Neuron, Volume *75*

Supplemental Information

The Role of Attention in Figure-Ground Segregation in Areas V1 and V4 of the Visual Cortex

Jasper Poort, Florian Raudies, Aurel Wannig, Victor A.F. Lamme, Heiko Neumann, and Pieter R. Roelfsema

Inventory of Supplemental Information

Supplemental Data

Supplemental Experimental Procedures

6. Eye position controls
7. Effect of difficulty of the curve tracing task on FGM and attentional modulation
8. Functional role of FGM in V1 and V4
9. Latency of the visual response, FGM and attentional modulation
10. Model of cortico-cortical interactions for figure-ground segregation

Supplemental References

Supplemental Data

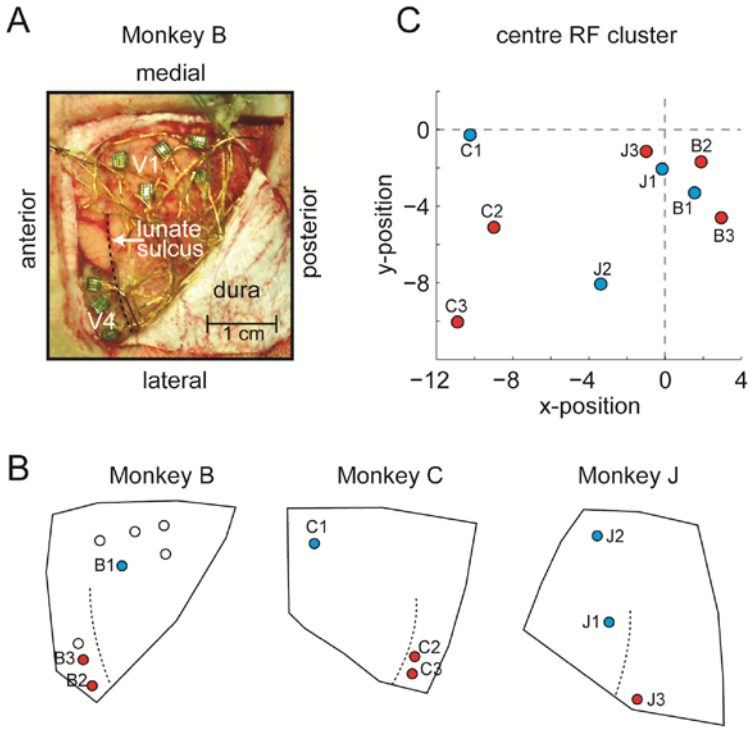

Figure S1. Position of the electrode arrays in cortex and receptive field locations (related to Figure 2).

(A) Picture taken during surgery in monkey B. The arrays are black and the wires that connect the arrays to connectors yellow. Dashed line, lunate sulcus.

(B) Schematic that shows the location of the arrays. The left hemisphere was implanted in Monkey B and the right hemisphere in monkeys C and J. Blue circles represent V1 arrays, red circles represent V4 arrays. White circles are arrays that were not recorded from in these experiments. Dashed line, lunate sulcus.

(C) Position of the receptive fields of the electrode arrays. Dashed lines represent the horizontal and vertical meridian. The RFs in monkey B are in the lower right quadrant of the visual field and the RFs in monkeys C and J in the lower left quadrant.

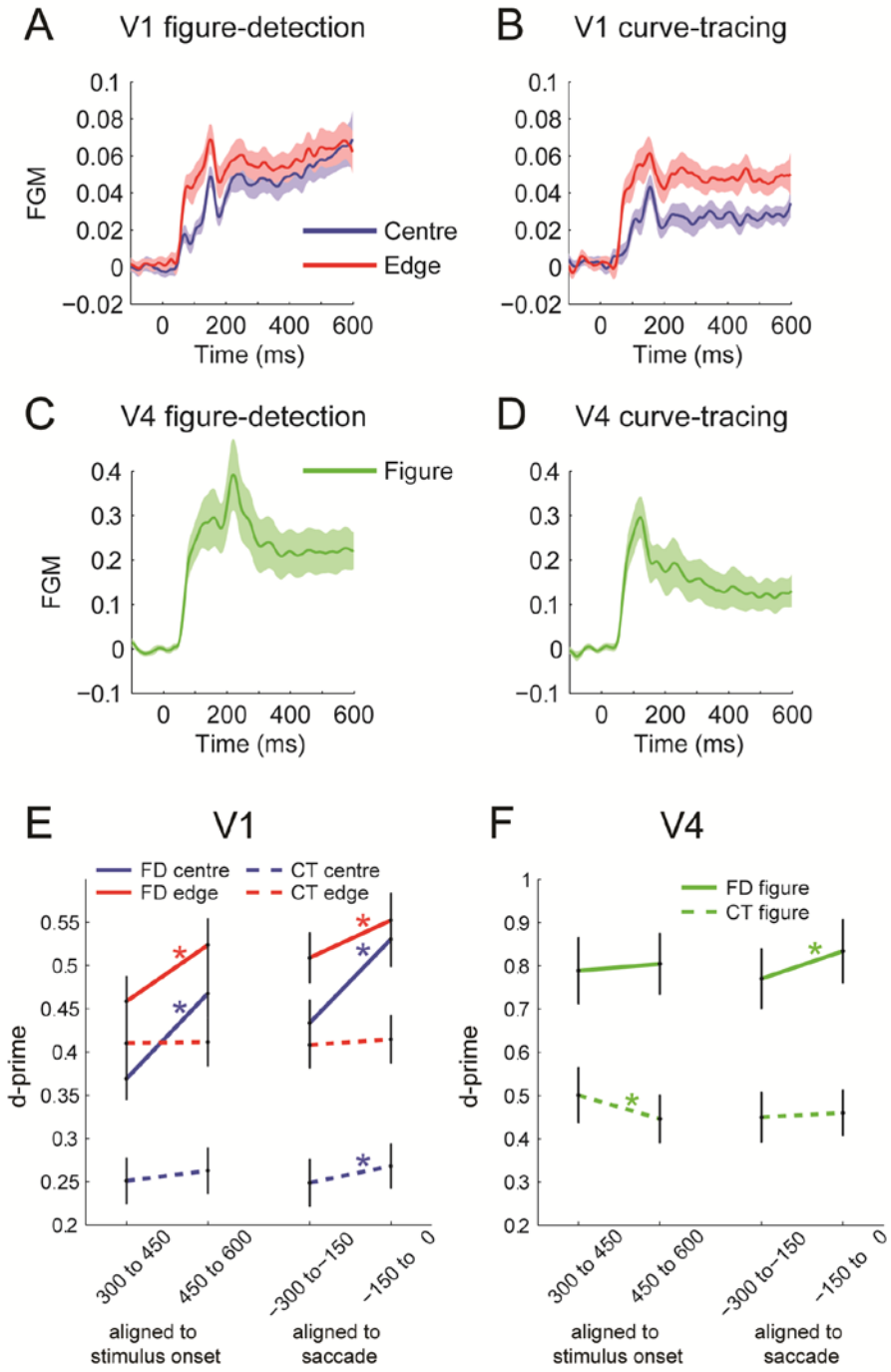

Figure S2. Time-course of figure-ground modulation in V1 and V4 (related to Figure 4 and 6).
(A,B) FGM in V1 at the centre and edge positions in the figure-detection (A) and curve-tracing task (B). Coloured areas around traces indicate ± 2 standard error of the mean.
(C,D) FGM in V4 in the figure-detection (C) task and curve-tracing task (D).
(E) Average d-prime of the V1 response in time windows aligned on stimulus onset and saccade in the figure detection (FD) and curve tracing task (CT), separately for RF positions on the centre and edge of the figure. Error bars indicate SEM.
(F) Average d-prime in V4.

