Suppressive and facilitatory spatial interactions in foveal vision: Foveal crowding is simple contrast masking

Dennis M. Levi
School of Optometry, University of California, Berkeley, CA, USA

Stanley A. Klein
School of Optometry, University of California, Berkeley, CA, USA

Srividhya Hariharan
College of Optometry, University of Houston, Houston, TX, USA

Spatial interactions are a critical and ubiquitous feature of spatial vision. These interactions may be inhibitory (reducing sensitivity as occurs in crowding) or facilitatory (enhancing sensitivity). In this work, we had four goals. 1. To test the hypothesis that foveal crowding depends on target size by measuring the extent of crowding for novel targets that were limited in their spatial frequency content. We used a large range of target sizes and spatial frequencies. 2. To assess whether the critical spatial frequency model (Hess, Dakin, & Kapoor, 2000) provides a general model for foveal crowding. To test this model, we measured crowding for a direction-identification task that did not require judging the orientation of the gap. 3. To test the hypothesis that foveal crowding is simply contrast masking by remote flanks we measured and compared crowding in a direction-identification experiment with masking by remote flanks in a detection experiment. In each of the experiments, our targets and flanks were composed of Gabor features, thus allowing us to control the feature contrast, spatial frequency, and orientation. 4. To assess the relationship between suppressive and facilitatory spatial interactions in foveal vision. Our results show that (1) foveal crowding is proportional to feature size over the more than 50-fold range of target sizes that we examined. Over this large range, foveal crowding is scale invariant. Our results also show it is the size of the envelope (SD) rather than the carrier (SF) that determines the extent of crowding in the fovea. 2. Crowding that occurs in the direction-identification task is quite similar to crowding where orientation information is available. Thus we conclude that the critical spatial frequency model does not provide a general explanation for foveal crowding. 3. Threshold elevation for crowding is similar to threshold elevation for masking as predicted by our test-pedestal model. Thus we conclude that foveal crowding is simple contrast masking. 4. Based on our comparison of threshold changes in crowding and masking, we conclude that in foveal vision, the suppressive spatial interactions due to nearby flanks are similar in the two tasks. However, the facilitatory interactions are quite different. In the crowding task, we find very little evidence for facilitation by flankers, whereas in the detection task, we find strong facilitation. We suggest that facilitation of detection by remote flanks may be, at least in part, a consequence of uncertainty reduction.

Keywords: contrast sensitivity, spatial vision, masking

Introduction

Spatial interactions are a critical and ubiquitous feature of spatial vision that serve to sharpen perception of form and enable features to be grouped into forms. Spatial interactions may be inhibitory (reducing sensitivity) or facilitatory (enhancing sensitivity).

Crowding, the deleterious influence of adjacent flanks on visual discrimination, is a form of inhibitory interaction that is ubiquituous in spatial vision. Crowding effects occur in a variety of tasks, including letter identification (Bouma, 1970; Flom, Weymouth, & Kahneman, 1963; Toet & Levi, 1992), Vernier acuity (Westheimer & Hauske, 1975; Levi, Klein, & Atsibaamo, 1985), stereoacuity (Butler & Westheimer, 1978), and orientation discrimination (Westheimer, Shimamura, & McKee, 1976). In foveal vision, crowding typically only occurs over very small distances (2-6 arc min.; Flom, Weymouth, & Kahneman, 1963; Toet & Levi, 1992; Liu & Ariditi, 2000) or is reported not to occur at all (Strasburger, Harvey, & Rentschler, 1991). In contrast, crowding in peripheral vision, and in the central field of strabismic amblyopes, occurs over very large distances (in the periphery, up to half the eccentricity; Bouma, 1970; Toet & Levi, 1992; Kooi, Toet, Tripathy, & Levi, 1994) where the spread functions of the target and flanks are clearly separate.

