

Separate Processing Dynamics for Texture Elements, Boundaries and Surfaces in Primary Visual Cortex of the Macaque Monkey

Victor A.F. Lamme^{1,2}, Valia Rodriguez-Rodriguez³ and Henk Spekreijse¹

¹Graduate School of Neurosciences, Department of Visual System Analysis, AMC, University of Amsterdam, PO Box 12011, 1100 AA Amsterdam, ²The Netherlands Ophthalmic Research Institute, PO Box 12141, 1100 AC Amsterdam, The Netherlands and ³Laboratory of Visual Perception, Cuban Neuroscience Center, Apartado 6880, Cubanacan, Habana, Cuba

A visual scene is rapidly segmented into the regions that are occupied by different objects and background. Segmentation may be initiated from the detection of boundaries, followed by the filling-in of the surfaces between these boundaries to render them visible. Alternatively, segmentation may be based on grouping of surface elements that are similar, so that boundaries are (implicitly) identified as the borders between elements that are grouped into objects. Here, we present recordings from awake monkey primary visual cortex that show that in late (>80 ms) components of the neural responses a correlate of boundary formation is expressed, followed by a filling-in (also called colouring) between the edges. These data favour a model of segmentation where boundary formation initiates surface filling-in.

Introduction

In the process of perceptual organization the visual system groups and segments the elements of an image into coherent objects and their surroundings. Figure 1, for example, looks like a textured square, overlying a textured background. Apparently, our visual system groups all the line segments of one orientation into a coherent region that segregates from the other line segments. In addition, the segmented regions attain different perceptual qualities, that of 'figure' and 'ground'. A boundary encloses the figure surface and separates it from the background while the background seems to continue behind the figure. This implies that perceptually the boundary belongs to the figure and not to the background, or in other terms, is *intrinsic* to the figure (Nakayama *et al.*, 1989).

Theoretically, boundaries and regions of grouped elements might be interrelated in two different ways. Boundary detection can precede grouping. In the case of Figure 1, that would imply that first a boundary is detected on the basis of local orientation discontinuities, which is followed by the 'filling in' or 'colouring' of the regions inside and outside of this boundary. Alternatively, grouping might precede the boundary formation; similar line segments group together, and boundaries are (implicitly) formed between regions of grouping discontinuity [for a discussion see, for example, Paradiso and Hahn (Paradiso and Hahn, 1996) and Moller and Hurlbert (Moller and Hurlbert, 1997)]. The former view is supported by various experiments on brightness and colour perception. When two concentric circles of different colours are presented, and the border of the inner one is stabilized on the retina (and hence no longer conveys a signal to the brain), the colour of the outer ring spreads over the inner circle, which becomes invisible (Krauskopf, 1963). The perception of brightness of a white patch, presented on a dark background seems to evolve from the edge inwards (Paradiso and Hahn, 1996; Rossi and Paradiso, 1996), and this brightness filling-in is halted by the consecutive presentation of a contour within the patch (Paradiso and Nakayama, 1991). Such brightness filling-in phenomena are not limited to the domain of colour and

luminance. Similar observations were made for a textured figure segregating from the background on the basis of orientation differences (Caputo, 1998). Moreover, it seems that local discontinuities in texture, i.e. texture boundaries, are more important in segregation than similarity of textures within the segregating regions (Nothdurft, 1985, 1992, 1994; Landy and Bergen, 1991). This suggests that also in texture segregation (Julesz, 1994), boundary detection precedes filling-in.

Gestalt psychologists, on the other hand, have emphasized the role of grouping laws in perceptual organization (Rock and Palmer, 1990). This view is supported by findings that global similarity influences the strength of local feature discontinuities in texture segregation (Enns, 1986; Nothdurft, 1994), and also by the finding that similarities may interfere with segregation (Callaghan, 1989; Rivest and Cavanagh, 1996; Moller and Hurlbert, 1997). Segregation furthermore depends on information about surface layout defined by binocular disparity (He and Nakayama, 1994). Also, thresholds for motion and colour segregation are lower, and segregation is faster, for broad vertical target strips than for thin ones (i.e. with identical boundary lengths), suggesting a role for fast region-based segmentation processes (Moller and Hurlbert, 1996).

The relative roles of boundary formation and surface filling-in or grouping have been particularly addressed in the models of Grossberg on preattentive vision, segmentation and figure-ground segregation. In these models, a boundary contour system (BCS) detects boundaries between regions that are filled-in by the feature contour system (FCS). In older versions of the model, the BCS leads the filling-in by the FCS (Grossberg and Mingolla, 1985; Grossberg *et al.*, 1989). In later versions (Grossberg, 1994), however, the two systems interact to form boundary and surface representations that are mutually consistent, and that may explain filling-in phenomenology (Arrington, 1994). In other models, surface signals are used to sharpen boundary signals (Poggio *et al.*, 1985; Lee, 1995).

Neuronal correlates of segregation and grouping have been studied in recent times: neuronal synchrony (Singer and Grey, 1995), as well as response rate modulation in early visual areas (Kapadia *et al.*, 1995; Lamme *et al.*, 1993, 1998a), seem to play a role as correlates of perceptual grouping. Activity mimicking perceptual filling-in has been found in areas V3 and V2 (DeWeerd *et al.*, 1995). Neuronal correlates of boundary detection on the basis of feature differences (Sillito *et al.*, 1995; Chaudhuri and Albright, 1997) or discontinuities (Grosz *et al.*, 1993) have been found in V1. Also, a correlate of figure-ground segregation and global scene perception is found in response modulations in V1 (Lamme, 1995; Zipser *et al.*, 1996). Primary visual cortex obviously also plays an important role in the encoding of basic stimulus features (Hubel and Wiesel, 1968; Schiller *et al.*, 1976). This area therefore seems an ideal substrate

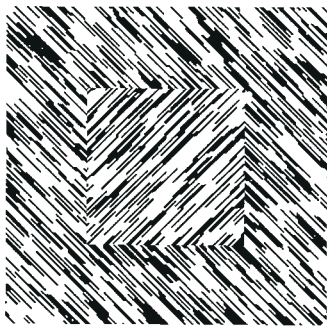


Figure 1. Example of the texture segregation figure-ground displays used in the experiments. In the actual displays, background size was $28^\circ \times 21^\circ$ of visual angle, and figure size was 4° .

to investigate the interrelations between feature detection and grouping, and boundary detection and filling-in.