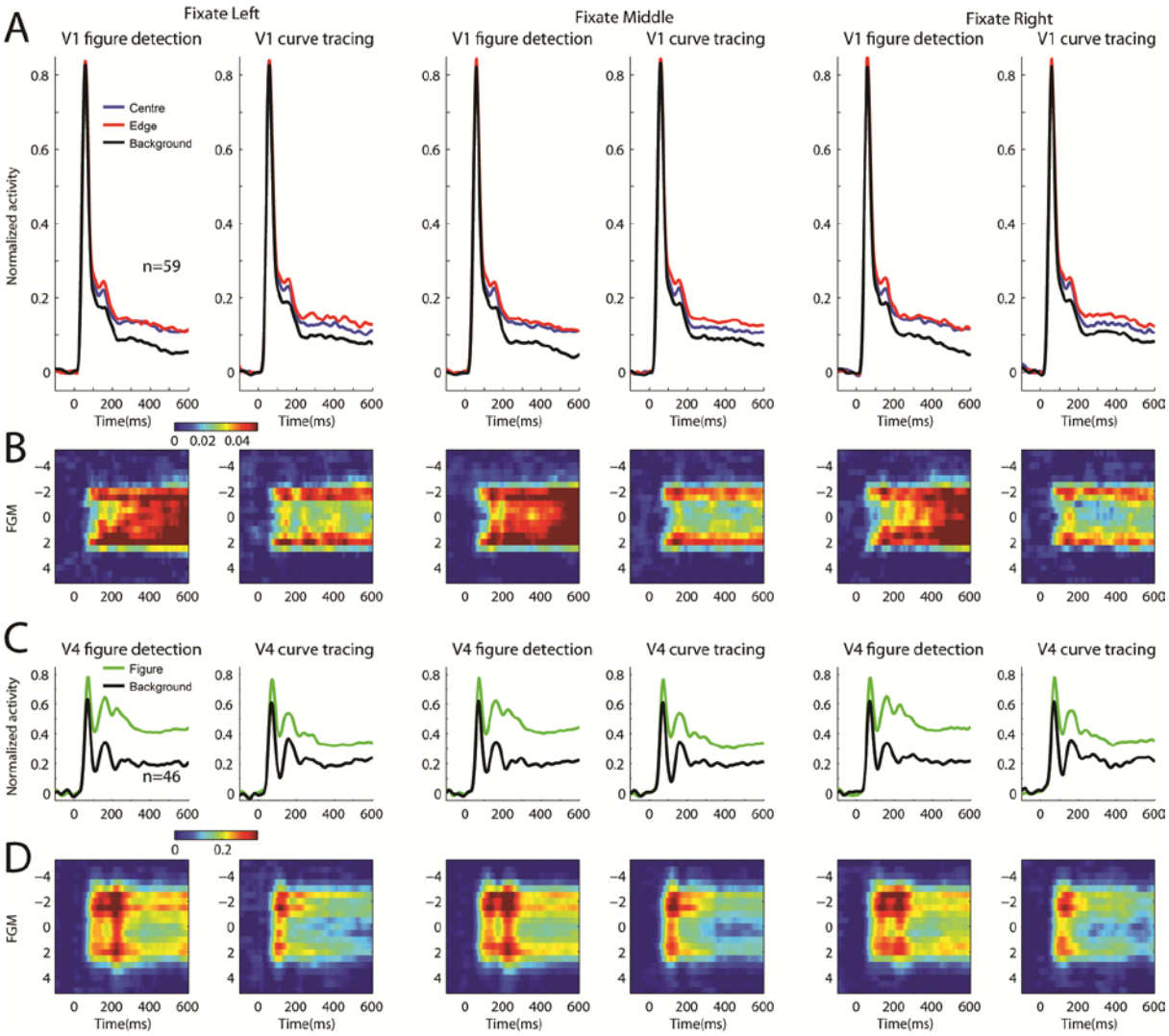

Figure S3. Activity in V1 and V4 as function of fixation position (related to Figure 4 and 6).

(A) Neuronal responses in V1 evoked by image elements of the figure-centre, edge and background (N= 59) in trials where fixation was in the left, middle or right side of the fixation window.

(B) Space-time profile of the FGM in V1.

(C) Responses evoked by the figure and background in area V4 (N=46).

(D) Space-time profile of the FGM in V4.

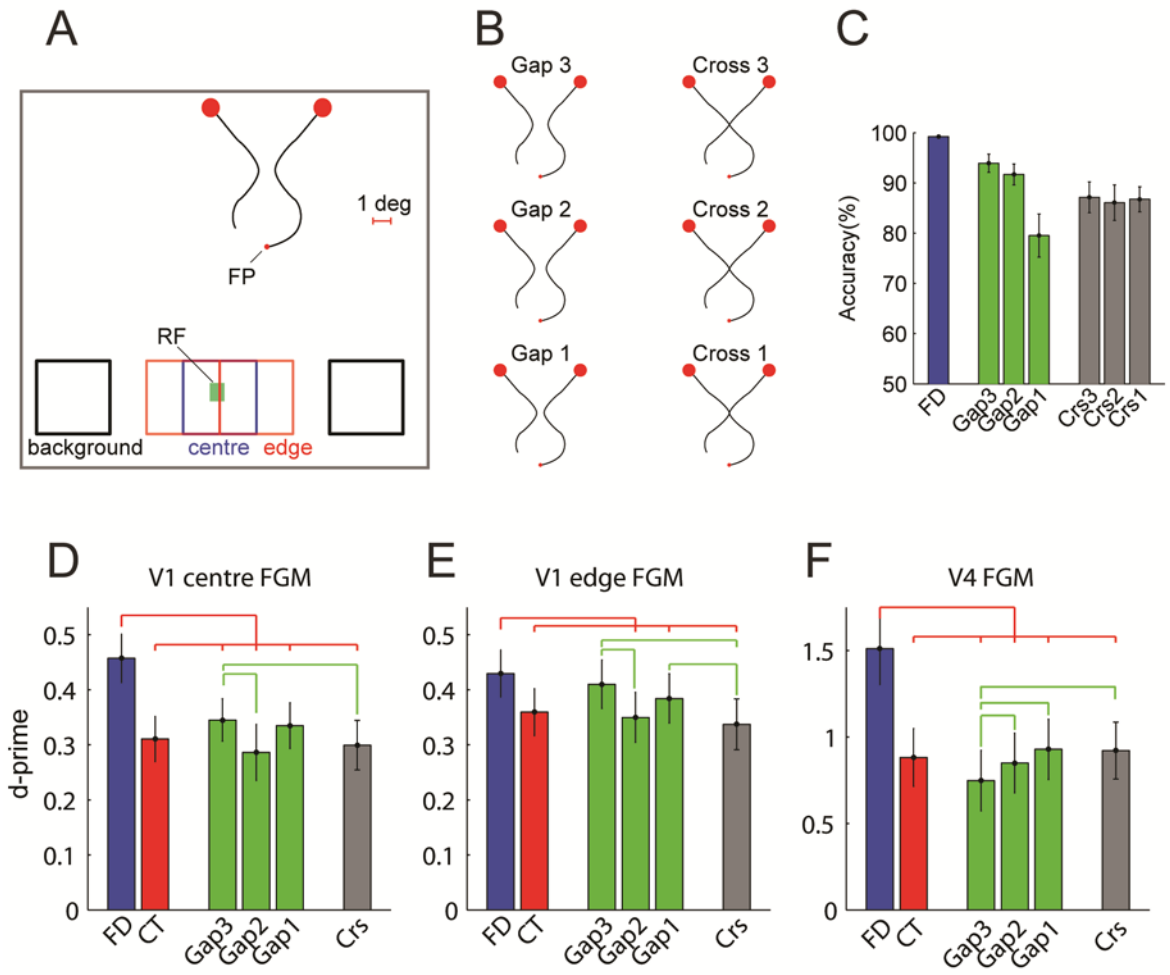

Figure S4. FGM in V1 and V4 as function of the difficulty of the curve-tracing task (related to Figure 4 and 6).

(A) We measured neuronal activity evoked by image elements of the figure-centre, background in the figure-detection task in 2 monkeys while we varied the difficulty of the curve-tracing task.

(B) The different curve tracing stimuli. Left, stimuli with non-intersecting curves, gap size decreases from top to bottom. Right, stimuli with intersecting curves. The angle of the intersection decreases from top to bottom.

(C) Accuracy in the figure detection (FD, blue bar) and curve-tracing task (green and gray bars).

(D) d-primes in V1 for FGM in the figure centre in the figure detection (FD) and curve tracing (CT) task. Green bars show the d-prime separately for the conditions with gap-sizes 1, 2 and 3. The gray bar is the average d-prime in conditions with intersecting curves. Error bars show ± 1 SEM. Red lines indicate significant differences in d-prime between the FD and CT tasks. Green lines indicate differences between CT conditions.

(E) d-prime of edge FMG in V1.

(F) FGM d-prime in V4.

