

Temporal Dynamics of Figure-Ground Segregation in Human Vision

Peter Neri and Dennis M. Levi

J Neurophysiol 97:951-957, 2007. First published Sep 27, 2006; doi:10.1152/jn.00753.2006

You might find this additional information useful...

This article cites 30 articles, 11 of which you can access free at:

<http://jn.physiology.org/cgi/content/full/97/1/951#BIBL>

Updated information and services including high-resolution figures, can be found at:

<http://jn.physiology.org/cgi/content/full/97/1/951>

Additional material and information about *Journal of Neurophysiology* can be found at:

<http://www.the-aps.org/publications/jn>

This information is current as of January 17, 2007 .

Temporal Dynamics of Figure-Ground Segregation in Human Vision

Peter Neri and Dennis M. Levi

School of Optometry, University of California at Berkeley, Berkeley, California

Submitted 20 July 2006; accepted in final form 26 September 2006

Neri P, Levi DM. Temporal dynamics of figure-ground segregation in human vision. *J Neurophysiol* 97: 951–957, 2007. First published September 27, 2006; doi:10.1152/jn.00753.2006. The segregation of figure from ground is arguably one of the most fundamental operations in human vision. Neural signals reflecting this operation appear in cortex as early as 50 ms and as late as 300 ms after presentation of a visual stimulus, but it is not known when these signals are used by the brain to construct the percepts of figure and ground. We used psychophysical reverse correlation to identify the temporal window for figure-ground signals in human perception and found it to lie within the range of 100–160 ms. Figure enhancement within this narrow temporal window was transient rather than sustained as may be expected from measurements in single neurons. These psychophysical results prompt and guide further electrophysiological studies.

INTRODUCTION

Image interpretation relies on the ability to extract and isolate objects from their surroundings. At the most basic and fundamental level of image analysis, objects consist of figures and their surroundings make up the background (Marr 1982). The segregation of figure from background cannot be achieved by analyzing only small image regions like those corresponding to the classical receptive fields (RFs) of neurons in primary visual cortex (V1). The fragmented information obtained by local analysis must be integrated at a more global level, either by horizontal connectivity within the same cortical stage (Zhaoping 2005) or by later processing at subsequent stages (Hupé et al. 1998; Lamme et al. 1998).

In line with these considerations, the response of V1 neurons does not depend exclusively on the image region presented within their RFs (Albright and Stoner 2002). Even if this region is left unchanged, V1/V2 neurons respond more effectively when the region is part of a larger figure as opposed to a background (Lamme 1995), even though information about the presence of the figure may only be available $20 \times$ RF-size away from the neuron's RF (Qiu and von der Heydt 2005; Zhou et al. 2000). This selectivity appears after ~ 100 ms of stimulus presentation (Zipser et al. 1996). When the RF is on the edge of the figure, the effect is typically faster at ~ 50 ms (Lamme et al. 1998; Zhou et al. 2000). Interestingly, figure-ground specific signals are abolished in the presence of anesthesia, suggesting that their origin may be feedback dependent (Lamme et al. 1998).

Figure-ground segregation is involved in texture processing. Brain signals specific to texture segregation can be measured at the surface of the human scalp using electroencephalographic (EEG) technology and become available after ~ 160 – 200 ms of stimulus presentation (Bach and Meigen 1999; Lamme et al.

1992). If disruptive signals are delivered to the scalp using transcranial magnetic stimulation (TMS), the largest effect on figure-ground segregation is observed within two periods at ~ 160 and ~ 260 ms (Heinen et al. 2005). Activity related to these phenomena has also been measured using functional MRI (fMRI) (Kastner et al. 2000; Skiera et al. 2000), but the time scale of texture segregation far exceeds the intrinsically poor temporal resolution of this technique.

Neural findings from the electrophysiological literature are summarized by the notion that figure-ground specific signals appear in visual cortex at various stages within a large time window spanning 50–300 ms. The presence of these signals in itself does not imply that they are also used for perceptual purposes (Vul and MacLeod 2006). Of course, figure and ground are eventually represented as behavioral constructs, but relevant neural signals may be used for this purpose at any point within the 50- to 300-ms time span during which they are known to be available. Our psychophysical experiments showed that figure and ground are assigned different perceptual gains by human observers within a narrow temporal window between 100 and 160 ms after stimulus presentation.

METHODS

Stimuli and tasks

Each trial could be of either figure-ground type (Fig. 1, *A* and *B*) or neutral-edge type (Fig. 1, *C* and *D*). Both trial types consisted of two 200-ms intervals separated by an 800-ms gap.

FIGURE-GROUND TRIALS. Figure-ground stimuli (Fig. 1, *A* and *B*) consisted of a $6^\circ \times 6^\circ$ square presented at an eccentricity (distance between fixation and stimulus center) of 4.3° and could appear at any of four locations around fixation (Fig. 1*C*, dashed outlines). This square contained a relevant region in the middle ($0.86^\circ \times 0.86^\circ$; Fig. 1*B*, orange outline). The portion of the square outside the relevant region defined a figure-ground configuration (Fig. 1, *A* and *B*). We used the relevant region to probe figure-ground perception. Specifically, the relevant region contained a spatiotemporal signal (an edge described below) across the figure-ground border. This probe could be either vertical (Fig. 1*A*) or horizontal (Fig. 1*B*), and the surrounding stimulus could be of either polarity (bright figure/dark background or vice versa). For example, in Fig. 1*B*, the figure ($1.72^\circ \times 1.72^\circ$) is bright ($+18.4$ cd/m²) and the background is dark (-18.4 cd/m²), but the opposite polarity (Fig. 1*A*, dark figure on bright background) could appear with equal probability. All luminance values are defined in relation to monitor background luminance (52 cd/m²).