We recorded from awake macaque monkey primary visual cortex while the animals were viewing displays like the one in Figure 1. Modulation of responses were observed depending on whether the receptive fields of the neurons responded to line elements of the figure or the background, as we have shown before (Lamme, 1995; Zipser *et al.*, 1996). In this study, we focus on the temporal aspects of these modulations, in particular with respect to the modulations evoked by the figure-ground boundary and surfaces. The use of implanted electrodes enabled us to obtain new and very detailed spatio-temporal profiles of the neural image that is cast on V1 when the figure-ground display is presented. It appeared that the modulations exhibit a temporal sequence of processing, going from local feature detection, followed by the detection of texture boundaries, to a neural representation of the relative figure-ground relationships of the surfaces in the scene.

Materials and Methods

Visual Stimulation and Behavioural Control

Stimuli were presented on a 21 in. computer monitor, driven by a #9 GxiTC TIGA graphics board. The display resolution was 1024×768 pixels, the refresh rate was 72.4 Hz. The screen subtended $28^\circ \times 21^\circ$ of visual angle. Trial initiation consisted of the appearance of a 0.2° red fixation spot on a texture of randomly oriented line segments. Monkeys were trained to maintain fixation on this spot. Fixation was considered maintained when the eyes did not at any time leave an imaginary $1.0^\circ \times 1.0^\circ$ window centred around the spot. Eye movements were monitored with scleral search coils, according to the modified double magnetic induction method (Bour *et al.*, 1984). When stored on disk, eye movements were digitized at 400 Hz. Three hundred milliseconds after fixation, the stimulus appeared on the screen, consisting of oriented line segments, so organized that a 4° square figure could be segregated from a full screen background. One stimulation sequence occurred per trial. The animals were rewarded with a drop of apple juice for maintaining fixation until the fixation spot turned off (500 ms after stimulus onset), and subsequently making a saccadic eye movement to the position of the square figure. In part of the recording sessions, the animals were also rewarded simply after having maintained fixation.

Recording of Neuronal Activity

Neural activity was recorded with surgically implanted, Trimel-coated platinum-iridium wires of 25 μm diameter, with exposed tips of 50–150 μm . Impedances ranged from 100 to 350 k Ω , at 1000 Hz. These wires were implanted in the operculum of area 17 (V1) in two macaque monkeys. The obtained signals were amplified (40 000 \times), band-pass filtered (750–5000 Hz), full-wave rectified, and then low-pass filtered (<200 Hz). This resulted in a low-frequency signal, representing the amount (or envelope) of high-frequency (i.e. spiking) activity (Legatt *et*

al., 1980), without any bias for high-amplitude spiking neurons, as might be the case when (arbitrary) amplitude thresholds are used to record multi-unit activity (MUA). This low-frequency signal was digitized (400 Hz), stored on disk and analysed off-line. For further analysis, 16 channels, selected from the implanted electrodes on the basis of signal-to-noise quality of the responses, were recorded simultaneously in each of the two monkeys. Aggregate receptive fields (RFs) of the neurons contributing to each channel were assessed with moving dark bars over a bright background while the animal was fixating. Exact RF positions and sizes were determined off-line from these responses. First, the peak of the activity that was evoked by the moving bar was timed. To compensate for response latency, 50 ms was subtracted, and the position of the bar at that moment in time was calculated. Eight or 16 orientations were used, and each orientation that evoked a response thus resulted in an estimate of the RF position. These position estimates were averaged to obtain the final RF positions that were used in this study. The size of the RF was determined from the response to a bar at the optimal orientation. It was calculated at what range of positions of the bar a response was obtained that exceeded background activity. The responses were considered to exceed background activity when 10% of the peak activity was reached. The width of the bar (0.2°) was subtracted from the thus obtained size. RF eccentricity ranged from 1.3° to 5.45° , and diameter from 0.18° to 1.4° (mean 0.52°). Orientation selectivity was moderately expressed in the MUA. The median orientation selectivity ratio (average response level while a bar of optimal orientation moved over the RF divided by average response for least effective orientation) was 1.94 (mean 2.19, range 1.16–9.03). All recording sites could be driven from either of the two eyes. Strong ocular dominance, as has been reported for layer 4C cells (Hubel and Wiesel, 1977) was absent. Given that electrodes were implanted at a range of depths, and given the binocularity of the signals, it is highly unlikely that the moderate orientation tuning should be regarded as a sign that the recordings expressed mainly layer 4C activity. Instead, taking the RF sizes, tuning ratio and ocular dominance together, we roughly estimate that the electrodes sampled neuronal activity over a distance of some 200–300 μm .

Receptive field positions and sizes stayed stable (within $\sim 0.2^\circ$) for many months of recording in these animals. Tuning characteristics could slowly change over periods of weeks or months. The results presented here were collected during weeks of recording in each monkey. The sets of stimuli were presented in randomized or interleaved ways, so as to avoid possible electrode drifts biasing the results.

Stimulus Positioning and Complementary Stimulus Pairs

We presented the figure-ground stimuli with the figure at various positions relative to the aggregate RFs. As a result, many responses were obtained with a RF either inside or outside the square figure (Fig. 2*a*). Figure and background are composed of orthogonal orientations. This would result in different RF stimulation, depending on the position of the figure relative to RF. Therefore, we always used complementary stimulus pairs, i.e. a particular orientation that was in one trial used for the figure was in another trial used for the background, and vice versa (Lamme, 1995; Lamme *et al.*, 1998b). To achieve this, two full-screen video pages were generated for each orientation, of which a small part was used for the figure and a large part of the other one for the background. Then, only the boundary between the figure texture and the background texture was changed for the different positions of the figure. Thus, by measuring and averaging V1 responses to the complimentary stimuli, we could ensure that, regardless of whether the RF falls in the figure or on the background, the RF was exposed on average to the same set of local features. Notice, however, that when the RF falls on the figure-ground edge, the complementary stimulation is only strictly balanced if the RFs behave like linear spatial filters; we will come back to this point in the Results section.