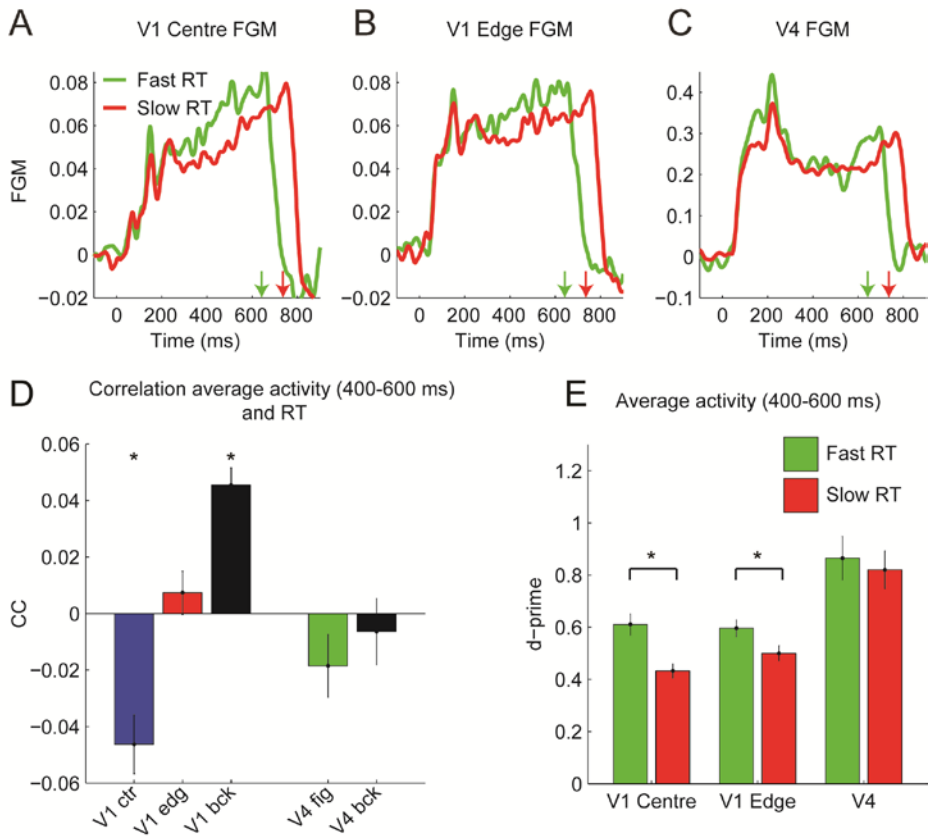

Figure S5. FGM in V1 and V4 for fast and slow responses (related to Figure 7).
(A-C) FGM in V1 at the figure centre (A) and edge (B) and V4 (C) for trials with reaction faster and slower than 100 ms. Arrows indicate the average onset of the saccade for the fast (green) and slow (red) trials.
(D) Average correlation between activity in V1 and V4 recording sites and reaction time. Error bars indicate ±1 SEM. Asterisks, significant correlations (Wilcoxon signed rank test, $P<0.05$).
(E) D-prime of FGM for trials with reaction faster (green bars) and slower than 100 ms (red bars) in a time window of 400 to 600 ms after stimulus onset. Error bars indicate ±1 SEM. Asterisks denote significant differences (paired t-test, $P<0.05$).

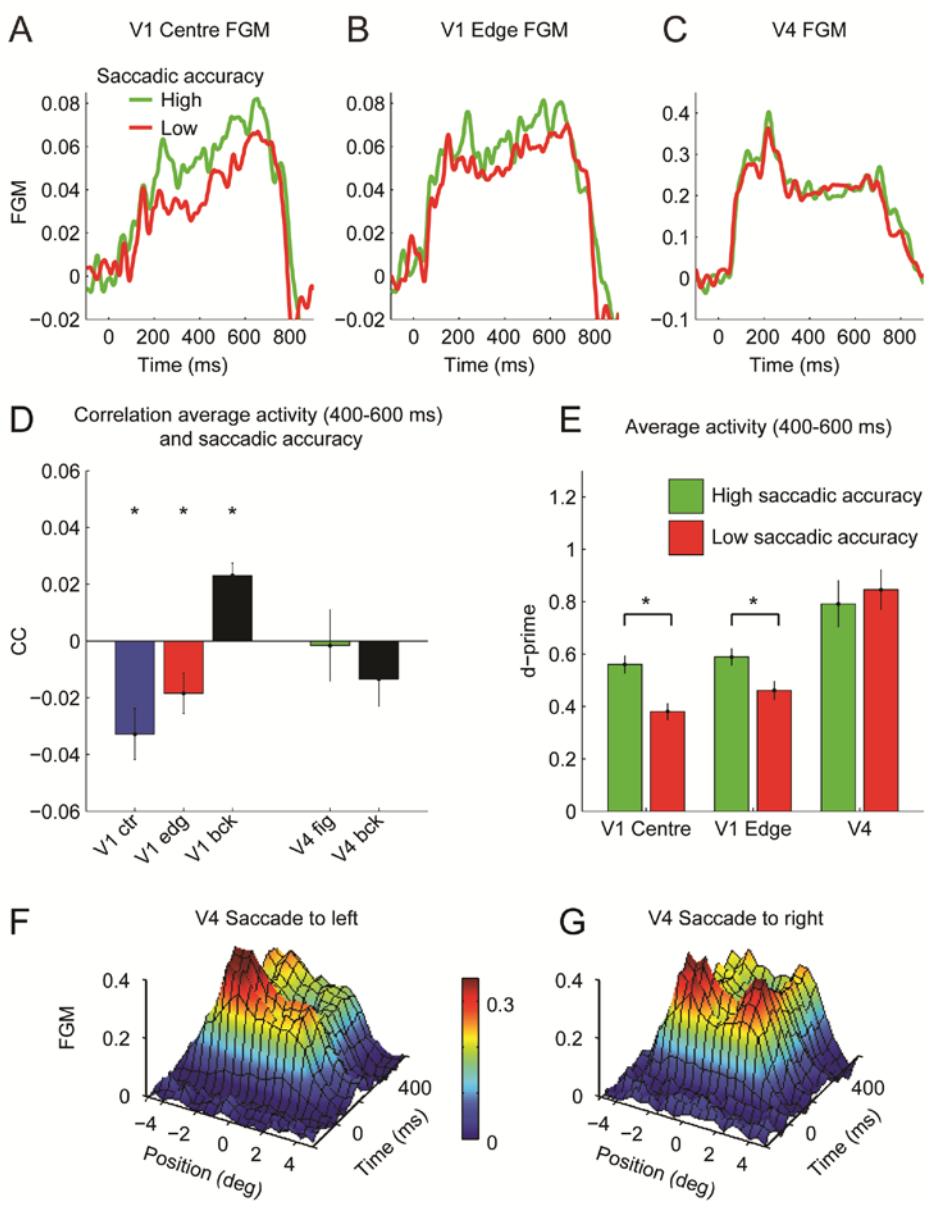

Figure S6. FGM in V1 and V4 in trials with high and low accuracy saccades (related to Figure 7).
(A-C) FGM in V1 at the figure centre (A) and edge (B) and V4 (C) for trials with high (25% most accurate trials, green curve) and low saccadic accuracy (25% least accurate saccades, red curve).
(D) Average correlation between activity in V1 and V4 recording sites and reaction time. Error bars indicate ±1 SEM. Asterisks, significant correlations (Wilcoxon signed rank test, $P<0.05$).
(E) D-prime in V1 at the figure centre and edge and in V4 for trials with high (green bars) and low saccadic accuracy (red bars). Error bars indicate SEM. Significant differences are indicated by asterisks (paired t-test, $P<0.05$).
(F,G) Spatiotemporal profile of V4 FGM in the figure-detection task for trials with saccades deviating to the left (F), and saccades deviating to the right (G), averaged across 37 V4 recording sites. X-axis, time relative to saccade onset. Y-axis, figure position relative to the centre of the RF.

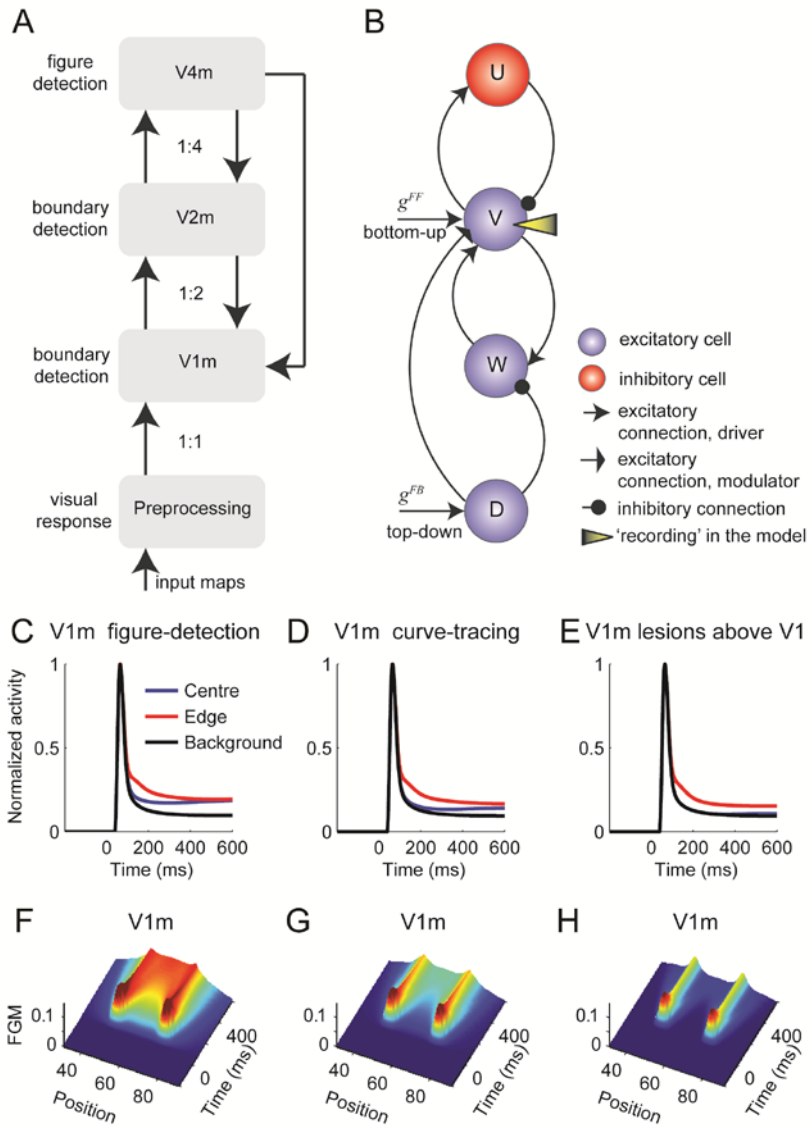

Figure S7. Schematic drawing of the model architecture and model activity in area $V1_m$ (related to Figure 9).

