another Gestalt rule stating that image elements with similar features are grouped together by the visual system (Rock & Palmer, 1990). Grouping is achieved in the feedback pathway, which labels the entire figure with an enhanced neuronal response, in accordance with cortical physiology (Lamme et al., 1999; Zipser et al., 1996; Lamme, 1995). Two major assumptions were made regarding the specificity of feedback interactions. First, it was assumed that feedback to neurons that are tuned to similar features is predominantly excitatory, whereas feedback to neurons with a different feature preference is inhibitory (see also Hahnloser et al., 1999; Chey et al., 1997; Finkel & Edelman, 1989). In contrast, Rao and Ballard (1999) suggested the opposite connection scheme in a related model. In their model, feedback rather suppresses the activity of neurons with a similar tuning, and we conjecture that such a layout of connections would not reproduce the enhancement of responses to the interior of a figure, as is observed in physiology. To our knowledge, there are no data regarding the functional specificity of feedback connections, so this disagreement awaits to be resolved experimentally. Second, it was assumed that the feedback projection is gated by feedforward activation. Areas that represent the image at a relatively course resolution feed back to areas that contain a more fine-grained representation. Gating prevents the excitatory feedback to spread beyond the region that is occupied by the figure, and as a result, lower areas demarcate the region occupied by the figure at an ever-increasing resolution. Physiological studies on contextual effects as well as studies on the effects of visual attention on neuronal responses support such a gating process. Influences from outside the classical receptive field are particularly strong for neurons that are well-driven by the visual stimulus, and much weaker for cells that receive little bottom up activation

(McAdams & Maunsell, 1999; Treue & Martínez Trujillo, 1999; Zipser et al., 1996; Kapadia et al., 1995; Sillito et al., 1995; Knierim & Van Essen, 1992; Schein & Desimone, 1990; Allman et al., 1985).

In conclusion, the assignment of different roles to feedforward, horizontal, and feedback connections allows the present model to resolve the grouping–segmentation paradox and to account for neurophysiological and psychophysical data on texture segregation in subjects with and without cortical lesions. There is some physiological data implying functional differences between these types of connections (Shao & Burkhalter, 1996). A recent study blocked the activity in one cortical region by cooling, and assessed the effects on other areas (Vanduffel, Payne, Lomber, & Orban, 1997). The activity in higher areas was strongly reduced, but cooling had much weaker effects on lower areas. This indicates that feedforward connections provide the visual drive for their postsynaptic targets, whereas feedback connections have a more subtle, modulator effect (Lamme & Roelfsema, 2000; Crick & Koch, 1998; Hupé et al., 1998). Nevertheless, our understanding of the strength and feature selectivity of feedforward, lateral, and feedback connections is far from complete. We predict that further knowledge about the functional differences between connection types will advance our understanding of important problems in visual perception, of which the grouping–segmentation paradox is but an example.

METHODS

The network is composed of five areas. Each area is subdivided into feedforward and feedback layers. The receptive fields of neighboring neurons are overlapping (by 25%), in all areas higher than V1. Area V1 contains 64×64 units tuned to the left oblique orientation, and the same number of units tuned to the opposite orientation. In higher areas, the image is represented with a decreasing resolution, since at each hierarchical level the number of units is reduced by a factor of four (Burt & Adelson, 1983).

The activity of the network units is described by continuous variables, which would correspond to the mean activity of a group of functionally similar neurons in physiology. Activity in the feedforward pathway is updated according to the following two equations:

$$\tau_1 \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{FF}_{L,i}^A = -\mathrm{FF}_{L,i}^A + w_1 f_{15}^{0.2} \left(\sum_{j \in U} \mathrm{FF}_{L-1,j}^A \right)$$

$$-\frac{w_4 \sum_{k \in V} \mathrm{FF}_{L,k}^A}{1 + w_6 \mathrm{FB}_{L,i}^A} - 3\mathrm{FA}_{L,i}^A \qquad (1)$$

$$\tau_2 \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{FA}_{L,i}^A = -\mathrm{FA}_{L,i}^A + \mathrm{FF}_{L,i}^A \qquad (2)$$

where $\mathrm{FF}_{L,i}^A$ indicates the activity of unit i with feature preference A in the feedforward layer of area L. f is a squashing function:

$$f(x)_s^\theta = 0.5(1 + \tanh(s(x - \theta))) \qquad (3)$$

The connections $(w_1 - w_6)$ are numbered in the equations as in Figure 2. Feedforward input (connection type 1) is provided by U, a neighborhood in area $L - 1$ with eight units that define the unit's receptive field. In the lowest area (V1), however, U corresponds to a single pixel in the input. Lateral inhibition (connection type 4) is provided by a neighborhood V, which contains eight units in feedforward layer that have the same feature selectivity. A further source of input is unit $\mathrm{FB}_{L,i}^A$, the corresponding unit of the feedback layer (connection type 6). Unit $\mathrm{FB}_{L,i}^A$ does not directly excite $\mathrm{FF}_{L,i}^A$ but influences it indirectly, by reducing the impact of lateral inhibition. Thereby, recurrent excitation between layers is avoided (Crick & Koch, 1998). The slope of the initial response transient to a newly presented image 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, $\mathrm{FA}_{L,i}^A$, which has a longer time constant τ_2 (τ_2 was set to 50). This local inhibitory process was included to model the transient responses of visual cortical neurons, but it is not important for figure–ground segmentation.