Data Analysis

To calculate post-stimulus time responses, 500 ms epochs following stimulus onset were averaged from those trials where the animal had fixated and responded with a correct saccade. The mean of the signal obtained at 0–30 ms after stimulus presentation was subtracted from the signal. For all practical purposes, this can be considered the amount of activity that was present while the animal fixated the pre-stimulus texture

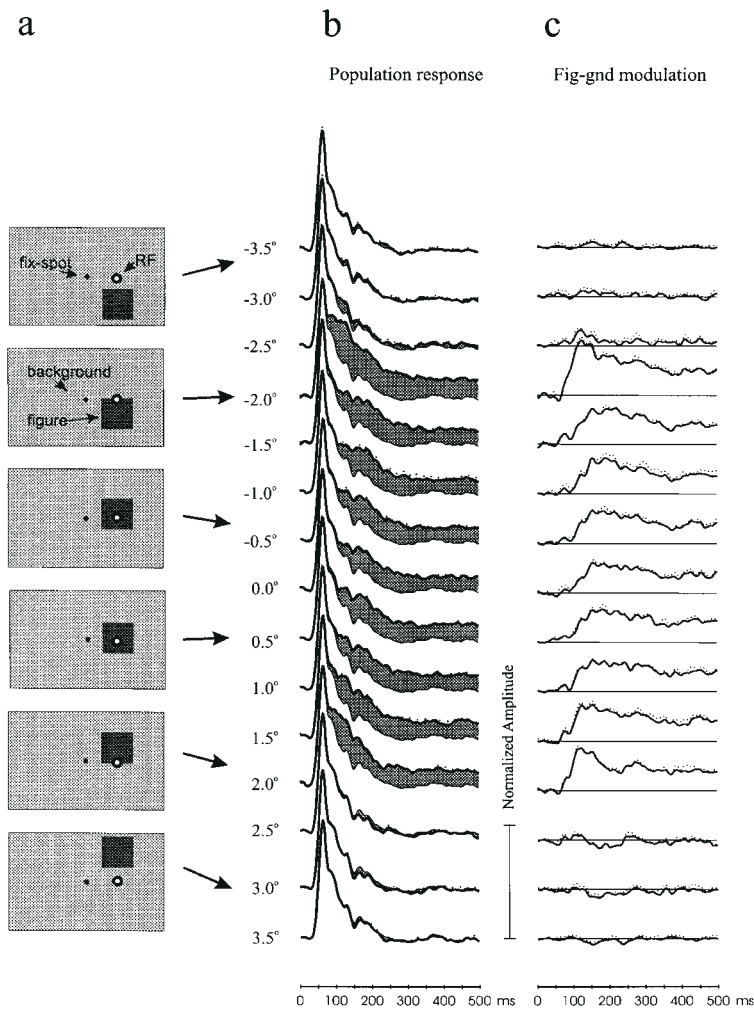


Figure 2. (a) Responses to texture stimuli were sampled with the figure at various positions relative to the RF. Different shades of grey represent different texture orientations. To balance local RF stimulation, complementary stimuli were used (Materials and Methods), only one of which is shown. (b) Average population response (32 recording sites; two monkeys), with the square figure at 15 different vertical positions relative to the RF centre. Position of the centre of the 4° wide figure is given relative to the RF centre. The top and bottom three responses are for positions when the RF is on background, the centre nine responses for the RF on figure or figure edge. Thick lines are responses, thin lines are the response to background (1.5° or more away from the edge). The difference between the two responses are indicated by grey shading. (c) The difference between figure and ground responses (grey shading in b) plotted in isolation. In all plots, dotted lines give standard errors of mean.

(Lamme, 1995; Lamme *et al.*, 1998b). Some sites exhibited activity that was locked to the monitor frame rate (72.4 Hz). Therefore a digital 72.4 Hz notch filter was applied. The displayed responses were additionally smoothed with 1-2-1 windows.

For all electrodes, we wanted to sample 15 positions of the RF relative to figure and ground (Fig. 2). However, for any specific position of the figure some RFs were inside the figure, while others were on the boundary or the background. We therefore used a more extensive set of figure positions that enabled simultaneous recording from electrodes, such that each electrode was used an equal number of times (~200 averages per position, per monkey) for all 15 positions. To get a population average response for all positions of 'RF relative to figure' (Figs 2 and 3), responses from the different electrodes were averaged. This was done such that those responses were averaged together that are identical with respect to the relative positions of figure and RF. In other words, RF₁ (3.5° relative to figure centre) is averaged with RF₂ (3.5° relative to figure centre) and with RF₃ (3.5° relative to figure centre), etc., which may all be different *absolute* positions of the figure. In this alignment procedure, vertical positions of the RF relative to the figure were rounded to the figure position step interval (0.5°).

Also, at different electrodes, responses were obtained that differed strongly in their magnitude. Suppose that at one electrode the responses that are obtained are twice as big as at another electrode; with normal

response averaging the response of the first electrode would have twice the effect on the average as the second. We wanted all electrodes to have an equal contribution to the population average. Therefore we normalized at each electrode all responses obtained for all stimulus conditions to the maximum response obtained at that electrode. In this way all electrodes contributed equally, but relative response differences between conditions (e.g. the different figure positions) at each electrode remained unaltered. Also, timing differences could not have been influenced by this procedure.