(A) Three model areas ($V1_m$, $V2_m$, $V4_m$) form a hierarchy of processing stages. The width (and height) of the RF of a unit in area V4 is four times the width of a V2 RF, which is in turn twice the width of a V1 RF. All areas connect to the next area in the hierarchy by feedforward connections. In addition, cortical areas provide feedback signals to lower areas. (B) Each area in panel A contains a population of four coupled cell types for every retinotopic position. Excitatory cells (blue circles) receive a driving bottom-up (*V*) or modulatory top-down input (*D*). The inhibitory cell (*U*, red circle) receives input from the excitatory cell and in turn inhibits it. Boundary-detection is realized by local competition provided by *W*. The figures display the activities of the *V* units. (C-E) Responses at the centre (blue curve), edge (red) and background positions (black), in the figure-detection task (C), the curve-tracing task (D), and in a lesion experiment (E). Note the difference in FGM at the centre and the edge. In the figure-detection task FGM at the centre becomes equal to edge-FGM, in the curve-tracing task the centre-FGM is weaker, and in the lesion experiment it largely disappears. These results demonstrate that centre-FGM in the model is generated by feedback, whereas edge-FGM depends on a local process. (F-H) Space-time profiles of the FGM in $V1_m$ in the figure-detection (F), curve-tracing (G), and lesion experiment (H).

Identifier	V1	V2	V4
C_1	1	1	1
C_2	1	1	5
C_3	5	5	5
C_4	15	15	-
E_1	5	3	5
E_2	1	5	2
E_3	35	35	15 / 5 (*)
E_4	2 / 0 (**)	2 / 0 (**)	2
E_5	2	2	-
G_1^2	45	45	25
G_2^2	1	1	1
G_3^2	2	2	1
G_4^2	1/15	1/15	1
G_5	1	1	1
G_6^2	1	1	1
G_7	1	1	-
A	5	5	5
K	2	2	-

$\sigma_{FB}\ (pixel)$	1	1	-
$suport\ \Lambda_{\sigma FB}\ (pixel)$	5	5	5
$\sigma_{FF}\ (pixel)$	-	1	1
$suport\ \Lambda_{\sigma FF}\ (pixel)$	-	5	5
$\sigma+(pixel)$	1	1	1
$suport\ \Lambda_{\sigma+}\ (pixel)$	5	5	5
$\sigma-(pixel)$	2	2	2
$suport\ \Lambda_{\sigma-}\ (pixel)$	9	9	9
Area specific identifiers and FF/FB connectivity between areas			
V	V_{V1}	V_{V2}	V_{V4}
$g^{FF}(t)$	$V_{\Pr e}$	V_{V1}	V_{V2}
$g^{FB}(t)$	$V_{V2}+V_{V4}$	V_{V2}	-

Table S1. Model parameters

(*) In the figure-detection task E_3 was set to 15 and it was set to 5 in the curve-tracing task. (**) To simulate the lesion experiment, E_4 was set to zero. See below for the specification of model parameters.

Supplemental Experimental Procedures

1. Implantation technique and electrode positions

Figure S1A shows a picture that was taken during the array implantation surgery in monkey B. In order to access the primary visual cortex (V1) and V4 a bone flap was removed over one hemisphere, and the dura was opened. Arrays of 4x5 electrodes (Cyberkinetics Neurotechnology Systems Inc.) were implanted in the cortex. In monkey B, 5 arrays were implanted in V1 and 3 in V4. Then the dura was sutured and the bone flap was placed back and fixed to the rest of the skull with titanium strips and screws. The position of the electrodes in V1 and V4 and the location of the RFs in the three monkeys have been illustrated in Figure S1B and S1C.

2. Recording of neuronal activity

We recorded multi-unit spiking activity (MUA) from the chronically implanted electrode arrays in areas V1 and V4 with TDT (Tucker Davis Technologies) recording equipment. As in previous studies (Legatt et al., 1980;Logothetis et al., 2001;Supèr and Roelfsema, 2005;Xing et al., 2009), the MUA signals from the electrodes were amplified, band-pass filtered (500-5000 Hz), full-wave rectified and then low-pass filtered at 500 Hz and sampled at a rate of 763 Hz. The MUA represents the pooled activity of a number of single units in the vicinity of the tip of the electrode and the population response obtained with this method is therefore expected to be identical to the population response obtained by pooling across single units. A recent study demonstrated that the MUA signal indeed provides a reliable estimate of the average single-unit response (Supèr and Roelfsema, 2005).

3. Mapping of receptive fields in V1 and V4

We measured the receptive field dimensions of every V1 recording site by determining the onset and offset of the response to a slowly moving light bar for each of eight movement directions (Kato et al., 1978). The median V1 receptive field area was 1.1 deg^2 (range 0.12 to 5.8 deg^2). The median eccentricity was 3.7 deg (range 1.6 to 10.9 deg). In V4 the MUA RFs were mapped by presenting white dots (0.5 deg, luminance 82$cd \cdot m^{-2}$) on a grey background (luminance 14$cd \cdot m^{-2}$) at different positions of a grid (0.5 deg spacing). We defined the hotspot of the V4 RF as the position with the maximum response and the RF borders as the locations where activity fell below 50% of the maximum (Motter, 2009). With this definition, the median V4 RF area was 9.6 deg^2 (range 3.3 to 96 deg^2). Setting the borders at 25% instead of

50% of the maximum resulted in a median RF area of 25 deg^2. The hotspots of the V4 RFs had a median eccentricity of 3.2 deg (range 1.0 to 18.6 deg).

4. Visual stimulus

The figure-ground stimulus consisted of a square figure with oriented line elements (16 pixels long, 0.44°, and 1 pixel wide) on a background with an orthogonal orientation (Figure 2A). To construct this figure-ground stimulus, we first made two base-textures. One base-texture was composed of line elements with an orientation of 45° and the other of line elements with an orientation of 135°. A base texture was made by first creating a full screen grid with a horizontal and vertical spacing of 5 pixels. The oriented line elements were placed on the grid with a random offset added to both the horizontal and vertical position (1 to 20 pixels). We then made one texture by copying a square 4°x4° region of the 45° base-texture onto the 135° base-texture and a complimentary texture by copying the same square region of the 135° base-texture onto the 45° base-texture. We averaged neuronal responses evoked by the complimentary stimuli, thereby ensuring that the RF was stimulated on average by the same set of local features, regardless of whether it fell on the figure or background. In the special case when the RF was on the figure edge, stimulation of the RF in the two complimentary conditions is balanced if the RF behaves like a linear spatial filter, linearly summing up responses to the line elements (Lamme et al., 1999). Although the conditions (edge or no edge in the RF) are in this case pixel-wise balanced, this balancing is not possible for higher order statistics including the change in orientation.

5. Time-course of FGM

Figure S2 shows the average time-course of FGM at the figure centre and edge in V1 and in V4 in the two tasks. When we aligned the data on the saccade, we found that the FGM ramped up (Figure 4C). To further quantify the time-course of FGM we measured d-prime in an early (300-450ms after stimulus onset) and a later window (450-600ms) after stimulus onset, and also in two windows aligned on the saccade (300-150 ms and 150-0ms before saccade onset) (Figure S2E,F). In the figure-detection task, FGM in V1 increased over time for both the centre ($d'_{300\ to\ 450}$=0.37 vs. $d'_{450\ to\ 600}$=0.47, $P<10^{-6}$) and edge positions ($d'_{300\ to\ 450}$=0.46 vs $d'_{450\ to\ 600}$=0.52, $P<10^{-6}$), but it did not increase in the curve tracing task (centre $d'_{300\ to\ 450}$=0.25 vs $d'_{450\ to\ 600}$=0.26, P=0.26, edge $d'_{300\ to\ 450}$=0.41 vs. $d'_{450\ to\ 600}$=0.41, P=0.89). When we aligned the V1 response to the onset of the saccade we found the same effect: in the figure detection task the d-prime task increased for both the centre and edge positions (centre

$d'_{-300 \text{ to } -150}$=0.43 vs $d'_{-150 \text{ to } 0}$=0.53, $P<10^{-6}$, edge $d'_{-300 \text{ to } -150}$=0.51 vs $d'_{-150 \text{ to } 0}$=0.55, $P<10^{-4}$). In the curve tracing task the d-prime stayed constant at the edge but there was a small increase with time at the centre (centre $d'_{-300 \text{ to } -150}$=0.25 vs $d'_{-150 \text{ to } 0}$=0.27, $P<0.05$, edge $d'_{-300 \text{ to } -150}$=0.41 vs $d'_{-150 \text{ to } 0}$=0.41, P=0.5).