There is not yet a widely accepted explanation for crowding. High level (i.e., attentional; He, Cavanagh, & Intriligator, 1996; Leat, Li, & Epp, 1999), low level...
(through lateral neural connections; Flom, Weymouth, & Kahneman, 1963; Polat & Sagi, 1993, 1994; Tripathy & Levi, 1994), and pattern masking (Chung, Levi, & Legge, 2001) have been proposed. However, a recent suggestion is that crowding is a consequence of the physics of the stimulus (Liu & Arditi, 2000; Hess, Dakin, & Kapoor, 2000). For example, the foveal degradation effect has, at least in part, been ascribed to the effect of the eye's point spread function when the letters are small and closely spaced (Liu & Arditi, 2000). It has also been argued that in foveal vision nearby flanks displace the critical spatial frequency band used to detect the orientation of the gap (horizontal vs. vertical) in a Landolt C to higher spatial frequencies, thereby reducing the visibility of the cue (Hess, Dakin, & Kapoor, 2000).

The optical explanation would predict that crowding occurs only for small targets near the limit of visual acuity and does not occur for large blurred stimuli. The predictions of the critical spatial frequency band explanation depend on whether it is the retinal spatial frequency or the object spatial frequency band that is critical. If it is the retinal spatial frequency band that is critical, then an upward shift of the critical spatial frequency band would shift the cue to higher retinal spatial frequencies that are less visible. This explanation would predict that crowding only occurs for targets near the limit of visual acuity and does not occur if the critical spatial frequency band is at low retinal spatial frequencies. On the other hand, if it is the object spatial frequency band that is critical, (e.g., 1.25 - 1.5 c/letter), then if flanks cause a shift to higher object frequencies, this shift may degrade discrimination (even at low retinal spatial frequencies) by shifting the letter outside the critical band. Indeed, there is a good deal of evidence that suggests that the critical spatial frequency band for letter recognition peaks between about 1 and 3 c/letter (Parish & Sperling, 1991; Solomon & Pelli, 1994; Alexander, Xie, & Derlacki, 1994; Chung, Legge, & Tjan, 2002). Because Hess, Dakin, and Kapoor (2000) used only a single (near the acuity limit) letter size, it is not clear whether their crowding effect was due to a shift in retinal or object spatial frequencies.

Another plausible but rather different explanation is that crowding occurs when the target and flank overlap within the same neural unit (e.g., both fall within a single receptive field). Note that this is more general than the critical spatial frequency band model (Hess, Dakin, & Kapoor, 2000) because it is not limited to the orientation cue (discussed in more detail later). Both of these explanations (shift in object frequency and overlap) predict that crowding would occur over a range of target sizes, rather than just at the acuity limit, and that the flanking distance would be proportional to the target size. In their classical study, Flom, Weymouth, and Kahneman (1963) estimated the extent of crowding by having normal and amblyopic observers judge the orientation of a Landolt C and varied the distance of surrounding flanks from the C. They found that the extent of crowding (i.e., the distance over which the flanks interfered with performance) was proportional to the observers' acuity and concluded that crowding is related to the size of the receptive fields that are most sensitive to the target. Their results led to an important but largely untested principle of acuity chart design—the idea that letter spacing should be proportional to letter size in order to keep the effect of contour interaction consistent across acuity levels.

However, because Flom, Weymouth, and Kahneman (1963) always used targets near the acuity limit, it is not clear whether crowding depends on target size.

The term crowding is ordinarily used to describe the fact that adjacent objects (letters or flanks) reduce the discriminability of a target. Several aspects of crowding make it mysterious; for example, in peripheral vision, crowding extends over very long distances (Bouma, 1970; Toet & Levi, 1992; Levi, Hariharan, & Klein, 2002) where the target and flanks do not overlap. The term masking is often used to describe the fact that a pattern (referred to as a mask) can reduce the discriminability of a target. Masking generally occurs for targets and masks that overlap, and the phenomenon of masking is reasonably well understood (Legge & Foley, 1980; Foley, 1994). Crowding and masking impair visual discrimination; thus, it is reasonable to ask whether they are two sides of the same coin (i.e., whether they share a common mechanism). Although several recent studies have addressed this question (Pelli & Palomares, 2000; Chung et al., 2001), it has been difficult to compare the two because they are measured with very different stimuli and tasks. One exception is the recent study by Parkes, Lund, Angelucci, Solomon, and Morgan (2001). They found that in peripheral vision, observers are unable to report the orientation of a target patch surrounded by flanking patches, but can accurately estimate the average orientation of an ensemble of such targets. They concluded that the local orientation signals are not lost (as would occur if there were masking) but instead are pooled before reaching consciousness. Interestingly, they found that this did not take place when the target was always presented at a known location in the fovea.