Results

Figure-Ground Modulation at the Population Response Level

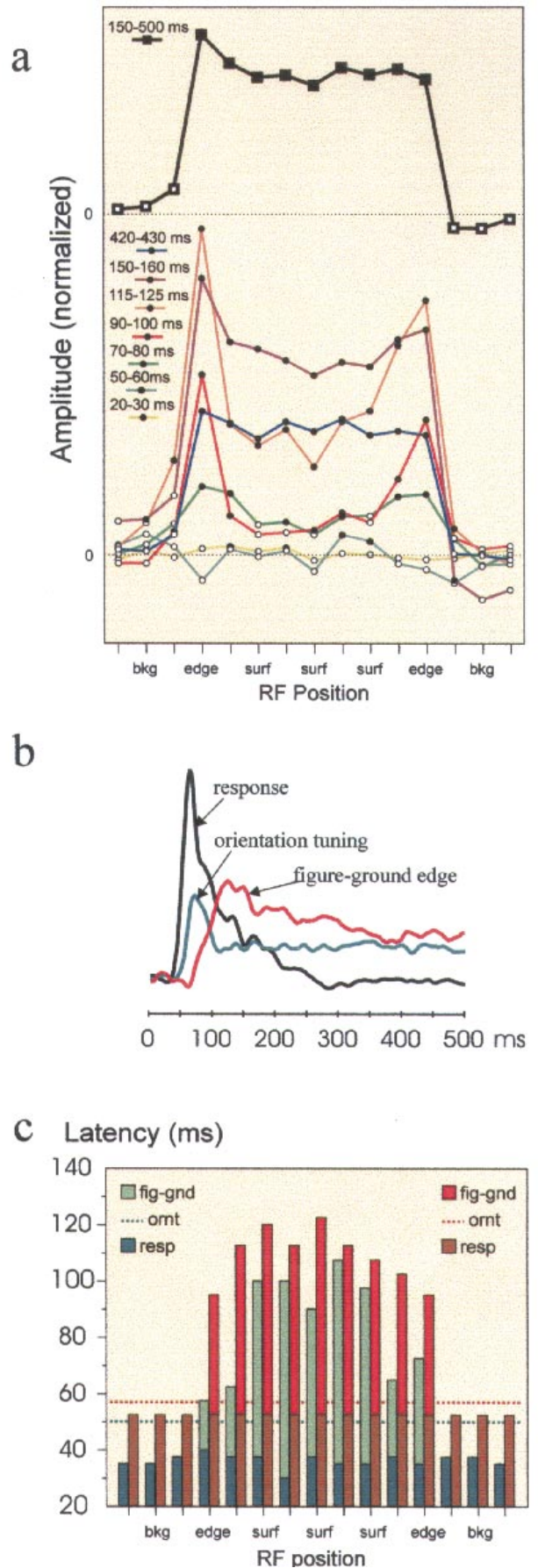
Neural activity was recorded with a square texture figure (Fig. 1) at various positions (Fig. 2a) relative to the V1 RFs. Responses from all electrodes of both monkeys were averaged to obtain population average responses for each position of the figure relative to the RF (Materials and Methods: data analysis). Complementary stimulus pairs (Materials and Methods) were used in order to have both orientations contribute equally to figure and

Figure 3. (a) (Top half): Average response strengths of the difference responses between 150 and 500 ms of Figure 2c (black squares and line), for the different positions of figure versus RF. Filled squares indicate samples that are significantly ($P < 0.01$) different from zero. (Lower half): Time slices (10 ms each) of the responses of Figure 2c. Filled circles represent samples that are significantly ($P < 0.01$) different from zero. (b) Population response to background ('response'), compared to the difference between response to optimal and suboptimal orientation of texture ('orientation tuning'), and to the difference response of Figure 2c for position -2.0° ('figure-ground edge'). (c) Latencies of response (resp), orientation tuning (ornt) and figure-ground enhancement (fig-gnd) were estimated by determining the onset of the first consecutively significant 50 ms epoch (green shades), or by determining the time at which 50% of peak value was obtained (red shades). These are shown for the 15 positions of figure relative to RF.

ground responses (note that at the boundary between figure and ground this would only strictly hold if the RFs behave like linear spatial filters; we will come back to this point below). Figure 2b shows these average population responses. Fifteen positions can be discerned, three such that the background is overlying the RFs, nine such that the figure (or its boundary) is overlying the RFs and three again with the background overlying the RFs. Thin lines in Figure 2b show the average response to background far away from the figure ($>1.5^\circ$ from the edge). Control experiments showed, in line with previous results (Lamme, 1995) that the background response 1.5° away from the edge is not different from background responses obtained much further away (up to 8°) or responses to background without any figure present in the display. The difference between the background response and the responses to the 15 positions is indicated by grey shading. These differences are plotted for the 15 positions in Figure 2c.

Several features of these responses confirm our earlier results on figure-ground modulation (Lamme, 1995). When the RF is on the background, responses are uniformly lower than when the receptive field is on the figure. In Figure 3a, we plotted the average of the amplitude between 150 to 500 ms of the difference responses obtained at the different positions (black line and squares). Figure versus ground response modulation is as large (after ~ 150 ms) for figure responses close to the boundary between figure and ground as for responses within the centre of the figure. Also, there is no modulation at the immediate outside of the boundary between figure and ground (open squares).

What we are interested in here is how this modulation evolves temporally from stimulus onset; in particular, the time interval before 150 ms shows specific features that have not been observed before. In Figure 3a we show (in various colors) 10 ms time slices of the difference responses of Figure 2c. The figure could be read as having a vertical scrolling bar going through the responses of Figure 2c and plotting the response amplitude at each position. Points that deviate significantly ($P < 0.01$) from zero have been marked as solid circles, non-significant deviations as open circles. Before the neurons start to respond (20–30 ms) the slice is flat. But also at the peak of the neuronal response (50–60 ms) the slice is flat, indicating that the figure-ground modulation is strongly delayed with respect to the response itself. Between 70 and 80 ms the first hints of modulation are observed, which are strongest at the boundaries between figure and ground. Between 90 and 100 ms, we observe an intriguing phenomenon; the modulation obtained at the boundaries is now strongly present, while the modulation for the remainder of the figure surface is still almost absent. The edge modulation peaks at 115–125 ms, while at that point figure surface modulation is still evolving. Only at 150–160 ms is the modulation for the whole figure surface at full strength. Modulation subsequently decays very slowly, but remains uniform for all positions inside and including the figure boundaries until the end (420–430 ms).



Some other points in Figure 3a than those discussed above are marked as significant deviations from zero, for example some of the 'surface' responses of the 70–80 or the 90–100 ms intervals. Does that imply that some form of surface modulation already occurs very early, almost at the same time as boundary modulation? Note that some of these points are significant in the 70–80 ms stretch, but then again fall out of significance at 90–100 ms, or the other way around. Some of the points in Figure 3a will be false positively assigned as significantly different from zero, simply because we are dealing with multiple comparisons (200 samples/position). However, it cannot be fully excluded that some very weak and transient early surface enhancement occurs. But to conclude that surface enhancement occurs before 100 ms would also not be warranted (see also below).