Each unit in the feedback path has a corresponding unit in the feedforward path from which it receives excitatory input (connection type 5). Activity in the feedback pathway is updated according to:

$$\tau_3 \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{FB}_{L,i}^A = -0.5\mathrm{FB}_{L,i}^A$$

$$+ f_{35}^{0.65} \left(\mathrm{FF}_{L,i}^A \left(w_5 + \sum_{j \in W} \left(w_3 \mathrm{FB}_{L+1,j}^A - w_2 \mathrm{FB}_{L+1,j}^{\bar{A}} \right) \right) \right)$$

$$(4)$$

where W corresponds to a neighborhood in the next higher area that contains eight units with a similar feature preference (connection type 3) and eight units with the opposite feature preference (connection type 2). Feedback from neurons with the same feature preference (A) is excitatory, whereas feedback from neurons with the opposite feature preference ($\bar{A}$) is inhibitory. Note that $\mathrm{FF}_{L,i}^A$, the corresponding neuron of the feedforward pathway, has a dual effect on $\mathrm{FB}_{L,i}^A$, since it activates the unit and also gates the input from the next higher level through a multiplicative interaction. The time constant of the feedback pathway, τ_3, is larger than that of the feedforward pathway and equals 50. Connection weights were as follows: $w_1 = 1.5$, $w_2 = 2.5$, $w_3 = 1.5$, $w_4 = 1.5$, $w_5 = 1$, and $w_6 = 1$.

The model was updated synchronously in the simulations. All simulations were performed with a single set of parameters as indicated above (the entire model is

controlled by a total of 13 parameters). A reasonable fit to the data was obtained if each time step of the model was set to 1.25 msec. In the figures, 40 msec was added to all data points to account for delays before area V1 (Nowak, Munk, Girard, & Bullier, 1995), since the retina and LGN were not included in the model.

Acknowledgments

Note

1. The interactions among V1 neurons that are tuned to the same orientation also depend on the relative location of their receptive fields. Neurons tuned to the same orientation typically inhibit each other, but they excite each other if they are tuned to collinear configurations (Kapadia et al., 2000). The detection of singletons and boundaries is possible in such a scheme as long as the total inhibitory input from neurons tuned to the same orientation exceeds the excitatory input (see also Li, 1999).

REFERENCES

Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local–global comparisons in visual neurons. *Annual Reviews of Neuroscience, 8,* 407–430.

Bergen, J. R., & Adelson, E. H. (1988). Early vision and texture perception. *Nature, 333,* 363–364.

Born, R. T., & Tootell, R. B. H. (1992). Segregation of global and local motion processing in primate middle temporal visual area. *Nature, 357,* 497–499.

Burt, P. J., & Adelson, E. H. (1983). The laplacian pyramid as a compact image code. *IEEE Transaction on Communications, 31,* 532–540.

Caelli, T. M. (1985). Three processing characteristics of visual texture segmentation. *Spatial Vision, 1,* 19–30.

Chey, J., Grossberg, S., & Mingolla, E. (1997). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 14,* 2570–2594.

Cohen, M. A., & Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception and Psychophysics, 36,* 428–456.

Crick, F., & Koch, C. (1998). Constraints on cortical and thalamic projections: The no-strong-loops hypothesis. *Nature, 391,* 245–250.

De Weerd, P., Desimone, R., & Ungerleider, L. G. (1996). Cue-dependent deficits in grating orientation discrimination after V4 lesions in macaques. *Visual Neuroscience, 13,* 529–538.

De Weerd, P., Sprague, J. M., Vandenbussche, E., & Orban, G. A. (1994). Two stages in visual texture segregation: A lesion study in the cat. *Journal of Neuroscience, 14,* 929–948.

Finkel, L. H., & Edelman, G. M. (1989). Integration of distributed cortical systems by reentry: A computer simulation of interactive functionally segregated visual areas. *Journal of Neuroscience, 9,* 3188–3208.

Fukushima, K. (1988). A neural network for visual pattern recognition. *IEEE Computer, 21,* 65–75.

Gibson, B. S. (1994). Visual attention and objects: One versus two or convex versus concave? *Journal of Experimental Psychology, Human Perception and Performance, 20,* 203–207.

Grossberg, S. (1994). 3-D vision and figure–ground separation by visual cortex. *Perception and Psychophysics, 55,* 48–120.

Grossberg, S. (1999). The link between brain learning, attention, and consciousness. *Consciousness and Cognition, 8,* 1–44.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92,* 173–211.