In area V4 FGM stayed constant in the figure detection task and decreased in time in the curve tracing task (FD $d'_{300 \text{ to } 450}$=0.79 vs $d'_{450 \text{ to } 600}$=0.80, P=0.5 , CT $d'_{300 \text{ to } 450}$=0.50 vs $d'_{450 \text{ to } 600}$=0.45, $P<0.01$). When we aligned the responses to the saccade, we found that the d-prime increased in the figure detection task and remained constant in the curve tracing task (FD $d'_{-300 \text{ to } -150}$=0.77 vs $d'_{-150 \text{ to } 0}$=0.83, P=0.001, CT $d'_{-300 \text{ to } -150}$=0.45 vs $d'_{-150 \text{ to } 0}$=0.46, P=0.5).

To investigate if the attentional effect on FGM is stronger around the time of the saccade, we computed the increase in d-prime $(d'_{FD}-d'_{CT})/ d'_{CT}$ in an early (300 -150ms before the saccade) and a late time window (150-0ms before the saccade). For the V1 centre FGM, the attentional effect indeed increased from 74% in the early time window to 98% in the later window. For the V1 edge modulation, the attentional effect also increased from 25% to 33%, and we observed the same effect in V4 where the attentional effect increased from 71% to 81%. In summary, V1 FGM increased with time and the attentional effect was strongest just before the saccade. In V4, FGM also increased just before the saccade and the V4 attentional effect was slightly stronger before the saccade, just as in V1.

6. Eye position controls

The eye position was measured with a double magnetic induction technique and sampled at a rate of 1017 Hz (Bour et al., 1984). Small differences in eye position between the conditions could influence neuronal activity, especially in area V1 where RFs are small. To investigate the potential effect of eye position, we first carried out a simplified analysis of the interaction between the effects of attention and texture segregation in two time windows, and we then repeated this analysis while controlling for eye position.

In the simplified analysis we defined two time windows. The first window included the initial response (50 to 150 ms after stimulus onset) and the second window the sustained response (200 to 600 ms). We performed a repeated measures ANOVA with factors RF position (centre or edge) and task (figure-detection or curve-tracing) on the FGM in these time windows across recording sites. In the initial response window in V1, we replicated the main effect of position (repeated measures ANOVA $F_{1,58}$=77.31, $P<10^{-6}$) because the response to the figure edge was stronger than the response to the centre, and found again no main effect of attention ($F_{1,58}$=1.6, $P>0.2$), and no interaction ($F_{1,58}$=0.03, $P>0.4$). In the

sustained response window in V1, we found significant main effects of position (repeated measures ANOVA $F_{1,58}$=37, $P<10^{-6}$) and task ($F_{1,58}$=46, $P<10^{-6}$) and also a significant interaction between position and task ($F_{1,58}$=52, $P<10^{-6}$). In area V4, FGM in the initial window did not depend on the task (paired t-test $P>0.4$), but FGM was stronger in the figure-detection task than in the curve-tracing task in the sustained response window (paired t-test $P<10^{-6}$).

We then performed a stratification procedure to investigate the effect of eye position that was first described in Roelfsema et al. (1998, see Figure 2 therein), and used in several later studies (e.g. Khayat et al., 2006;Moro et al., 2010;Roelfsema et al., 2007;Roelfsema and Spekreijse, 2001) to investigate if systematic differences in eye position contribute to the differences in response magnitude.

We first computed the average horizontal and vertical eye-position during the visual stimulation period (0-600 ms after stimulus onset) when the monkey maintained fixation. We then rejected trials with microsaccades (small eye movements within the fixation window) so that only trials with a stationary eye position remained. We divided the fixation window (1°×1°) in 4x4 bins of 0.25°x0.25° and assigned every trial to one of these bins on the basis of the average eye-position. This resulted in a 2D distribution of fixation positions in the vicinity of the fixation point for every stimulus condition (RF on figure centre, the edge or on the background). We made these distributions identical (stratification) by randomly removing surplus trials until the number of trials in each bin was the same in all conditions. We repeated our analysis in two time windows with the reduced data set.

In the initial window we found a main effect of RF position (repeated measures ANOVA $F_{1,58}$=71, $P<10^{-6}$), no main effect of attention ($F_{1,58}$=0.01, $P>0.4$), and no interaction ($F_{1,58}$=0.03, $P>0.4$). In the sustained response window we found a main effect of RF position (repeated measures ANOVA $F_{1,58}$=40, $P<10^{-6}$) and attention ($F_{1,58}$=28, $P<10^{-6}$) and also a significant interaction between RF position and attention ($F_{1,58}$=36, $P<10^{-6}$), replicating the effects that were observed without stratification. Also in area V4 stratification did not change the results. There was no effect of task in the initial window (paired *t*-test $P>0.4$), but significantly stronger FGM in the figure-detection task than in the curve-tracing task in the sustained period ($P<10^{-6}$).

In an additional control analysis we divided trials in 3 groups with different ranges of eye positions in the direction in which the figure was shifted across conditions (we rotated data in monkey C by 90 deg because we varied figure position in the vertical direction). The first group contained the 25% trials with most leftward fixation (median fixation position -

0.16), the second group the 50% trials with central fixation (median fixation position 0) and the third group the 25% trials with the most rightward fixation (median fixation position 0.18). Figure S3 shows that our results are replicated in every trial group. We then performed the simplified analysis described above, separately for these three groups of trials.

In the initial response window (50 to 150 ms after stimulus onset) in V1 we obtained a main effect of RF position: a difference in the response evoked by figure and background ($F_{1,58}$=41, 79, 57, all *Ps*<10^{-6}), and no effect of attention ($F_{1,58}$=0.09; 3.8; 0.02, all *Ps*>0.05), and no interaction ($F_{1,58}$=0.26, 0.92, 0.17, all *Ps*>0.05). In the sustained response window (200 to 600 ms after stimulus onset) we found significant main effects of RF position ($F_{1,58}$=26, 35, 33, all *Ps*<10^{-6}) and attention ($F_{1,58}$=19.6, 43, 30, all *Ps*<10^{-4}) in V1, and also a significant interaction between RF position and attention ($F_{1,58}$=28, 26, 28, all *Ps*<10^{-5}). In area V4, FGM in the initial window did not depend on attention (paired t-test all *Ps*>0.05), but FGM was stronger in the figure-detection task than in the curve-tracing task in the sustained response window (paired t-test all *Ps*<10^{-5}). All results were the same in all three groups and replicated the results of our main analysis. We therefore conclude that our results cannot be explained by differences in eye position between conditions.

7. Effect of difficulty of the curve tracing task on FGM and attentional modulation

The curve-tracing task demands attention (Scholte et al., 2001), and a recent study in human observers (Houtkamp and Roelfsema, 2010) demonstrated that subjects cannot trace more than one curve at a time, even if they are placed in opposite hemifields. The monkeys accuracy was high (94%) in our standard curve tracing task, and we therefore considered the possibility that we might have seen larger differences in FGM had we used a more demanding curve-tracing task. We note however that we did observe a clear difference in FGM between tasks *in spite of* this high accuracy.

We carried out an additional experiment with the same texture stimuli but we now varied the difficulty of the curve tracing-task. The monkeys performed the curve-tracing task or the texture-segregation task on alternating days, just as in our main experiment. To make the curve-tracing task more difficult, we introduced intersections between the curves on 50% of the trials and made it difficult to discriminate between intersections and non-intersections (Figure S4). If the curves did not cross, we varied the size of the gap between them (1.3, 0.8 or 0.5deg) because a small gap causes the configuration to look more like an intersection. If the curves crossed, we varied the angle of the intersection (30, 60 or 90 deg) because sharper angles look more like non-intersections (see Roelfsema and Spekreijse, 2001). On curve-

tracing days, accuracy indeed decreased for smaller gaps. Accuracy was 94% with the largest gap and decreased to 91% for the intermediate gap and to 80% for the smallest gap (Figure S4B,C, repeated measures ANOVA, $F_{2,10}$=10.95, P<0.001). For crossing curves, the angle of the intersection did not result in reliable differences in accuracy (unlike in Roelfsema and Spekreijse, 2001), but the accuracy in all conditions with intersections was lower than in the easiest condition with non-crossing curves. Thus, the intersections made the curve-tracing task substantially more difficult.