Responses are thus identical for all 15 positions up to ~70 ms after stimulus onset. The early activity is thus only determined by the local features presented in the receptive field, which are identical for all positions by means of the complementary stimulus pairs. Remarkably, responses up to 70 ms are also identical for the two positions where the figure–ground boundary is overlying the RF. The complementary stimulus design in this situation would only strictly hold in case of linear summation within RFs. Apparently, the aggregate RFs behave like linear spatial filters at the population response level, at least up to 70 ms.

An important question is whether the effects observed between 70 and 100 ms at the boundary are mediated by local RF tuning mechanisms, or by other local or global mechanisms. An obvious local mechanism would be selectivity for the orientation (or any other feature) of the locally present texture. We averaged background responses to the texture that yielded the largest response and responses from the least effective texture. The difference between these two is shown in Figure 3b, in composite with the average response to background, and the difference responses obtained for the figure–ground edge. Tuning to the orientation of the texture is somewhat later than the response itself, but clearly earlier than the figure–ground boundary enhancement. From this we conclude that the figure–ground boundary enhancement is caused by a mechanism that is different from local RF tuning.

To get objective measures of the latency differences between the various responses we used two methods: one based on a statistical significance criterion, the other based on a time to peak criterion, which is independent of the number of averages used. For the first method, we determined for each sample of the responses whether it was significantly different from zero at the $P = 0.01$ level. Because there are 200 samples/response, many 'significant' samples can be expected to occur just by chance. Simply looking for the first significant sample thus would lead to erroneous latency estimates. An often used and more robust measure is to look for consecutive stretches of significant samples (Maunsell and Gibson, 1992; Munk *et al.*, 1995; Roelfsema *et al.*, 1998). The start of such a stretch would be a much more reliable estimate of latency. Here, we determined the start of the first continuous 50 ms epoch of response (or enhancement) that was significant at the 1% level. The results of this analysis are shown in shades of green in Figure 3c. The second method of latency estimation was to calculate the time at which the amplitude of the response (or the enhancement) reached 50% of its peak value. The results of this calculation are shown in shades of red in Figure 3c. Both methods yielded similar results in three respects: latencies of the response itself (30–40 ms, method 1; 50 ms, method 2) are always shorter than the latency of orientation tuning (50 ms; 57.5 ms), which is in

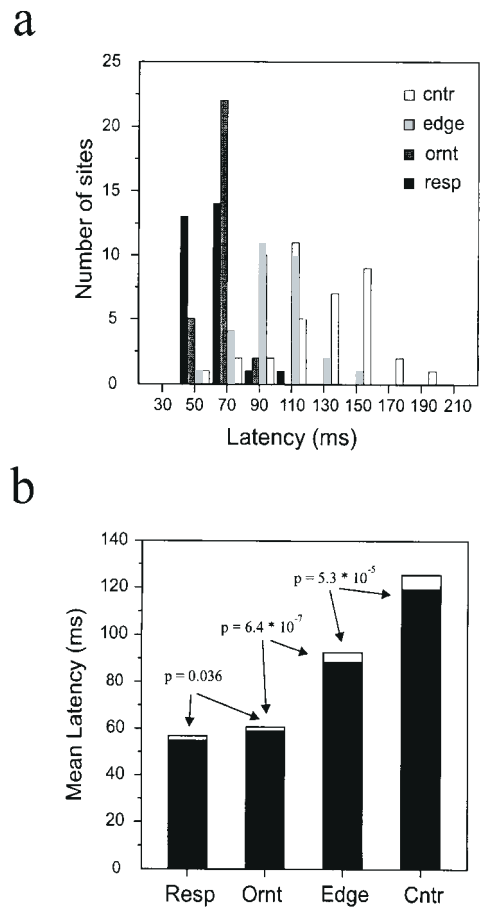


Figure 4. (a) Distribution of latencies at individual recording sites for the background response (resp), orientation tuning (ornt), figure–ground boundary enhancement (edge) and figure–ground surface enhancement obtained at the centre of the figure. (b) Means of the distributions given in (a) (black bars). White segment of each bar indicates the SEM. Significance levels are shown for paired *t*-tests.

turn always shorter than the latencies of figure–ground edge enhancement (57.5–70 ms; 90 ms). Figure–ground edge enhancement is always faster than enhancement obtained for central figure surface positions (90–107.5 ms; 112.5–122.5 ms).

A remarkable feature is observed for the latency measures that are obtained with the second method (Fig. 3c, red bars): latencies of the figure–ground enhancement are shortest at the boundary (90 ms), and increase in ~7.5 ms steps for every 0.5° towards the centre of the figure (120 ms). In other words, there is a gradual increase in latency, going from the figure–ground edge towards the centre.

Feature, Boundary and Figure Surface Relationships at Individual Sites

Like we did in Figure 3 for the population response, we calculated latencies of response, orientation selectivity and figure–ground enhancement at the boundary and the centre of the figure for individual recording sites. In individual cases, method 1 could not be used due to signal-to-noise limitations; although figure–ground enhancement was significant at the $P < 0.01$ level in 25 of the 32 cases, too few responses showed reliable consecutive stretches of significant samples. Latencies of all four response types could be calculated for 29 of the 32 sites according to method 2 (when 50% of peak was reached before 30 ms or after 250 ms, results were discarded). Latency of

response was calculated from the average of the two most extreme background positions, latency of orientation tuning from the difference in response for two texture orientations at background, latency of boundary enhancement from the two edge positions, and latency of surface centre enhancement from the three most central surface positions. The distributions of these four types of latencies are shown in Figure 4a, their means in Figure 4b. Significant differences are found between the latencies of all four types of responses (one-way ANOVA $P = 6 \times 10^{-12}$; for individual comparisons, see Fig. 4b). At the population mean level we thus observe a sequence of processing, starting with RF-based orientation-selective responses, followed by response enhancement caused by the figure-ground boundary, which is then again followed by response enhancement related to the figure-ground surface relationships.

Although the distributions of Figure 4a have significantly different means (Fig. 4b) they do overlap. At individual sites, therefore, the sequential processing of feature, boundary and surface that is observed at the population level might not strictly hold. For example, edge enhancement from site to site ranges between 50 and 150 ms, while centre enhancement ranges between 50 and 190 ms (Fig. 4a). So at some individual sites it might occur that centre enhancement occurs *before* edge enhancement. If, however, centre enhancement is induced by edge enhancement, it should always follow it. We have plotted for individual sites the latencies of response versus orientation tuning (Fig. 5a), of orientation tuning versus figure-ground edge enhancement (Fig. 5b), and of edge versus surface enhancement (Fig. 5c). In 25/29 cases, local RF tuning precedes figure-ground boundary enhancement (Fig. 5b), strengthening our earlier conclusions that boundary enhancement is caused by a mechanism that is distinct from orientation tuning, probably sensitivity to orientation contrast.