This work had four goals. The first goal was to test the hypothesis that foveal crowding depends on target size. To test this hypothesis, we measured the extent of crowding for targets that were limited in their spatial frequency content using a large range of target sizes and spatial frequencies. Second, we asked whether the Hess, Dakin, and Kapoor (2000) critical spatial frequency model provides a general model for foveal crowding. To test their model, we measured crowding for a task that did not require judging the orientation of the Landolt C gap. The third goal was to test the hypothesis that foveal crowding is simply masking by remote flanks (Chung et al., 2001). To test this hypothesis, we measured and compared crowding in a direction-identification experiment with masking in a detection experiment.
(similar to the experiments of Polat & Sagi, 1993, 1994). In each of the experiments, our targets and flanks were composed of Gabor features, thus allowing us to control the feature contrast, spatial frequency, and orientation. These experiments provide a bridge between traditional crowding experiments, and recent studies on spatial interactions using Gabor targets and flanks (Polat & Sagi, 1993, 1994; Zenger & Sagi, 1996).

As noted above, spatial interactions may also be facilitatory (enhancing sensitivity). For example, Polat & Sagi (1993, 1994) using Gabor patches suggested that there may be both excitatory and inhibitory interactions in contrast detection. They suggested that the facilitation that they observed in normal foveal vision was due to long-range neural connections. A number of physiological studies have shown that responses of neurons in V1 can be modulated (either increased or decreased) by surrounding stimuli outside the classical receptive field (Gilbert, 1998; Fitzpatrick, 2000) due to long and short-range interactions. Thus, the fourth goal was to assess the relationship between suppressive and facilitatory spatial interactions in foveal vision. To test whether the facilitation was due to stimulus uncertainty reduction, we investigated whether flanks reduced the slope of the psychometric function.

### General Methods

The stimuli, composed of Gabor or Gaussian patches, were displayed on one of two video monitors (a Monoray high brightness monitor with a mean luminance of approximately 80 cd/m² or a Mitsubishi Diamond Scan 20H monitor with a mean luminance of ≈ 56 cd/m²) using a Cambridge Research Systems (Cambridge, UK) VSG 2/3 graphics card with 15-bit contrast resolution. Six observers (including two of the authors) with normal or corrected-to-normal vision participated in one or more of the experiments. For all observers, viewing was monocular, with the untested eye occluded with a black patch. All observers were well practiced in making psychophysical judgments.

### Experiment 1: Crowding Depends on Size

#### Methods

The target was an E-like figure composed of 17 circular Gabor patches (five per side; Figure 1). On each trial the target was briefly presented (for 195 msec) with one of four orientations (up, down, left, or right) selected at random. The observer’s task was to identify the orientation. The target patches always contained a horizontal carrier, and each patch was separated from its neighbor by 3 standard deviations (center-to-center). Unless otherwise specified, the bandwidth of the patches composing the target was 0.825 octaves. The carrier was always in sine phase and the spatial period was typically equal to half of the separation so that there was phase coherence across samples. A horizontal bar thus consisted of a high contrast cycle alternating with a lower contrast cycle (approximately 60% of the higher contrast), which is described by the following equation:

$$\sin\left(2\pi x/\lambda\right) \left\{ G(x - 4\lambda, y) + G(x - 2\lambda, y) + G(x + 2\lambda, y) + G(x + 4\lambda, y) \right\}$$

where

$$G(x, y) = \exp\left(-1.5^2 \left( \frac{x^2 + y^2}{2\lambda^2} \right) \right)$$

We chose this E-pattern for several reasons: 1. Like letters, it is localized and highly familiar. 2. It is quite robust to the effects of jitter and undersampling and performance in normal (uncrowded) foveal vision can be well understood on the basis of an ideal observer model (Levi, Sharma, & Klein, 1997; Levi, Klein, & Sharma, 1999). 3. By varying the viewing distance, it is easy to study crowding over a wide range of spatial scales (note that in our study, the gaps in the E pattern were always much larger than the observers' resolution limit).