To obtain a sufficient number of trials we restricted the measurement of FGM to 5 positions of the RF relative to the figure (our main experiment used 23 positions). In one condition we placed the figure center on the RF, in two conditions the RF fell on the edge, and in two conditions the RF fell on the background (see Figure S4A). We recorded from 25 recording sites in V1 (monkey J, n=10, Monkey C, n=15), and from 24 recording sites in V4 (monkey C).

We first investigated how FGM depended on the task, pooling activity across all configurations of the two curves (blue and red bars in Figure S4D,E,F). We reproduced all findings of our main experiment. First, centre-FGM in area V1 was stronger in the figure-detection task (average d'=0.46) than in the curve-tracing task (average d'=0.31, paired t-test $P<10^{-5}$). Second, the effect of top-down attention on V1 FGM was weaker if the RF fell on the edge because the average d' was 0.43 in the figure-detection task and 0.36 in curve-tracing task. Third, we observed a significant interaction between task and RF position in V1 (edge or centre, repeated measures ANOVA, $F_{1,24}$=20.54, P<0.001). Fourth, FGM in V4 in the figure-detection task (average d'=1.51) was weaker than in the curve-tracing task (average d'=0.88, $P<10^{-5}$).

We next investigated the effect of the difficulty of the curve-tracing task by comparing FGM d' across the 3 conditions with non-intersections (Figure S4B) where smaller gap-sizes caused more errors and may have captured more attentional resources.

- *Effect of difficulty on centre FGM in V1.* The effect of gap size in the curve-tracing task on V1 FGM d-prime in the figure centre was small (d' gap3=0.35, d' gap2=0.29 and d' gap1=0.33). The d' for the largest gap (gap 3) was higher than the d' for the middle gap (gap 2, P<0.01), but the other comparisons did not yield significant differences (gap 3 vs. 1, P>0.4; gap 2 vs. 1, P>0.1).

- *Effect of difficulty on edge FGM in V1.* Variations in V1 edge FGM d' were small (d'gap3=0.41, d'gap2=0.35 and d'gap1=0.38) and the only significant difference was between

gap 2 and 3 (P<0.01), whereas FGM did not differ significantly between the other difficulty levels (gap 3 vs. 1, P>0.2; gap 2 vs. 1, P>0.1).

- *Effect of difficulty on FGM in V4.* In V4, FGM d' was larger for the smaller gap sizes (d' gap 3=0.75, d' gap 2=0.85 and d' gap 1=0.93). FGM for gap 3 was smaller than that for gap 2 and 1, but the difference in FGM between gap 1 and 2 was not significant (gap 2 vs. 3, P<0.05: gap 1 vs. 3, P<0.001; gap 1 vs. 2, P>0.11).

Thus, we have reproduced all of our main findings with a more demanding version of the curve-tracing task. We found comparatively small effects of task-difficulty on FGM in V1 and V4. In V4, there was even a small increase in FGM if the curve-tracing task became more difficult. We therefore conclude that our main results do not depend strongly on the difficulty of the curve-tracing task.

8. Functional role of FGM in V1 and V4

We predicted that if FGM plays a role in saccade planning (1) FGM should predict the timing of the saccade, (2) it should distinguish between saccades with higher and lower accuracy and (3) the spatial profile of FGM should predict the saccadic landing point

(1) FGM predicts the timing of the saccade

Figure 7A (also shown as Figure S5A) demonstrated that V1 center FGM predicts the timing of the saccade. We found a similar albeit weaker effect of saccade timing on the V1 edge modulation (Figure S5B), but this effect was not apparent in V4 (Figure S5C). To investigate the relationship between the activity of individual recording sites (in a time window from 400 to 600 ms after stimulus onset) and reaction time, we computed correlation coefficients. We found a significant *negative* correlation between the activity at the figure centre in V1 and the reaction time, and a *positive* correlation between V1 background responses and the reaction time (Figure S5D). The correlation between the reaction time and V1 edge responses was not significant and the same held true for the correlation with V4 activity evoked by either figure or background. We also found no significant correlations between V1 activity and reaction time in the curve-tracing task. Thus, in the figure-detection task, short response times are associated with a strong V1 response to the figure centre, and a weak response to the background.

To determine the significance of the earlier onset of the ramping in fast and slow trials, we computed the d-prime in a window of 400 to 600 ms (Figure S5E). The d-prime of the V1

centre responses was significantly higher on trials with fast reaction times (d' fast =0.61, d' slow =0.43, paired t-test, $P<10^{-6}$) and a similar effect was found for the V1 edge FGM (d' fast=0.60, d' slow =0.50, $P<10^{-5}$), but not in V4 (d' fast =0.87,d' slow=0.82, P=0.17). The effect of FGM on RT was more pronounced for the V1 centre FGM than for the edge FGM (Δd' V1 centre=0.18, Δd' V1 edge=0.10, P<0.01).

(2) FGM distinguishes between saccades with higher and lower accuracy

We measured saccadic accuracy as the Euclidean distance between the saccadic endpoint on a particular trial and the median saccadic endpoint in that condition and divided trials into those with a low (25% trials with largest distance from median saccadic endpoint) and high saccadic accuracy (25% trials with smallest distance).

V1 FGM was stronger on trials with saccades of a high accuracy than on trials with a lower accuracy saccade (Figure S6). This effect was strongest for the V1 representation of the figure centre, weaker for the representation of the edge and absent in area V4. To test significance we computed d-primes in a window from 400 to 600 ms after stimulus onset (Figure S6E). The d-prime of centre and edge FGM was significantly higher in trials with accurate saccades (V1 centre: d' accurate =0.56, d' less accurate=0.38, $P<10^{-6}$, V1 edge: d' accurate=0.59, d' less accurate=0.46, $P<10^{-5}$). In addition, the effect of accuracy on centre FGM was stronger than the effect of accuracy on edge FGM (Δd' V1 centre=0.18, Δd' V1 edge=0.12, P<0.05). The V4 FGM d' did not depend on the accuracy of the saccade (d' high RT=0.82, d' low RT=0.86, P=0.35).

To investigate the relation between neuronal activity at individual recording sites and the offset of the saccade vector from its median, we computed the correlation (Figure S6D). If the V1 RF fell on the figure centre or on the edge, the correlation between the response and saccade offset (a measure for inaccuracy) was significant and negative. If the RF fell on the background, however, the correlation was positive. Thus, accurate saccades are associated with strong responses at the figure centre and edge and weak responses in the background. In V4 these correlations did not differ significantly from zero.

(3) The spatial profile of FGM predicts the saccadic landing point

The relation between the saccadic landing point and neuronal activity in V1 has been illustrated in Figure 7. If the saccade deviated to the left, V1 FGM was higher on the left side of the figure and if the saccade deviated to the right FGM was strong on the right side of the figure. We observed the same effect in V4 where an increased FGM on the left predicted that

the saccade would deviate to the left, and an increased FGM on the right predicted a deviation of the saccade to the right (Figure S6F,G).

To test significance, we first selected stimulus conditions with the RF either on the left or right side of the figure (*posL*: -2, -1.5, -1 deg; *posR*, 1, 1.5, 2 deg), and computed FGM in a time window of 400 to 600 ms, separately for *sacL,* and *sacR* trials. We first determined if the bias to the left or right could be predicted from the FGM and we found that V1 FGM was stronger if the saccade deviated towards the side of the figure on which the RF was located ($FGM_{posR,sacR} > FGM_{posR,sacL}$, paired t-test $P<0.01$ and $FGM_{posL,sacR} < FGM_{posL,sacL}$, $P<0.05$). As a measure of the overall strength of these effects, we computed a modulation index:

$$MI_{LR} = \frac{(FGM_{posR,sacR} - FGM_{posR,sacL})}{(FGM_{posR,sacR} + FGM_{posR,sacL})/2} + \frac{(FGM_{posL,sacL} - FGM_{posL,sacR})}{(FGM_{posL,sacL} + FGM_{posL,sacR})/2}$$

The average value of this index was 0.44 in V1, which is significantly larger than zero (Wilcoxon signed rank test, $P<10^{-5}$). Also in V4 the FGM was greatest when the saccade deviation was towards the side on which the RF was located (both *Ps*<0.05) and MI_{LR} was 0.10, which was significantly greater than zero ($P<0.05$).