Grossberg, S., & Mingolla, E. (1993). Neural dynamics of motion perception: Direction fields, apertures, and resonant grouping. *Perception and Psychophysics, 53,* 243–278.

Hahnloser, R., Douglas, R., Mahowald, M., & Hepp, K. (1999). Feedback interactions between neuronal pointers and maps for attentional processing. *Nature Neuroscience, 2,* 746–752.

Hupé, J.-M., James, A. C., Girard, P., & Bullier, J. (2001). Response modulations by static texture surround in area V1 of the macaque monkey do not depend on feedback connections from V2. *Journal of Neurophysiology, 85,* 146–163.

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.

Huxlin, K. R., Saunders, R. C., Marchionini, D., Pham, H.-A., & Merigan, W. H. (2000). Perceptual deficits after lesions of inferotemporal cortex in macaques. *Cerebral Cortex, 10,* 671–683.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290,* 91–97.

Kanizsa, G., & Gerbino, W. (1976). Convexity and symmetry in figure–ground organization. In M. Henle (Ed.), *Vision and artifact* (pp. 25–32). New York: Springer.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron, 15,* 843–856.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology, 84,* 2048–2062.

Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1997). Neuronal correlates of pop-out in cat striate cortex. *Vision Research, 37,* 371–376.

Knierim, J. J., & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology, 67,* 961–980.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York: Harcourt Brace & Co.

Lagae, L., Gulyas, B., Raiguel, S., & Orban, G. A. (1989). Laminar analysis of motion information processing in area V5 of the macaque. *Brain Research, 496,* 361–367.

Lamme, V. A. F. (1995). The neurophysiology of figure–ground segregation in primary visual cortex. *Journal of Neuroscience, 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. *Cerebral Cortex, 9,* 406–413.

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

Lamme, V. A. F., Supèr, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology, 8,* 529–535.

Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences, U.S.A., 96,* 10539–10535.

Livingstone, M. S., & Hubel, D. H. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240,* 740–749.

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 7,* 923–932.

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque area V4. *Journal of Neuroscience, 19,* 431–441.

Merigan, W. H. (1996). Basic visual capacities and shape discrimination after lesions of extrastriate area V4 in macaques. *Visual Neuroscience, 13,* 51–60.

Merigan, W. H., Nealey, T. A., & Maunsell, J. H. R. (1993). Visual effects of lesions of cortical area V2 in macaques. *Journal of Neuroscience, 13,* 3180–3191.

Moraglia, G. (1989). Display organization and the detection of horizontal line segments. *Perception and Psychophysics, 45,* 265–272.

Nothdurft, H. C. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception and Psychophysics, 52,* 355–375.

Nowak, L. G., Munk, M. H. J., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual Neuroscience, 12,* 371–384.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience, 13,* 4700–4719.

Peterson, M. A., Harvey, E. M., & Weidenbacher, H. J. (1991). Shape recognition contributions to figure–ground reversal: Which route counts? *Journal of Experimental Psychology,Human Perception and Performance, 17,* 1075–1089.

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 Neuroscience, 2,* 79–87.

Rock, I., & Palmer, S. (1990). The legacy of Gestalt psychology. *Scientific American, 263,* 48–61.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2000). The implementation of visual routines. *Vision Research, 40,* 1385–1411.

Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiological Review, 75,* 107–154.

Schein, S. J., & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *Journal of Neuroscience, 10,* 3369–3389.

Shao, Z., & Burkhalter, A. (1996). Different balance of excitation and inhibition in forward and feedback circuits of rat visual cortex. *Journal of Neuroscience, 15,* 7353–7365.

Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J., & Davis, J. (1995). Visual cortical mechanisms detecting focal discontinuities. *Nature, 378,* 492–496.

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

Tanaka, K. (1995). Neuronal mechanisms of object recognition. *Science, 262,* 685–688.

Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y., & Iwai, E. (1986). Analysis of local and wide-field motion in the superior temporal visual areas of the macaque monkey. *Journal of Neuroscience, 6,* 134–144.

Tanimoto, S. L. (1985). Image processing with hierarchical cellular logic. In H. Freeman & G. G. Pieroni (Eds.), *NATO ASI series,* vol. F18, *Computer architectures for spatially distributed data* (pp. 279–294). Berlin: Springer-Verlag.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception and Psychophysics, 51,* 599–606.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treue, S., & Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399,* 575–579.

Ullman, S. (1984). Visual routines. *Cognition, 18,* 97–159.

Vanduffel, W., Payne, B. R., Lomber, S. G., & Orban, G. A. (1997). Functional impact of cerebral connections. *Proceedings of the National Academy of Sciences, U.S.A., 94,* 7617–7620.

Vecera, S. P., & O'Reilly, R. C. (1998). Figure–ground organization and object recognition processes: An interactive account. *Journal of Experimental Psychology, Human Perception and Performance, 24,* 441–462.

Zeki, S. (1980). The representation of colours in the cerebral cortex. *Nature, 284,* 412–418.

Zipser, K., Lamme, V. A. F., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience, 16,* 7376–7389.