To assess the influence of the flanks on pattern perception, we measured the contrast threshold for identifying the orientation of the target using a four-alternative method of constant stimuli. Each of the four surrounding flanking bars was composed of five Gabor patches (Figure 1). Unless otherwise specified, the size, separation, and spatial frequency of the flanks were identical to those of the target, and flank contrast was 90%. From trial to trial, the target was presented at one of four near-threshold contrast levels (based on pilot experiments), and the resulting psychometric functions were fit with a Weibull function to estimate threshold for identifying the orientation of the target. Each threshold estimate, corresponding to the contrast resulting in 72.4% correct performance (d’ ≈ 1.6), was based on 100 trials. The contrast thresholds presented in “Results” are the weighted means of at least four individual threshold estimates. From run to run, we varied the flank distance (including infinity, which provided a measure of the unflanked performance) and the viewing distance to vary the target size. The flank distance was specified as the distance from the center of the flank to the center of the adjacent limb of the target. Figure 1 shows examples of the target with flanks at distances corresponding to 9, 4.5, 3, and 2 times the patch standard deviation. In some experiments, the E-patterns were composed of dark Gaussian rather than Gabor patches (Figure 2). In control experiments, we also varied the orientation and spatial frequency of the flanks (at a fixed viewing distance).
Figure 1. Examples of our Gabor E stimuli. The top left panel is the isolated E, which served as the target. The other panels show the E target surrounded by high-contrast flanks at separations equal to 9, 4.5, 3, and 2 standard deviations from the target. The bottom right panel shows flanks with the carrier oriented orthogonal to the target carrier. Both targets and flanks were composed of identical Gabor patches.
Figure 2. Examples of our Gaussian E stimuli. The top left panel is the isolated E, which served as the target. The other panels show the E target surrounded by high-contrast flanks at separations equal to 9, 4.5, 3, and 2 standard deviations from the target. The bottom right panel shows each flank consisting of just two patches placed in line with the cue (i.e., the gap locations).
Results

Foveal crowding depends on target size

Nearby flanks elevate thresholds for identifying the orientation of the E pattern. This effect of flanks is the hallmark of crowding. In normal foveal vision, the unflanked contrast threshold and the flank-to-target distance at which thresholds begin to rise depend on the target size. This can be seen in Figures 3 and 4, which show foveal performance for Gabor (Figure 3) and Gaussian (Figure 4) Es for a range of target sizes (the target size is specified by the standard deviation of the Gaussian envelope of the patches comprising the target). Note that it is target size (standard deviation), not spatial frequency, that determines crowding (Figure 5A); therefore, from here on we specify the flank distance in standard deviation units (SDUs). Thresholds for different spatial frequencies (1.67 and 3.33 c/degree) but the same standard deviations (12 arc min) are similar; however, thresholds for different standard deviations (12 and 24 arc min) but the same spatial frequency (1.67 c/degree) are quite different (Figure 5A). Moreover, foveal crowding does not occur when the targets and flanks have orthogonal carrier orientations (Figure 5B).