In 27/29 sites, boundary enhancement has a shorter or equal latency than figure centre enhancement (Fig. 5c). What we wanted to know is whether this implies that at individual sites, the sequential relationship between edge and centre enhancement also exists. Our second test is therefore aimed at trying to establish how significant it is to find that in 27/29 sites centre enhancement does not occur before edge enhancement. Obviously this depends on the distributions of both latencies. For example, if we have totally separate distributions, just because of that, centre modulation would always follow edge modulation. A finding of all sites showing this behaviour would not give extra information. If we have fully overlapping distributions, just by chance half would normally follow and half would not. A finding of >50% of sites showing that centre modulation follows edge modulation would give us extra information. In the case of the partially overlapping distributions found here the probability 'an sich' of centre enhancement following edge enhancement is 0.794. This probability can be fed into a binomial distribution calculation which then tells us how high the probability is of finding 27 (or more) of 29 cases of centre modulation following edge modulation. This is 0.045. This means that the finding of centre modulation following edge modulation in 27/29 cases is significantly more than can be expected by chance. This is a remarkable finding, which indicates that, at each site, boundary modulation might occur at different times (sometimes even later than the average latency of centre modulation), but nevertheless is always *followed* by centre modulation.

This finding is further strengthened by the following consideration. The boundary between figure and ground was placed from one position to the next in discrete 0.5° steps. Because of

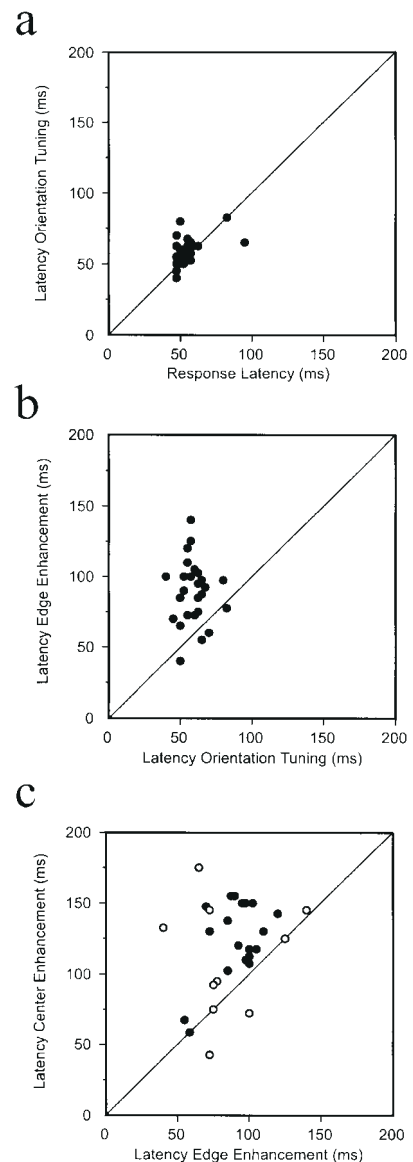


Figure 5. Correlations for individual recording sites between response latency and latency of orientation selectivity (a), between latency of orientation selectivity and of figure-ground edge enhancement (b), and between latency of figure-ground edge enhancement and of enhancement at the centre of the figure (c). In the latter graph, grey-filled symbols represent sites for which the figure-ground edge was not presented on the RF centre.

this, for some small and awkwardly positioned RFs the boundary was projected only in the periphery of the RF, instead of its centre, or in some cases might even have 'missed' the RF. For these electrodes, the 'edge' responses could therefore be diluted by 'surface' responses. Electrodes for which this might have occurred are shown in light grey in Figure 5c. When these are excluded, *all* sites show that edge enhancement occurs before figure surface enhancement.

These results show that for the population of V1 neurons, the filling-in process is induced by the boundary detection. However, whether they also indicate that boundary detection at each individual site induces the filling-in at that site remains unresolved, since we were unable to establish whether the latency difference between boundary and surface signals was significant at each site.

Summary of Results

At the population response level, a clear temporal sequence is observed in the processing of a figure-ground display like Figure 1. First, cells start to respond, and this is very shortly followed by orientation selectivity for texture elements. At considerable longer latency the boundary between figure and ground elicits a response enhancement. This is followed by response enhancement for the figure surface, with longer latencies for positions further away from the figure-ground edge (Fig. 3c). This suggests a process of filling-in starting from the edge. This is corroborated by the finding that at individual recording sites edge enhancement may occur at a range of times, but is always followed by surface enhancement.

Eye Movement Controls

Our results were obtained in monkeys that had to fixate within a window of $1^\circ \times 1^\circ$. At first sight, it might therefore be surprising that dramatic changes in response were observed for shifts in stimulus position as small as 0.5° . However, monkeys fixated much more precisely on the 0.2° fixation spot than demanded. On average, 95% of fixations were within 0.2° . This can also be observed in our finding that some of the RFs were as small as 0.2° .

Another concern is whether the stimuli induce small eye movements within the window of fixation that might contribute to the effects reported. We have published many controls for this, showing that this is not the case (Lamme, 1995; Zipser *et al.*, 1996; Lamme *et al.*, 1998b). The eye movement controls that we performed on the data presented here have been published elsewhere (Lamme *et al.*, 1998b), together with additional experiments showing that the effects are not attributable to eye movements or spatially focused attention.

Feature, Boundary and Surface Detection

Three qualitatively different phases can thus be observed in the spatio-temporal response profiles that are unrolling in V1 after the presentation of a segregating texture. We argue that three different processes, separated by different temporal dynamics and brought about by different properties of the neurons and their connections within V1 underlie these phases. It appears that local feature detection governs the initial response phase in V1: almost as soon as cells start to fire, orientation selectivity is expressed in their responses. This is in line with other evidence showing that orientation selectivity is present in the early spikes of V1 neurons (Celebrini *et al.*, 1993), and in early synaptic potentials, that arise from lateral geniculate nucleus (LGN) input alone (Ferster *et al.*, 1996). This response phase is therefore most likely generated from feedforward processing from LGN to V1 and through the cortical pathways within V1 (Mitzdorf and Singer, 1979; Lund, 1998).