FIGURE-GROUND TASK. Our task was a two-interval forced choice. One interval contained a target stimulus; the other interval contained a nontarget stimulus. Observers had to identify the target interval and received immediate feedback. Whether an interval is target or non-

Address for reprint requests and other correspondence: P. Neri, Department of Optometry and Visual Science, City University, Northampton Square, London EC1V 0HB, United Kingdom (E-mail: pn@white.stanford.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

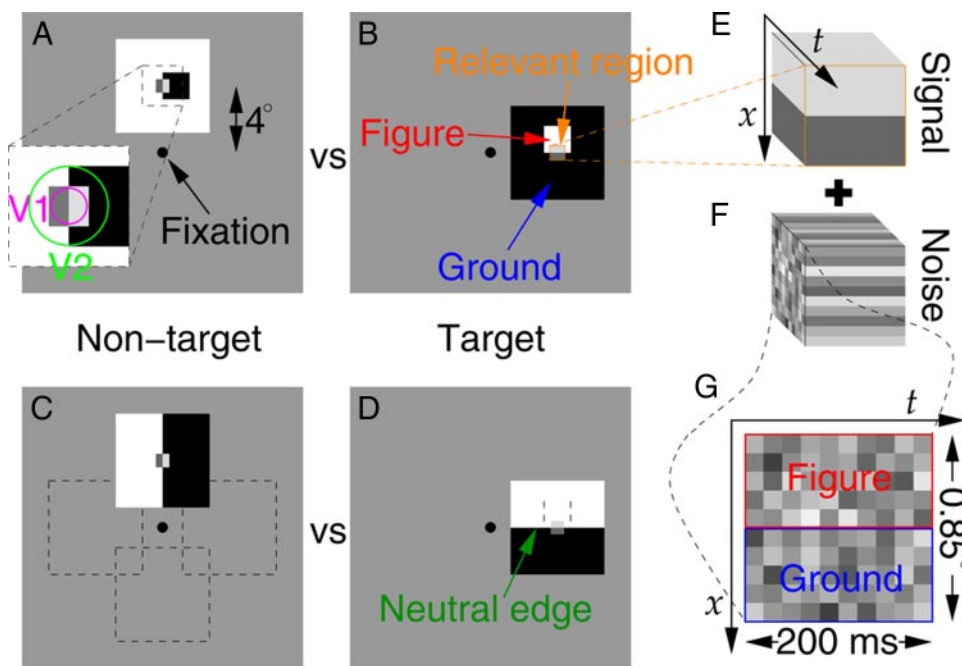


FIG. 1. Stimuli and task. *A* and *B*: stimulus consisted of a large ground square containing a smaller figure square (labeled in *B*) and could appear at any of 4 locations (dashed outlines in *C*) around fixation. Figure contained a central relevant region (orange outline). Each trial consisted of 2 intervals: in the nontarget interval, polarity of the edge within the relevant region was inconsistent with polarity of remaining stimulus (*A*); in the target interval, it was consistent (*B*). Each trial could be of either figure-ground type (*A* and *B*) or neutral-edge type (*C* and *D*). Size of relevant region was comparable with estimated receptive field (RF) size in V1/V2 (magnifying inset in *A*). Edge presented within the relevant region is the signal that defines whether interval is target or nontarget and can be thought of as the spatiotemporal object in *E*. In every interval, we added to this signal a different sample of spatiotemporal noise (*F*). Because both *E* and *F* do not vary across the spatial dimension that runs along the edge, *F* can be represented as the 2-dimensional object shown in *G*.

target depends on the relationship between the relevant region and the rest of the stimulus. As noted above, the relevant region contains a spatiotemporal signal $S(x, t)$ like that shown in Fig. 1*E*, divided across space so that one side is bright and the other side is dark. If we take $x = 0$ to be at the level of the edge in Fig. 1*E*, we can describe this signal as $S = k$ for $x < 0$ and $S = -k$ for $x > 0$, where x ranges between -0.46 and $+0.46^\circ$, t goes from 0 to 200 ms, and $k = 1.84$ (subjects S2 and S3) or 2.3 (S1 and S4) cd/m^2 (the contrast of the edge within the relevant region was lower than the contrast of the rest of the stimulus as visible in Fig. 1*B*). This signal did not vary across time. In the target interval, this signal was consistent with the stimulus outside the relevant region like in Fig. 1*B* (i.e., the bright side within the relevant region was on the bright side of the rest of the stimulus), whereas in the nontarget interval, it took the opposite polarity (Fig. 1*A*). In each interval on every trial, we added to the signal an independently generated sample of spatiotemporal Gaussian noise $n_i(x, t)$ with SD $\sigma_n = 6.9 \text{ cd/m}^2$ varying every 20 ms across 10 different spatial positions spanning the relevant region (Fig. 1, *F* and *G*).