As will be quantified below, in foveal vision, the extent of crowding depends on target size over a wide range of target sizes (an approximately 50-fold range of target sizes). In order to quantify the extent of crowding, we estimated the critical distance (CD) by fitting the threshold versus flank distance (FD) data with Gaussian functions (curves in Figures 3-5) of the form:

$$\text{Th}_f = \text{Th}_{\text{unf}} \left(1 + \frac{(\text{FD})^2}{(\text{CD})^2}\right)$$

where Th is the flanked threshold; Th_unf is the unflanked threshold, and Peak is the amplitude of the Gaussian (its height in unmasked threshold units for a flank distance of 0). Nonlinear regression was used to estimate the three parameters, Th_unf, Peak, and CD. The Gaussian function provides a good fit to the data, and our novel parameterization specifies the critical distance for crowding as the flank distance that causes the unflanked threshold to double. This critical distance (specified in arc min) is proportional to the overall target size (Figure 6) for both Gabor (open circles) and Gaussian (gray symbols) targets. It is of interest that the extent of crowding in the fovea is similar when the flanks consist of five patches (solid circles) or just two (diamonds) placed in line with the cue (i.e., the gaps). The best-fitting power function (shown in gray) has an exponent of 0.99 ± 0.06. This figure clearly shows that in foveal vision, the critical distance is about one sixth of the overall target size, or about 2.5 times the target standard deviation (Figure 6, top abscissa) or approximately 0.9 times the separation. At this distance, the target and flanks clearly overlap (see lower panels of Figures 1 and 2).

Figure 3. Contrast thresholds versus flank distance. Performance for Gabor Es for a range of target sizes for three observers (the target size is coded by symbol size). The lines are the right side tails of Gaussian functions (see text) fit to the data.
Figure 4. Contrast thresholds versus flank distance. Performance for Gaussian Es for a range of target sizes for two observers (the target size is coded by symbol size). The lines are Gaussian tails (see text) fit to the data.

Interestingly, the data of Hess, Dakin, & Kapoor (2000) fall closely in line with our E data. We have fit their results (their Figure 1) with Gaussians (Equation 2) to extract a comparable measure of critical distance, and these are plotted at the sizes of their C targets in Figure 6 (red triangles). Despite the differences in stimuli, methods, and observers, it is clear that their results fall closely in line with ours.

In normal foveal vision, crowding is scale invariant, and is primarily determined by target size (SD). When replotted as threshold elevation (i.e., flanked threshold/unflanked threshold) versus target-to-flank distance expressed in SDUs (i.e., target-to-flank distance [in arc min], divided by patch SD [in arc min]), foveal performance over a wide range of pattern sizes collapses into a more or less unitary function (Figure 7).
Crowding causes 180-degree errors

To learn more about the effects of crowding, we analyzed the errors made by our observers. Specifically, the types of errors can be classified as being either 180-degree errors, in which the observer’s report is the mirror image of the actual orientation (i.e., the observer confuses up for down or left for right), or 90-degree errors, in which the observer confuses, for example, up or down with left. These errors are illustrated in Levi et al. (1999; Figure 11 [top]). An ideal observer model predicts different error rates for 180-degree errors than for 90-degree errors (Levi et al., 1999).

Figure 8 shows the proportion of 90- and 180-degree errors under conditions where crowding occurs (small flank distances, top panels) and under conditions where there is little or no crowding (large flank distances, lower panels) for observer D.L. Similar results were obtained for the other observers. Note that under conditions of crowding, there is a preponderance of 180-degree errors. Random performance (as would be expected for targets near their contrast thresholds) would result in twice as many 90-degree errors because there are twice the number of possible 90-degree confusions (e.g., up with left and up with right) as 180-degree confusions (up with down). Thus, the predominance of 180-degree errors seems to be
Our results show that for our E-like targets, crowding causes a specific loss of 180-degree (mirror image) discrimination. A very simple model explains the 180-degree loss for the E target. The crowding occurs at the outside of the E, so the outside border is masked. The middle bar of the E is less masked and it contains 90-degree but not 180-degree information. Thus, our result might be specific to the E target. A difficulty with the critical band (Bondarko & Danilova, 1997; Anderson & Thibos, 1999; Hess, Dakin, & Kapoor, 2000) or Fourier hypothesis is that it fails to make a clear distinction between inner features and outer features. Specifically, the critical band hypothesis posits that discrimination of the orientation of a target (such as a Landolt C or an illiterate E) involves two stages: the first stage is the selection of the spatial frequency channel that gives the maximum differential response to horizontal and vertical (Bondarko & Danilova, 1997; Hess, Dakin, & Kapoor, 2000) (i.e., the first stage determines the orientation of the gap [horizontal or vertical]). The second stage determines the position (i.e., left vs. right or up vs. down). Hess et al. make the assumption that for foveal viewing, positional accuracy is high, and that the visual system uses some representation of amplitude within a critical orientation/spatial frequency band to determine the orientation. According to their model, crowding will occur when the flanks shift the critical frequency away from the most sensitive spatial frequency band, and thus impair the orientation judgment. Thus, the model incorrectly predicts that the errors made under conditions of crowding will be predominantly 90-degree errors (i.e., observers should be unable to correctly identify the orientation). However, our error analysis indicates that observers make predominantly 180-degree errors under conditions of crowding, and the readers can easily verify for themselves that it is much easier to discriminate the orientation than the position of the gaps under conditions of crowding (Figures 1 and 2, with flanks at 2 SD). Clearly, the critical band hypothesis does not provide a general explanation for crowding effects.