9. Latency of the visual responses, FGM and attentional modulation

There are different methods to measure to latency of a neuronal response and the latency of a response difference. One common method is to take the first of a number of time bins that satisfy a significance criterion (Lennie, 1981;Maunsell and Gibson, 1992). The disadvantage of this method is that it is sensitive to the amount of data that was collected because more trials yield higher significances and therefore shorter latencies. To avoid this confound, we used an alternative latency estimate that is independent of the number of trials, and is derived by fitting a function *f(t)* to the response or response difference (Roelfsema et al., 2003;Thompson et al., 1996). The function was derived from the following two assumptions: firstly, the onset of the response (or the response difference) has a Gaussian distribution across trials and neurons, and secondly a fraction of the response dissipates exponentially. The assumptions yield two differential equations:

$$\partial m_1(t)/\partial(t) = -\alpha m_1(t) + g(t,\mu,\sigma) \quad (1)$$

$$\partial m_2(t)/\partial(t) = g(t,\mu,\sigma) \quad (2)$$

Equation (1) models the dissipating response (modulation) and equation (2) the non-dissipating response (modulation), where $g(t,\mu,\sigma)$ is a Gaussian density with mean μ and standard deviation σ, a^{-1} is the time constant of the dissipation. The total response (modulation) is taken as the sum of m_1 and m_2. The solution to these equation is the sum of an ex-Gaussian and a cumulative Gaussian (Luce, 1986):

$$f(t) = d(\exp(\mu\alpha + 0.5\sigma^2\alpha^2 - \alpha t)(G(t,\mu+\sigma^2\alpha,\sigma) + cG(t,\mu,\sigma),$$

where c and d are the contributions of the non-dissipating and dissipating activity (modulation). We fitted $f(t)$ to the responses using the Matlab curve fitting toolbox. In our main analysis we (arbitrarily) defined the latency as the point in time where the fitted function reached 33% of its maximum (lat_{33}).

Our population analysis fitted the function to the response (modulation) averaged across recording sites. To determine the confidence intervals of the latency estimates we performed a bootstrapping procedure: we randomly selected recording sites from our sample with replacement 1000 times and then fitted the function to estimate the 95% confidence interval (2.5 and 97.5 percentiles). To determine the significance of a latency difference between conditions we subtracted the bootstrapped latency estimates to obtain a distribution of the latency difference (two-sided p-value). We also determined the latencies at individual recording sites. The latency of FGM could only be determined for recording sites with significant FGM in the sustained time window (200-600 ms) in the figure-detection task ($p<0.05$, 64% of V1 recording sites [38 of 59] and 91% in V4 recording sites [42 of 46]).

10. Model of cortico-cortical interactions for figure-ground segregation

The aim of our model was to investigate if (1) an early and local boundary-detection process based on iso-orientation inhibition could be combined with a later process that implements iso-orientation excitation for region filling in feedback connections, and (2) if such a model can account for neuronal activity in areas V1 and V4. Our model uses mechanisms that are similar to a previous model (Roelfsema et al., 2002) and that are also in line with a more recent model by Bhatt et al. (2007), who used the term 'attentional shroud' for the filling of the interior of the figure with enhanced neuronal activity. Such an attentional shroud presumably corresponds to object-based attention that is directed to all the image elements of the figure.

The activity of many neurons in areas V1, V2 and V4 evoked by an edge depends on whether this edge is 'owned' by the region on one or on the other side of the edge (Zhou et al., 2000). These border ownership signals give a complementary approach to figure-ground segregation. Recent studies (Craft et al., 2007;Mihalas et al., 2011) modelled the mechanisms for border ownership as the interaction between lower and higher visual areas. We did not include border-ownership signals in our model, which rather focused on the mechanisms for region filling. However, future modelling studies could explore more elaborate schemes with interactions between border-ownership, attention and region filling for the optimal combination of region-based and edge-based cues for figure-ground segregation.

Stimulus input for the model

We used a visual image of 121×121 input elements (corresponding to 19°×19°, i.e. 1° of visual angle corresponds to 6 image elements). The input consisted of a figure (25×25 elements, 4°×4°) with 45° texture elements on a background with an orientation of 135°. The input is represented by model units in two separate activity maps, one for the 45° and the other for the 135° orientation. Input units have an activity of one if the texture orientation matches their preferred orientation. We added a delay of 40ms at the preprocessing stage to account for delays prior to $V1_m$[i].

Model overview and model dynamics

The model consists of cortical areas $V1_m$, $V2_m$, and $V4_m$, each containing a complete spatial representation of the visual image. Within each area, there are $N \times N \times 2$ units where N is the

[i] Model areas are indexed by 'm' to distinguish them from anatomical areas in visual cortex.

number of RFs in the X- and Y directions (V1, *N*=121; V2, *N*=61; V4, *N*=15) and there are two preferred orientations. The region of space over which units in the model integrate activity increases so that the width (and height) of the RF of a $V4_m$ unit is four times the width of the RF of a $V2_m$ unit, and the width of a RF of a $V2_m$ unit is twice the width of a RF in $V1_m$. Therefore the figure is processed at a high spatial resolution in $V1_m$, and at coarser resolutions in $V2_m$ and $V4_m$. Neurons in area $V1_m$ are bi-directionally connected to neurons in $V2_m$ and $V2_m$ in turn is bi-directionally coupled with $V4_m$. In addition, area $V4_m$ sends direct feedback connections to $V1_m$ (Figure S7A).

All model cells are simulated as single compartment units with an activity that is modeled as a continuous variable (rate coding). The model units receive excitatory and inhibitory input, and have a leak conductance. Changes in activity are induced by excitatory and inhibitory conductances, similar to the Hodgkin-Huxley model (Hodgkin and Huxley, 1952) and the activity of the units is determined by a set of first-order nonlinear ordinary differential equations, which is numerically solved with a second-order Runge-Kutta method (using a step size that corresponds to 1 ms). The units of the model receive bottom-up input from the previous area, top-down input from higher areas and they also interact with neighboring units of the same area.

Preprocessing of the visual input

The input from the LGN to the cortex consists of a strong transient when the stimulus appears, and this transient is followed by a weaker sustained response. We modeled these dynamics at the preprocessing stage by coupling fast excitatory units with slower inhibitory units with activities V_{Pre} and W_{Pre}, respectively. The excitatory preprocessing unit receives a bottom-up signal g^{in}, it then excites the inhibitory unit and receives feedback inhibition in return. The inhibitory unit has a time-constant that is five times larger than that of the excitatory unit. As a result, V_{Pre} builds up activation within about 50ms and it is then gradually inhibited by W_{Pre}. The dynamics of the preprocessing stage are governed by the following equations:

$$1 \cdot \dot{V}_{Pre} = -2 \cdot {W_{Pre}}^2 \cdot V_{Pre} + g^{in}(t) \cdot (10 - V_{Pre}) \quad (3a)$$

$$5 \cdot \dot{W}_{Pre} = -W_{Pre} + V_{Pre} \cdot (25 - W_{Pre}) \quad (3b)$$

In equation (3a) the excitatory unit integrates the input signal g^{in} (towards a maximum of 10) and its activity decays back to zero due to a leak conductance that depends quadratically on

the activity of the inhibitory unit, W_{Pre}. The inhibitory unit integrates the input from the excitatory unit towards a maximum of 25 (Equation 3b). The interaction between excitatory and inhibitory units cause a strong transient response followed by a weaker sustained response, just as is observed in the LGN and in visual cortex.

Model areas V1, V2, and V4: Connectivity scheme

Every area receives driving input from the next lower area (V1 from the preprocessing stage) and it also receives modulatory input from higher areas (Bullier, 2003) (Figure S7B). The areas contain four types of units arranged in a cortical column that is repeated across retinotopic positions. The first unit (*V* in Figure S7B) is an input unit and receives activity from the lower level. The second unit (*D*) receives the feedback input from higher areas, the third (*U*) is a slow inhibitory unit and the fourth unit (*W*) is responsible for the centre-surround interactions that give rise to boundary detection. The region in the previous layer that provides input to *V*-units is controlled by a Gaussian kernel $\Lambda_{\sigma,FF}$. Top-down signals from a region in the higher areas (determined by a Gaussian kernel $\Lambda_{\sigma,FB}$) are fed back to *D*-units to modulate the activity[ii].

Model areas V1, V2, and V4: Model dynamics

We implemented a local iso-orientation inhibition scheme for boundary detection in the feedforward pathway (using the *W*-units) and an interareal region filling scheme based on iso-orientation excitation in the feedback pathway (mediated by the *D*-units). The units filter the input that comes from the previous layer, they engage in local competitive interactions, and are modulated by feedback. These interactions were modeled as a system of four coupled, nonlinear ordinary differential equations (Figure S7). A set of three equations models the *forward pathway* while a single equation describes the *feedback pathway*.