NEUTRAL-EDGE TRIALS. As a control, we also presented neutral-edge trials (Fig. 1, *C* and *D*), which were interleaved with the figure-ground trials. Neutral-edge stimuli are shown in Fig. 1, *C* and *D*. Note that in the relevant region, the neutral-edge is identical to the figure-ground stimulus. The neutral edge could be either vertical (Fig. 1*C*) or horizontal (Fig. 1*D*) and could be of either polarity.

NEUTRAL-EDGE TASK. The neutral-edge task was identical to and was randomly interleaved with the figure-ground task. The observers' task was to identify the target interval and was followed by immediate feedback. In the target interval, this signal was consistent with the stimulus outside the relevant region as in Fig. 1*D* (i.e., the bright side within the relevant region was on the bright side of the rest of the stimulus), whereas in the nontarget interval, it took the opposite polarity (Fig. 1*C*). As with the figure-ground task, we added to the signal an independently generated sample of spatiotemporal Gaussian noise. Both trial types and all stimulus conditions were randomly interleaved. We tested four observers. All but S2 (P.N.) were naïve. All procedures received NE1 Human Subjects approval.

Reverse correlation analysis

The actual stimulus that was presented to our observers could contain the figure above, below, to the left, or the right of the

relevant region. For analysis, we rotated the stimulus so that the figure (whether bright or dark) was above as shown in Fig. 1*B*. For neutral-edge trials, the stimulus was similarly rotated by 90° when the edge was vertical to make it horizontal. We refer to this configuration (horizontal edge, figure above) in the following two paragraphs and in Fig. 2. The stimulus presented within the relevant region on the target interval on trial i is $I_i^T(x, t) = S(x, t) + n_i^T(x, t)$. Similarly, in the nontarget interval, $I_i^{NT} = -S + n_i^{NT}$. $\Delta n_i^c = p \cdot n_i^T - q \cdot n_i^{NT}$ is the difference between the noise sample presented in the two intervals for that trial. $c = 0$ if the observer responded incorrectly on that trial, 1 if correctly. $p = 1$ if the edge defined by the entire stimulus on that interval (whether figure-ground or neutral-edge) is bright above and dark below (as shown in Fig. 1, *B* and *D*), and $p = -1$ for the opposite polarity (Fig. 1, *A* and *C*). The same holds for q . The sign inversion introduced by p and q allows us to combine noise fields from stimuli of opposite polarity by effectively co-registering polarity across stimuli. The perceptual filter F is computed $F(x, t) = E[\Delta n_i^1] - E[\Delta n_i^0]$, where E is the mean across i (Abbey and Eckstein 2002). Because the initial rotation we perform for analysis is ambiguous in the case of neutral-edge trials (i.e., a vertical edge can be made horizontal by rotating it either clockwise or anticlockwise), the neutral-edge perceptual filter is inherently symmetric across space (Fig. 2*C*). In other words, once polarity is factored out (which is necessary to combine data from all trials), there is no logical distinction between one side and the other side of the neutral edge.

We constructed the temporal profiles shown in Fig. 2*E* by pooling data only from the spatial location in the filters that showed the largest modulation across conditions, shown by colored outlines in Fig. 2, *C* and *D*. This spatially restricted analysis was necessary to isolate informative regions within filters from uninformative noise. The symmetry of the neutral-edge filter means that we can only plot one curve for this condition. To plot gain for the background (blue trace), we reversed the sign for filter modulations within this region, because background polarity is always opposite to the figure.

Modeling

In each interval, each spatial position x in the image stimulus $I(x, t)$ (Fig. 3*A*) was convolved (\otimes) with the temporal impulse response function $f(t)$ shown in Fig. 3*B* [$f(t) = 0$ for $0 \leq t < 40$ ms, $f(t) = 0.103$