The initial feedforward phase is followed by the detection of surface boundaries. The modulation of the response that is caused by the figure-ground boundary occurs much later than the expression of orientation tuning (Fig. 3b). From this we conclude that a different mechanism underlies to it. The recent findings of some V1 cells being sensitive to orientation contrast (Sillito *et al.*, 1995) or abutting gratings (Grosz *et al.*, 1993) are most likely underlying this second response phase. Specific sensitivity for orientation contrast was also reported by Caputo (Caputo, 1998). Another candidate would be end-stopping (Hubel and Wiesel, 1968, 1977), as the line segments

are typically truncated at the boundary. End-stopping has also been coined as a mechanism to create selectivity for corner-like features, and the abutting orthogonal line segments at the boundary form such corner features. This phase may be mediated by feedforward mechanisms in combination with horizontal connections.

Finally the cells seem to represent the figure-ground relationships of surfaces in the scene with a higher response level for the figure than for the background. This higher level persists for as long as the stimulus is on. The third phase is clearly an expression of influences from beyond the classical receptive field (Allman *et al.*, 1985; Gilbert and Wiesel, 1990; Knierim and Van Essen, 1992; Lamme, 1995; Kapadia *et al.*, 1995; Zipser *et al.*, 1996). These have often been interpreted in terms of RF centres and (inhibitory or excitatory) surrounds. We have, however, shown that this representation is rather cue-invariant (Zipser *et al.*, 1996) and bears no direct relation to the RF properties of the V1 cells (Lamme, 1995). Moreover, it seems to be most closely related to perceptual interpretation of the scene (Kapadia *et al.*, 1995; Lamme, 1995). It is also absent in the anaesthetized animal (Lamme *et al.*, 1998b). It is very likely that the third phase is an expression of horizontal connections within V1 (Gilbert, 1992) in combination with feedback from extrastriate areas (Salin and Bullier, 1995). In fact, we have evidence that the surface signals, and not the boundary signals, are abolished by extra-striate lesions (Lamme *et al.*, 1998a).

Our results strongly argue in favour of a model of preattentive vision where boundary detection precedes and initiates surface filling-in (also called colouring). Boundary signals are very sharp right from the start, i.e. there seems to be no boundary contraction process (Lee, 1995). The interaction between boundary signals and surface signals seems to be mostly one way, from boundary to surface. That is not to say that surface signals never influence boundary formation. We used rather clear-cut, segmentable images, and it might be that when there is more noise or ambiguity, an influence of surface signals on the neural representation of the boundary is seen.

The Timing of Boundary versus Surface

In these experiments, the response modulation at the centre of the figure lagged the boundary enhancement by ~ 30 ms. Given the size of the figure, this would correspond to a speed of filling-in of $67^\circ/\text{s}$. Propagation speed of brightness filling-in was estimated in several studies. Speeds of 110 – $150^\circ/\text{s}$ (Paradiso and Nakayama, 1991), 5 – $10^\circ/\text{s}$ (Paradiso and Hahn, 1996), 140 – $180^\circ/\text{s}$ (Rossi and Paradiso, 1996) and $19^\circ/\text{s}$ (Davey *et al.*, 1998) have been reported. Our results are within the same range. Using so called ‘phantom contour’ stimuli, Rogers-Ramachandran and Ramachandran (Rogers-Ramachandran and Ramachandran, 1998) found a fast (15 Hz) texture contour system, and a slower (7 Hz) surface discrimination system. In his masking experiments, Caputo (Caputo, 1998) found a latency of 40–80 ms for segregation contours and a latency of 120 ms for the spreading of surface filling-in. Although these results are difficult to compare, they both suggest that boundary detection is twice as fast as surface-filling-in. We find something similar when we compare the latencies of the boundary and surface signals with the latency of the feature detection system; surface latency minus feature latency ($122.5 - 57.5 = 65$ ms) is about twice as long as boundary latency minus feature latency ($95.0 - 57.5 = 37.5$).

The kind of filling-in discussed here – also referred to as colouring – is a rather different process than the much slower type of filling-in that is observed when surrounding stimulus

properties 'invade' a region of the visual field that receives no visual input, as is the case for the blind spot, or for artificial scotomata (non-stimulated regions). In those cases, it takes ~5–10 s (Ramachandran and Gregory, 1991; De Weerd *et al.*, 1995) before this type of filling-in starts to operate.

Notes

We thank Kor Brandsma and Jacques de Feiter for biotechnical support, and Peter Brassinga and Hans Meester for technical support. This work was supported by a grant from the Royal Dutch Academy of Sciences (KNAW) to V.L.

Address correspondence to Victor A.F. Lamme, Graduate School of Neurosciences, Department of Visual System Analysis, AMC, University of Amsterdam, PO Box 12011, 1100 AA Amsterdam, The Netherlands. Email: v.lamme@amc.uva.nl.