The first Equation (4a) of the forward pathway determines the membrane potential *V* of *excitatory cells*. The model neurons integrate the bottom-up signal g^{FF} from the previous area and are modulated by feedback from higher areas, through the *D*-unit,

$$C_1 \cdot \dot{V} = -G_1^2 \cdot U \cdot V + A \cdot g^{FF}(t) \cdot (1 + K \cdot D) \cdot (E_1 - V) + G_2^2 \cdot W \cdot (E_2 - V). \qquad (4a)$$

[ii] These equations are the same for V1, V2, and V4.

Thus, three input conductances influence *V*. The first term ($G_1^2 \cdot U \cdot V$) denotes a leak conductance, which is influenced by the inhibitory *U*-cell (see Equation (4c) below). The second term ($A \cdot g^{FF} \cdot (1 + K \cdot D) \cdot (E_1 - V)$), determines the driving input g^{FF} and its modulation by the top-down feedback signal *D*. Note that the *D*-unit can only influence activated *V*-units and has no effect if the driving input ($g^{FF}(t)$) is zero. The feedback is therefore *modulatory*: it scales the activity but it cannot create activity in units not driven by feedforward connections (see Raudies and Neumann, 2010). The third term, $G_2^2 \cdot W \cdot (E_2 - V)$ determines the activation *W* from the boundary detection process (Equation (4b)), which causes the response enhancement at edges.

Boundaries between regions with different features are detected by a local competitive mechanism within the orientation maps (i.e. iso-orientation inhibition) defined by

$$C_2 \cdot \dot{W} = -G_3^2 \cdot D \cdot W + G_4^2 \cdot (\left| V - (V * \Lambda_{\sigma-}) \right| * \Lambda_{\sigma+}) \cdot (E_3 - W). \quad (4b)$$

The term $\left| V - (V * \Lambda_{\sigma-}) \right| * \Lambda_{\sigma+}$ represents an on-centre/off-surround filter that attains high or low values near boundaries. The inhibitory surround $\Lambda_{\sigma-}$ is a Gaussian kernel and * denotes a convolution. The mechanism is sensitive to boundaries in both feature maps because we compute the absolute value. Responses on both sides of the boundary are merged by the second convolution with the Gaussian kernel $\Lambda_{\sigma+}$. This boundary detection signal excites the *W*-unit through the term $G_4^2 \cdot (\left| V - (V * \Lambda_{\sigma-}) \right| * \Lambda_{\sigma+}) \cdot (E_3 - W)$. The leak conductance of the *W*-unit is controlled by the *D*-unit of the feedback pathway (through the term $G_3^2 \cdot D \cdot W$).

The activity of the inhibitory *U*-cell is given by

$$C_3 \cdot \dot{U} = -G_5 \cdot U + G_6^2 \cdot V \cdot (E_4 - U). \quad (4c)$$

The first term $-G_5 \cdot U$ denotes a passive leak conductance and the second term, $G_6^2 \cdot V \cdot (E_4 - U)$, represents the input from the excitatory *V*-cell.

Finally, the top-down modulation for region filling is mediated by the *D*-unit that receives excitation from units in higher areas tuned to the same orientation (i.e. iso-orientation excitation). The activity of *D*-units is determined by

$$C_4 \cdot \dot{D} = -G_7 \cdot D + g^{FB}(t) \cdot (E_5 - D) \qquad (4d)$$

The time constant C_4 is relatively large causing slower changes in activation than the effects of the driving forward pathway (Roelfsema et al., 2002) (Table S1). The first term ($-G_7 \cdot D$) denotes a leak conductance and the second term ($g^{FB}(t) \cdot (E_5 - D)$) determines the effect of the feedback input g^{FB} that can reach a maximal activity E_5. Feedback connections excite neurons tuned to the same orientation to fill the entire figural region in lower areas with FGM. The values of all relevant constants are listed in Table S1.

Modeling the effect of attention and the effect of a lesion of areas higher than V1

To model the effect of attention we changed a single parameter, the reversal potential of the boundary detection mechanisms in $V4_m$ (E_3, Table S1). A decrease in E_3 reduced the pop-out effect in area $V4_m$ (Figure 9E,F) and this reduces the feedback to area $V1_m$. The result of the weaker feedback signal from $V4_m$ (and $V2_m$ that is also influenced by $V4_m$) to area $V1_m$ is a reduction of the centre-FGM (Figure 9C,D). Note, however, that the reduced feedback has comparatively little effect on the representation of the figure boundaries that are detected by a mechanism intrinsic to area $V1_m$.

To investigate the role of feedback in the model, we carried out an additional experiment where we modeled the effect of a lesion of areas higher than $V1_m$. We abolished the effect of feedback by setting E_4 in Equation (4d) to zero. Even in the complete absence of feedback boundary-detection in area $V1_m$ was still present but now the centre FGM was completely absent (Figures S7E and S7H).

Reference List

Bhatt,R., Carpenter,G.A., and Grossberg,S. (2007). Texture segregation by visual cortex: perceptual grouping, attention, and learning. Vision Res. *47*, 3173-3211.

Bour,L.J., van Gisbergen,J.A., Bruijns,J., and Ottes,F.P. (1984). The double magnetic induction method for measuring eye movement--results in monkey and man. IEEE Trans. Biomed. Eng *31*, 419-427.

Bullier,J. (2003). Hierarchies of Cortical Areas. In The Primate Visual System, J.H. Kaas, and C.E. Collins, eds. CRC Press), pp. 181-204.

Craft,E., Schütze,H., Niebur,E., and von der Heydt,R. (2007). A neural model of figure-ground organization. J. Neurophysiol. *97*, 4310-4326.

Hodgkin,A.L., and Huxley,A.F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol *117*, 500-544.

Houtkamp,R., and Roelfsema,P.R. (2010). Parallel and serial grouping of image elements in visual perception. J. Exp. Psychol. Hum. Percept. Perform. *36*, 1443-1459.

Kato,H., Bishop,P.O., and Orban,G.A. (1978). Hypercomplex and simple/complex cell classifications in cat striate cortex. J. Neurophysiol. *41*, 1071-1095.

Khayat,P.S., Spekreijse,H., and Roelfsema,P.R. (2006). Attention lights up new object representations before the old ones fade away. J. Neurosci. *26*, 138-142.

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

Legatt,A.D., Arezzo,J., and Vaughan,H.G., Jr. (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.

Lennie,P. (1981). The physiological basis of variations in visual latency. Vision Res. *21*, 815-824.

Logothetis,N.K., Pauls,J., Augath,M., Trinath,T., and Oeltermann,A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412*, 150-157.

Luce,R.D. (1986). Response times (Oxford: Oxford University Press).

Maunsell,J.H., and Gibson,J.R. (1992). Visual response latencies in striate cortex of the macaque monkey. J. Neurophysiol. *68*, 1332-1344.

Mihalas,S., Dong,Y., von der Heydt,R., and Niebur,E. (2011). Mechanisms of perceptual organization provide auto-zoom and auto-localization for attention to objects. Proc. Natl. Acad. Sci. U. S. A *108*, 7583-7588.

Moro,S.I., Tolboom,M., Khayat,P.S., and Roelfsema,P.R. (2010). Neuronal activity in the visual cortex reveals the temporal order of cognitive operations. J. Neurosci. *30*, 16293-16303.

Motter,B.C. (2009). Central V4 receptive fields are scaled by the V1 cortical magnification and correspond to a constant-sized sampling of the V1 surface. J. Neurosci. *29*, 5749-5757.

Raudies,F., and Neumann,H. (2010). A neural model of the temporal dynamics of figure-ground segregation in motion perception. Neural Netw. *23*, 160-176.

Roelfsema,P.R., Khayat,P.S., and Spekreijse,H. (2003). Subtask sequencing in the primary visual cortex. Proc. Natl. Acad. Sci. U. S. A *100*, 5467-5472.

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

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

Roelfsema,P.R., and Spekreijse,H. (2001). The representation of erroneously perceived stimuli in the primary visual cortex. Neuron *31*, 853-863.

Roelfsema,P.R., Tolboom,M., and Khayat,P.S. (2007). Different Processing Phases for Features, Figures, and Selective Attention in the Primary Visual Cortex. Neuron *56*, 785-792.

Scholte,H.S., Spekreijse,H., and Roelfsema,P.R. (2001). The spatial profile of visual attention in mental curve tracing. Vision Res. *41*, 2569-2580.

Supèr,H., and Roelfsema,P.R. (2005). Chronic multiunit recordings in behaving animals: advantages and limitations. Prog. Brain Res. *147*, 263-282.

Thompson,K.G., Hanes,D.P., Bichot,N.P., and Schall,J.D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. J. Neurophysiol. *76*, 4040-4055.

Xing,D., Yeh,C.I., and Shapley,R.M. (2009). Spatial spread of the local field potential and its laminar variation in visual cortex. J. Neurosci. *29*, 11540-11549.

Zhou,H., Friedman,H.S., and von der Heydt,R. (2000). Coding of border ownership in monkey visual cortex. J. Neurosci. *20*, 6594-6611.