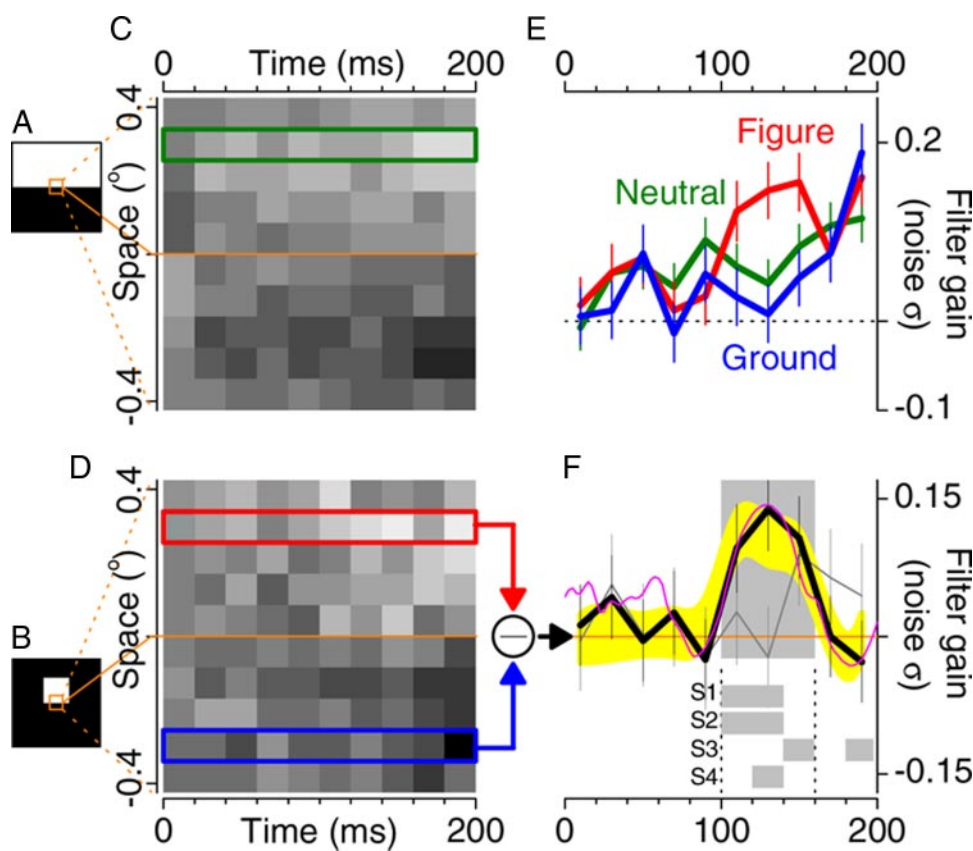


FIG. 2. Figure-ground enhancement occurs within 100–160 ms. Stimulus cartoons are shown in *A* and *B* for neutral-edge and figure-ground, respectively. These should not be interpreted to mean that data in this figure only refer to the specific configurations depicted here, in that we pooled data from all contrast polarities and orientations. *C*: perceptual filter for neutral-edge trials from 4 subjects. By analysis, the neutral-edge filter is symmetric around the edge. *D*: perceptual filter for figure-ground trials. *E*: filter values for temporal slices indicated by colored outlines in *C* and *D* (ordinate plots filter amplitude in units of noise SD). Values for the ground (blue) outline have been sign-reversed to plot gain. Only 1 trace (green) is plotted for the neutral-edge filter because of its inherent symmetry. *F*: black line plots difference between red and blue traces in *E* (all traces were first computed separately for each subject and then averaged across subjects). Gray box indicates temporal region for which this difference is significantly different from 0 at $P < 0.05$. Small boxes refer to individual observers (S1–S4). Yellow trace shows outcome of model simulations (Fig. 3). Gray trace refers to a control experiment where the figure was replaced by 2 lines at positions indicated by dashed lines in Fig. 1*D* (7,000 trials collected between S1 and S2). Pink trace is taken from blue trace (orbitofrontal cortex) in Fig. 3*C* of Bar et al. (2006), arbitrarily rescaled along the ordinate. We collected a total of ~20,000 trials (5,000 per subject on average). Raw data are shown, no smoothing was applied. Error bars show \pm SE for data and \pm SD for model simulation (yellow trace).

for $40 \leq t < 60$ ms, $f(t) = 0.021$ for $60 \leq t < 80$ ms, $f(t) = 0.003$ for $80 \leq t < 100$ ms, and $f(t) = 0$ for $t < 0$ or $t \geq 100$ ms], obtaining $o(x,t) = I(x,t) \otimes f(t)$. This front-end filtering stage simulates a V1 onset latency of 40 ms (Bair et al. 2002). We computed $o_{\text{gnd}}(t) = -E_{x>0}[o(x,t)]$ and $o_{\text{fig}}(t) = E_{x<0}[o]$ (Fig. 3*C*), where the average E is intended across x for $x > 0$ (within the background *gnd*) or $x < 0$ (within the figure *fig*). For the modeling results shown here, we only averaged two spatial positions within each region to simulate the partial spatial weighting observed in the data, but this choice was not critical. The final scalar output for the interval was $O = E_i[o_{\text{fig}}(t) \times W_{\text{fig}}(t) + o_{\text{gnd}}(t) \times W_{\text{gnd}}(t)]$, where $W_{\text{fig}}(t)$ and $W_{\text{gnd}}(t)$ are, respectively, the red and blue weighting functions in Fig. 3*E*. Internal noise (Gaussian distribution with SD $0.04 \times \sigma_n$) was added to O . The interval with largest O was selected by the model as the target interval.

We used $k = 2.1$ cd/m² (average of the 2 values used with human observers). We computed model responses to 20,000 trials of stimuli identical to those used in the psychophysics and derived the corresponding perceptive fields. We repeated this procedure 100 times. Model trace in Fig. 2*F* plots the mean \pm SD for the 100 estimates.