References

- Allman J, Miezin F, McGuiness E (1985) Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci* 8:407–430.
- Arrington KF (1994) The temporal dynamics of brightness filling-in. *Vis Res* 24:3371–3387.
- Bour LJ, Van Gisbergen JAM, Bruijns J, Ottes FP (1984) The double magnetic induction method for measuring eye movement – results in monkey and man. *IEEE Trans Biomed Engng* 31:419–427.
- Callaghan TC (1989) Interference and dominance in texture segregation: hue, geometric form, and line orientation. *Percept Psychophys* 46:299–311.
- Caputo G (1998) Texture brightness filling-in. *Vis Res* 38:841–851.
- Celebrini S, Thorpe S, Trotter Y, Imbert M (1993) Dynamics of orientation coding in area V1 of the awake primate. *Vis Neurosci* 10:811–825.
- Chaudhuri A, Albright TD (1997) Neuronal responses to edges defined by luminance vs. temporal texture in macaque area V1. *Vis Neurosci* 14:949–962.
- Davey MP, Madess T, Srinivasan MV (1998) The spatiotemporal properties of the Craik–O'Brein–Cornsweet effect are consistent with 'filling-in'. *Vis Res* 38:2037–2046.
- De Weerd P, Gattass R, Desimone R, Ungerleider LG (1995) Responses of cells in monkey visual cortex during perceptual filling in of an artificial scotoma. *Nature* 377:731–734.
- Enns J (1986) Seeing textons in context. *Percept Psychophys* 39:143–147.
- Ferster D, Chung S, Wheat H (1996) Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* 380:249–252.
- Gilbert CD (1992) Horizontal integration and cortical dynamics. *Neuron* 9:1–13.
- Gilbert CD, Wiesel TN (1990) The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vis Res* 30:1689–1701.
- Grosf DH, Shapley RM, Hawken MJ (1993) Macaque V1 neurons can signal 'illusory' contours. *Nature* 365:550–552.
- Grossberg S (1985) Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Percept Psychophys* 38:141–171.
- Grossberg S (1994) 3-D vision and figure-ground separation by visual cortex. *Percept Psychophys* 55:48–120.
- Grossberg S, Mingolla E, Todorovic D (1989) A neural network architecture for pre-attentive vision. *IEEE Trans Biomed Engng* 36:65–83.
- He ZJ, Nakayama K (1994) Perceiving textures: beyond filtering. *Vis Res* 34:151–162.
- Hubel DH, Wiesel TN (1977) Ferrier Lecture. Functional architecture of macaque monkey visual cortex. *Proc R Soc Lond B* 198:1–59.
- Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195:215–243.
- Julesz B (1984) A brief outline of the texton theory of human vision. *Trends Neurosci* 7:41–45.
- Kapadia MK, Ito M, Gilbert CD, 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.
- Knierim JJ, Van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J Neurophysiol* 67:961–980.
- Krauskopf J (1963) Effect of retinal image stabilization on the appearance of heterochromatic targets. *J Opt Soc Am* 53:741–744.
- Lamme VAF (1995) The neurophysiology of figure-ground segregation in primary visual cortex. *J Neurosci* 15:1605–1615.
- Lamme VAF, Van Dijk BW, Spekreijse H (1993) Contour from motion processing occurs in primary visual cortex. *Nature* 363:541–543.
- Lamme VAF, Supér H, Spekreijse H (1998a) Feedforward, horizontal, and feedback processing in the visual cortex. *Curr Opin Neurobiol* 8:529–535.
- Lamme VAF, Zipser K, Spekreijse H (1998b) Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proc Natl Acad Sci USA* 95:3263–3268.
- Landy MS, Bergen JR (1991) Texture segregation and orientation gradient. *Vis Res* 31:679–691.
- Lee T-S (1995) A Bayesian framework for understanding texture segmentation in the primary visual cortex. *Vis Res* 35:2643–2657.
- Legatt AD, Arezzo J, Vaughan HG (1980) Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: effects of volume conducted potentials. *J Neurosci Methods* 2:203–217.
- Lund JS (1988) Anatomical organization of macaque monkey striate visual cortex. *Annu Rev Neurosci* 11:253–288.
- Maunsell JHR, Gibson JR (1992) Visual response latencies in striate cortex of the macaque monkey. *J Neurophysiol* 4:1332–1334.
- Mitzdorf U, Singer W (1979) Excitatory synaptic ensemble properties in the visual cortex of the macaque monkey: a current source density analysis of electrically evoked potentials. *J Comp Neurol* 187:71–84.
- Moller P, Hurlbert AC (1996) Psychophysical evidence for fast region-based segmentation processes in motion and colour. *Proc Natl Acad Sci USA* 93:7421–7426.
- Moller P, Hurlbert AC (1997) Motion edges and regions guide image segmentation by colour. *Proc R Soc Lond B*, 264:1571–1577.
- Munk MHJ, Nowak LG, Girard P, Chounlamountri N, Bullier J (1995) Visual latencies in cytochrome oxidase bands of macaque area V2. *Proc Natl Acad Sci USA* 92:988–992.
- Nakayama K, Shimojo S, Silverman GH (1989) Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception* 18:55–68.
- Nothdurft HC (1985) Orientation sensitivity and texture segmentation in patterns with different line orientation. *Vis Res* 25:551–560.
- Nothdurft HC (1992) Feature analysis and the role of similarity in preattentive vision. *Vis Res* 52:355–375.
- Nothdurft HC (1994) Common properties of visual segmentation. In: Higher-order processing in the visual system (Bock R, Goode JA, eds), Ciba Foundation Symposium 184, pp. 245–268. Chichester: Wiley.
- Paradiso MA, Nakayama K (1991) Brightness perception and filling-in. *Vis Res* 31:1221–1236.
- Paradiso MA, Hahn S (1996) Filling-in percepts produced by luminance modulation. *Vis Res* 36:2657–2663.
- Poggio Y, Torre V, Koch C (1985) Computational vision and regularization theory. *Nature* 317.
- Ramachandran VS, Gregory RL (1991) Perceptual filling in of artificially induced scotomas in human vision. *Nature* 350:699–702.
- Rivest J, Cavanagh P (1996) Localizing contours defined by more than one attribute. *Vis Res* 36:53–66.
- Rock I, Palmer S (1990) The legacy of Gestalt psychology. *Scient Am* 263:48–61.
- Roelfsema PR, Lamme VAF, Spekreijse H (1998) Object based attention in primary visual cortex of the macaque monkey. *Nature* 395:376–381.
- Rogers-Ramachandran DC, Ramachandran VS (1998) Psychophysical evidence for boundary and surface systems in human vision. *Vis Res* 38:71–77.
- Rossi AF, Paradiso MA (1996) Temporal limits of brightness induction and mechanisms of brightness perception. *Vis Res* 36:1391–1398.
- Salin P, Bullier J (1995) Corticocortical connections in the visual system: structure and function. *Physiol Rev* 75:107–154.
- Schiller PH, Finlay BL, Volman SF (1976) Quantitative studies of single cell properties in monkey striate cortex. I–V. *J Neurophysiol* 39:1288–1374.
- Sillito AM, Grieve KL, Jones HE, Cudeiro J, Davis J (1995) Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* 378:492–496.
- Singer W, Grey CM (1995) Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci* 18:555–586.
- Zipser K, Lamme VAF, Schiller PH (1996) Contextual modulation in primary visual cortex. *J Neurosci* 16:7376–7389.