Ideal observer analysis

The ideal filter F_{ideal} matches the signal (Green and Swets 1966), i.e., $F_{\text{ideal}}(x,t) = g$ for $x < 0$ and $F_{\text{ideal}} = -g$ for $x > 0$, where $g > 0$. Its output on every trial is $O_{\text{ideal}}^T = E_{x,t}[F_{\text{ideal}}(x,t) \cdot I^T(x,t)]$ for the target interval and similarly for the nontarget interval (substitute I^{NT} for I^T). The ideal observer chooses the interval associated with the largest output, so the probability of a correct response is $P[(O_{\text{ideal}}^T - O_{\text{ideal}}^{\text{NT}}) > 0]$. The quantity within $()$ is $E_{x,t}[F_{\text{ideal}} \cdot (2S + n_i^T + n_i^{\text{NT}})]$, which is proportional to a Gaussian distribution with mean μ and SD $\sigma_n/\sqrt{200}$. For the lowest signal-to-noise ratio we used in the psychophysical experiment ($\mu = 0.27\sigma_n$), this SD is 0.26μ , i.e., $p \sim 1$ (100% correct). Because the ideal observer never makes errors, the formula for deriving F using psychophysical reverse correlation cannot be applied. However, we can make a prediction by assuming a lower SNR. In this case, the prediction is simply $F \propto F_{\text{ideal}} \propto S$, i.e., the filter is a scaled image of the signal (Ahumada 2002). We are assuming here that the ideal observer has perfect knowledge of the location, orientation, and polarity of the edge, because these parameters were unambiguously determined by the noise-free high-contrast stimulus area around the relevant region. Even if some uncertainty is assumed about these parameters, the ideal observer would suffer from it equally for figure-ground and neutral-edge trials, predicting no difference between them.

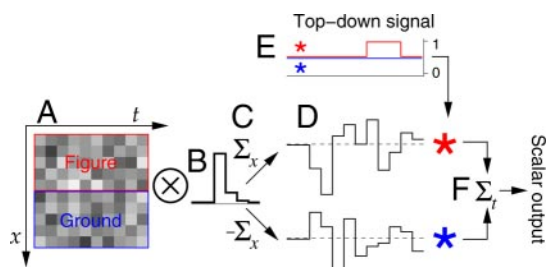


FIG. 3. Model structure. Spatiotemporal stimuli (*A*) were convolved (\otimes) with a temporal impulse response (trace in *B*), which simulated a neuronal onset latency of 40 ms. Spatial averaging was performed separately within figure and ground (*C*). Model weighed ($*$) 2 outputs from this temporal convolution (*D*) using the 2 weighting functions shown by red and blue traces in *E*. Final output associated with each stimulus was computed by averaging across time (*F*) the difference between figure and ground, and interval with largest output was selected by the model. Resulting perceptual filter is shown in Fig. 2*F*, plotted on the same axis scale used for human data.

RESULTS

Figure enhancement effect

The rationale for our experiments was the following. It is known that stimuli containing figure-ground configurations lead to enhanced firing of V1 neurons with RFs at the level of the figure (Zipser et al. 1996). If this neuronal enhancement is associated with an increased perceptual gain for the stimulus features that are processed by V1 neurons, we expect to see a measurable change in the perceptual filter used by observers to process the figure-ground target. To test this hypothesis, we performed experiments aimed at deriving figure-ground perceptual fields, the psychophysical analogs of neuronal RFs (Neri and Levi 2006).

We designed a task that required observers to compare the contrast polarity of a probe (relevant region in Fig. 1*B*) with the polarity of the surrounding stimulus, whether the surround defined a figure-ground configuration (Fig. 1, *A* and *B*) or a neutral edge (Fig. 1, *C* and *D*). To measure the perceptual filter used by observers for performing this task, we used a perturbation technique in which visual noise is added to the signal edge within the relevant region (Ahumada 2002; see METHODS). Figure 2 shows the two spatiotemporal filters obtained for neutral-edge trials (Fig. 2*C*) and figure-ground trials (Fig. 2*D*). Figure 2*E* shows temporal profiles along the spatial positions indicated by colored outlines in Fig. 2, *C* and *D*, where filter modulation was largest across space. It appears that, between 100 and 160 ms, the temporal profile for the figure lies above the profile for the ground (red > blue). To bring out this effect more clearly, we computed the difference between figure and ground, shown by the black line in Fig. 2*F*. Between 100 and 160 ms, observers were weighing the figure side of the stimulus more than the ground side (the large gray box shows significance at $P < 0.05$ averaged across observers; the small gray boxes show that each of the 4 observers has significant regions within this temporal window).

The figure > ground effect could occur as a consequence of figure enhancement or background suppression. Because we independently measured a perceptual filter for the neutral condition (green), we are in a position to compare figure versus neutral-edge on one side (Fig. 2*E*, red vs. green), and background versus neutral-edge on the other side (Fig. 2*E*, blue vs. green). It seems from these comparisons that the figure > ground effect results mainly from figure enhancement (red > green). However, we cannot rule out the possibility that we may have failed to detect a smaller effect of background suppression. Only one trace is shown for the neutral-edge condition because this is a necessary logical consequence of combining data from all trials (i.e., from all polarity configurations). This procedure results in inherent symmetry across space for the neutral-edge filter (see METHODS).

Modeling

To illustrate more clearly our interpretation of the effect shown in Fig. 2*F*, we constructed a straightforward model that implements the physiology of figure-ground segregation as currently understood. This model incorporates a top-down enhancement of figure gain that lasts for 60 ms starting at 140 ms after stimulus onset (Fig. 3). As shown in Fig. 2*F* (yellow trace), the model prediction closely matches that derived from

human observers (no rescaling was applied to the yellow trace).

We chose to model the top-down enhancement based on firing responses from neurons with RFs inside the figure (Zipser et al. 1996) rather than at the edge between figure and ground (Zhou et al. 2000). The difference in the onset of figure enhancement between these two conditions is ~ 50 ms (100–120 ms as opposed to 50–70 ms). The bulk of modulations within our spatiotemporal filters occurred at the spatial locations indicated by colored outlines in Fig. 2, *C* and *D*. These spatial locations are closer to the inside of the figure than they are to the edge. For this reason, the enhancement latency measured for neurons with RFs within the figure seemed more relevant to our analysis.

We emphasize that the model in Fig. 3 is not intended as a literal and detailed description of the neural processing stages occurring within the brain of the human observers. The purpose of this model is to show that the main result of a larger modulation in the reverse correlation image for figure as opposed to ground is generally consistent with expectations from what we presently know about neuronal responses relating to figure-ground segregation. For example, the exact time scale and characteristics of the impulse response in Fig. 3*B* or the feedback signal in Fig. 3*E* are not to be intended to reflect the direct implications of our data. Rather, the facts established by our experiments are the empirical traces in Fig. 2, *E* and *F*. Models with very different characteristics may be consistent with these traces. Our simple model shows that the traces conform to rough predictions from the physiological literature.

Performance metrics

Figure 4 plots the percentage of correct identifications of the target interval on neutral-edge trials (*x*-axis) and on figure-ground trials (*y*-axis) for all four observers. All values were

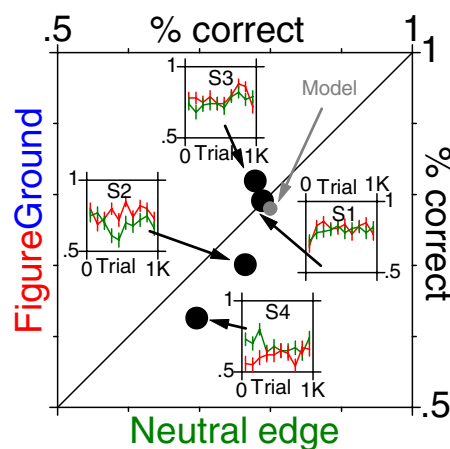


FIG. 4. Detection performance did not differ between neutral-edge and figure-ground (paired t -test, $P > 0.05$). Black points plot percentage of correct responses given by observers on figure-ground trials (*y*-axis) vs. neutral-edge trials (*x*-axis). Each point refers to a different observer. Human performance falls between chance (0.5) and perfect (1) for signal-to-noise ratio used in our experiments and does not differ between the 2 trial types. *Insets*: percent correct for the 1st 1,000 trials in each observer binned every 100 trials on both figure-ground (red) and neutral-edge trials (green). Observers learned the task very quickly (performance does not change systematically over time). Gray point shows model performance.

between 60 and 85% and fall very close to the unity line, indicating that there was no substantial difference in overall performance between the two trial types. The model (gray symbol) agrees nicely with the human data.

Control experiment

We worried that the effects reported thus far may have resulted as a consequence of the asymmetry introduced by the figure-ground stimulus (Fig. 1, *A* and *B*), rather than by the presence of the figure. The gray trace in Fig. 2*F* shows the result of a control experiment for which the stimulus was like the one we used on neutral-edge trials (Fig. 1, *C* and *D*) but with the addition of two clearly visible lines (black-on-white or white-on-black, same contrast as the rest of the stimulus) positioned at the locations occupied by the two sides of the figure in the figure-ground stimulus (Fig. 1*D*, dashed lines). No point is significantly different from 0. A related concern is that observers may have performed the task based on a simple contrast metric: in the vicinity of the relevant region, targets have lower contrast than nontargets because of the strong edges at the periphery of the relevant region in the nontarget stimulus. A strategy based on this distinction, however, predicts no difference between neutral-edge and figure-ground trials, contrary to our observations.

DISCUSSION

Relations to cortical physiology

The size of the relevant region in our stimuli was chosen to match the estimated RF size of V1/V2 neurons for the corresponding eccentricity (Gattass et al. 1981; see *inset* in Fig. 1*A*). Our task required observers to resolve the details of the edge within this region, and so we expect that they would need to rely on signals from early visual cortex, possibly as early as V1. The task also required that the polarity of the edge within the relevant region be compared with the polarity of the surrounding region, which roughly matched the estimated RF size of V2 neurons. It seems safe to conclude that observers relied at least partly on signals from V1/V2, because RF size is too coarse for this task in subsequent visual areas (Smith et al. 2001). However our stimuli were also designed to study the possibility that processing at larger scales may be involved. For the portion of stimulus contained within the two RFs in Fig. 1*A*, there is no difference between neutral-edge trials and figure-ground trials. Because we observed important differences, we speculate a role for information from larger regions.

We chose an overall duration of 200 ms for our stimuli to avoid eye movements and because it is comparable to the range typically spanned by the temporal impulse response of V1/V2 neurons (DeAngelis et al. 1995). Finally, 50–200 ms has been proposed as a plausible range for the psychological moment, the temporal scale below which no conscious scrutiny is possible (Von der Malsburg 1999). The effects we show in this study are within this range, lending strength to our speculation that they reflect physiological mechanisms underlying figure-ground segregation rather than higher-level thought processes.

Our result of enhanced gain for figure over ground is consistent with the properties of V1 neurons. Lamme (1995)

reported that neuronal gain is larger for neurons with RFs within a figure as opposed to background. If human observers rely on the output of several such neurons to perform the detection task used in this study, the filters returned by psychophysical reverse correlation are expected to show the figure > ground effect we observed. We verified this prediction using a simple model inspired by V1 physiology (Fig. 3). The difference in neuronal gain reported by these investigators appears after a delay of ~100 ms (Zipser et al. 1996), corresponding to the 100- to 160-ms range within which we observe the first significant sign of a figure > ground effect (Fig. 2*F*).

A puzzling feature of our data is that the figure > ground effect does not persist beyond 160 ms. V1 neurons show enhanced responses to the figure for the entire duration of the stimulus, typically up to 300 ms (Lamme 1995). Our stimuli lasted for 200 ms. From V1 physiology, we would expect that the black trace in Fig. 2*F* remains above zero up until the end of the 200-ms range we sampled. Interestingly, although the figure enhancement effect tends to remain more or less constant across time in area V1, it becomes less and less pronounced with time in area V4 (see Fig. 20 in Zhou et al. 2000). We therefore speculate that the transient nature observed in our data reflects processing at later stages in visual cortex. Indicative evidence supporting this speculation comes from magnetoencephalographic (MEG) recordings of top-down signals from orbitofrontal to visual cortex (Bar et al. 2006), which match very closely the transient figure > ground effect we report here (Fig. 2*F*, pink trace). Notice, however, that this comparison is not straightforward, in that the MEG top-down signal is more closely related to the gain enhancement shown by the red trace in Fig. 3*E* (rather than to the perceptual filter reported in Fig. 2*F*), and the two have slightly different time-courses because of the neuronal onset latency implemented by the temporal impulse response function in Fig. 3*B*.

Departures from the ideal detector

Our results are in general agreement with the properties of individual neurons in early visual cortex (V1–V2), but they depart markedly from the predictions of the ideal observer (Green and Swets 1966). The ideal prediction for Fig. 2*E* consists of a horizontal line at some positive value and is the same for all three colors. For Fig. 2*F*, the prediction is zero everywhere, and ideal performance is 100% correct in both trial types for Fig. 4 (see METHODS for technical specifications). Our data do not conform to these predictions, implying that human observers did not adopt an ideal strategy in performing the tasks despite receiving feedback on every trial. We speculate that the departures from optimality we measured psychophysically are inevitable consequences of the physiological mechanisms underlying the perceptual representations of figure and ground. Interestingly, this departure is restricted to the 100-to 160-ms window. Both before and after this temporal window, observers adopted a near optimal strategy (equal nonzero weights to figure and ground, Fig. 2*F*).

Despite having adopted a nonoptimal strategy on figure-ground trials, observers performed equally well on these trials as they did on neutral edge trials (Fig. 4). There is no inconsistency between these two results, as shown by our model,

which captures every aspect of them. However, it may be argued that the figure-ground effects exposed in Fig. 2 may not be of particular significance to figure-ground processing, given that they have no effect on detection performance. We believe that such a conclusion would be mistaken: it is not surprising that different metrics (reverse correlation averages as opposed to percentage of correct responses) may differ in their power to resolve specific behavioral phenomena. In our opinion, the failure of one should not diminish the value of results obtained from the other one.

The 100-ms timescale for perception

Various aspects of V1 responses that depend on information from outside the classical RF are characterized by an ~100-ms delay: line continuity (Roelfsema et al. 1998), target-distractor selectivity in multistimulus arrays (Constantinidis and Steinmetz 2001, 2005; Ipata et al. 2006), contour integration (Li et al. 2006), and a variety of electrophysiological studies of attentional selection (Spitzer et al. 1988; Supèr et al. 2001). The 100-ms offset is also relevant to several behavioral phenomena: exogenous cueing (Nothdurft 2002; Posner and Cohen 1984), attentional capture by high-contrast cues (Nakayama and Mackeben 1989), and detection versus identification of image segments (Neri and Heeger 2002). A recent study using a technique similar to ours has shown that contour interpolation occurs within a window of ~200 ms and peaks between 100 and 180 ms (Gold and Shubel 2006; although contour interpolation is related to figure-ground segmentation, this study did not address the figure-ground dichotomy as directly as we did here). The similarity in temporal dynamics for these different phenomena, including figure-ground segregation as shown here, suggests that they may share common neural mechanisms. Not all results obtained using psychophysical reverse correlation happen at this temporal scale: a recent study using this technique showed that center-surround interactions occur at significantly shorter scales, typically 0–50 ms (Tadin et al. 2006).

Our results call for more thorough explorations of the detailed time-course of figure-ground differences in the response of single neurons. Our prediction is that neurons in later visual areas will show increasingly transient figure-ground effects, such as those we measured in human observers (see also Bar et al. 2006) and suggested by the trend in Zhou et al. (2000). The transient nature of the figure-ground signal should play a central role in future electrophysiological studies, because it bears directly on perceptually relevant aspects of figure-ground processing in humans as shown in Fig. 2F.

ACKNOWLEDGMENTS

We thank M. Banks, S. Klein, R. Li, B. Olshausen, C. Schor, and J. Solomon for comments.

GRANTS

This research was supported by National Eye Institute Grant RO1 EY-01728.

REFERENCES

Abbey CK, Eckstein MP. Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. *J Vision* 2: 66–78, 2002.

- Ahumada AJ Jr. Classification image weights and internal noise level estimation. *J Vision* 2: 121–131, 2002.
- Albright TD, Stoner GR. Contextual influences on visual processing. *Annu Rev Neurosci* 25: 339–379, 2002.
- Bach M, Meigen T. Electrophysiological correlates of human texture segregation, an overview. *Doc Ophthalmol* 95: 335–347, 1999.
- Bair W, Cavanaugh JR, Smith MA, Movshon JA. The timing of response onset and offset in macaque visual neurons. *J Neurosci* 22: 3189–3205, 2002.
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hämäläinen MS, Marinkovic K, Schachter DL, Rosen BR, Halgren E. Top-down facilitation of visual recognition. *Proc Natl Acad Sci USA* 103: 449–454, 2006.
- Constantinidis C, Steinmetz MA. Neuronal responses in area 7a to multiple-stimulus displays: I. Neurons encode the location of the salient stimulus. *Cereb Cortex* 11: 581–591, 2001.
- Constantinidis C, Steinmetz MA. Posterior parietal cortex automatically encodes the location of salient stimuli. *J Neurosci* 11: 581–591, 2005.
- DeAngelis GC, Ohzawa I, Freeman RD. Receptive-field dynamics in the central visual pathways. *Trends Neurosci* 18: 451–458, 1995.
- Gattass R, Gross CG, Sandell JH. Visual topography of V2 in the macaque. *J Comp Neurol* 201: 519–539, 1981.
- Gold JM, Shubel E. The spatiotemporal properties of visual completion measured by response classification. *J Vision* 6: 356–365, 2006.
- Green DM, Swets JA. *Signal Detection Theory and Psychophysics*. New York: John Wiley, 1966.
- Heinen K, Jolij J, Lamme VAF. Figure-ground segregation requires two distinct periods of activity in V1: a transcranial magnetic stimulation study. *Neuroreport* 16: 1483–1487, 2005.
- Hupé JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394: 784–787, 1998.
- Ipata AE, Gee AL, Gottlieb J, Bisley JW, Goldberg ME. LIP responses to a popout stimulus are reduced if it is overtly ignored. *Nat Neurosci* 9: 1071–1076, 2006.
- Kastner S, De Weerd P, Ungerleider LG. Texture segregation in the human visual cortex: a functional MRI study. *J Neurophysiol* 83: 2453–2457, 2000.
- Lamme VAF. The neurophysiology of figure-ground segregation in primary visual cortex. *J Neurosci* 15: 1605–1615, 1995.
- Lamme VAF, van Dijk BW, Spekreijse H. Texture segregation is processed by primary visual cortex in man and monkey. Evidence from VEP experiments. *Vision Res* 32: 797–807, 1992.
- Lamme VAF, Zipser K, Spekreijse H. Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proc Natl Acad Sci USA* 95: 3263–3268, 1998.
- Li W, Piëch V, Gilbert CD. Contour saliency in primary visual cortex. *Neuron* 50: 951–962, 2006.
- Marr D. *Vision*. New York: Freeman, 1982.
- Nakayama K, Mackeben M. Sustained and transient components of focal visual attention. *Vision Res* 29: 1631–1647, 1989.
- Neri P, Heeger DJ. Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nat Neurosci* 5: 812–816, 2002.
- Neri P, Levi DM. Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vision Res* 46: 2465–2474, 2006.
- Nothdurft H-C. Attention shifts to salient targets. *Vision Res* 42: 1287–1306, 2002.
- Posner MI, Cohen Y. Components of visual orienting. In: *Attention and Performance X: Control of Language Processes*, edited by Bouma H and Bowhuis DG. Hillsdale, NJ: Erlbaum, 1984, p. 531–556.
- Qiu FT, von der Heydt R. Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. *Neuron* 47: 155–166, 2005.
- Roelfsema PR, Lamme VAF, Spekreijse H. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395: 376–381, 1998.
- Skiera G, Petersen D, Skalej M, Fahle M. Correlates of figure-ground segregation in fMRI. *Vision Res* 40: 2047–2056, 2000.
- Smith AT, Singh KD, Williams AL, Greenlee MW. Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cereb Cortex* 11: 1182–1190, 2001.
- Spitzer H, Desimone R, Moran J. Increased attention enhances both behavioral and neuronal performance. *Science* 240: 338–340, 1988.

- Supèr H, Spekreijse H, Lamme VAF.** Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nat Neurosci* 4: 304–310, 2001.
- Tadin D, Lappin JS, Blake R.** Fine temporal properties of center-surround interactions in motion revealed by reverse correlation. *J Neurosci* 26: 2614–2622, 2006.
- Von der Malsburg C.** The what and why of binding: the modeler's perspective. *Neuron* 24: 95–104, 1999.
- Vul E, MacLeod DI.** Contingent aftereffects distinguish conscious and pre-conscious color processing. *Nat Neurosci* 9: 873–874, 2006.
- Zhaoping L.** Border ownership from intracortical interactions in visual area V2. *Neuron* 47: 143–153, 2005.
- Zhou H, Friedman HS, von der Heydt R.** Coding of border ownership in monkey visual cortex. *J Neurosci* 20: 6594–6611, 2000.
- Zipser K, Lamme VAF, Schiller PH.** Contextual modulation in primary visual cortex. *J Neurosci* 16: 7376–7389, 1996.