Figure 8. Confusion analysis. We classified the errors as either 180-degree errors (mirror image errors) or 90-degree errors (nonmirror image errors). This figure plots the proportion of 90- and 180-degree errors under conditions where crowding occurs (small flank distances [2 SDUs], top panel) and under conditions where there is little or no crowding (large flank distances [6-9 SDUs], lower panel) for observer D.L. Note that under conditions of crowding, there is a preponderance of 180-degree errors. Under conditions of crowding, the observer is able to correctly judge whether the legs of the E are oriented vertically or horizontally, but is unable to correctly identify the location of the gaps (the reader can verify this in the lower left panels of Figures 1 and 2).

Experiment 2: Crowding in a Direction-Identification Task

To further test the Hess, Dakin, & Kapoor (2000) critical band hypothesis, in this experiment, we measured crowding using a 2-Alternative-Forced-Choice (AFC) direction-identification task, in which observers were required to make a 180-degree judgment, thus eliminating the need to extract the orientation of the gap.

Methods

The target was the same E-like figure composed of 17 circular Gabor patches that was used in Experiment 1, but in this experiment, we measured contrast thresholds.
for identifying the direction of the E-like pattern using a 2-alternative method of constant stimuli. In separate experiments, we measured contrast thresholds for left versus right discrimination and for up versus down discrimination. Each of the two flanking bars were composed of five Gabor patches (Figure 9). Unless otherwise specified, the size, separation, and spatial frequency of the flanks were identical to those of the target, and flank contrast was 90%. In the left versus right experiments, the flanks were placed on either side of the E. In the up versus down experiments, they were placed above and below it. From run to run, we varied the distance of the flanks from the target (specified as the distance from the center of the flank to the center of the adjacent limb of the target). Figure 9 shows examples of the target and flanks.

Figure 9. Examples of the E target and flanks used in Experiment 2. In the left versus right experiments, the flanks were placed on either side of the E. In the up versus down experiments, they were placed above and below it.
To test the generality of our results, we also measured contrast thresholds for identifying the position of the gap in a C-like pattern (Figure 10). The C pattern is actually a circle composed of 12 overlapped Gabor patches, with a gap produced by removing N patches. The patch overlap gives a ring contrast of about 1.48 times the contrast of the individual patches. The contrast at the gap is 0.47 times the patch contrast. Thus the gap has about 1/3 of the contrast of the ring. The patch spatial frequency (10 c/degree) and standard deviation (4') were identical to the E pattern. In our experiments, N = 1. This pattern has several advantages over the 'E' pattern. It varies smoothly in space, is more compact than the E (because the constituent patches are overlapped, it has a radius of 13.2'), and it has a single gap that does not provide the strong global orientation cue that is seen with the E. For the C pattern, the flanks consisted of a pair of single high-contrast (90%) patches (on either side for the left vs. right discrimination, and above and below for up vs. down discrimination) whose size, spatial frequency, and orientation were identical to the patches comprising C.

As in Experiment 1, the target was presented at one of four near-threshold contrast levels (based on pilot experiments), and the resulting psychometric functions were fit with a Weibull function to estimate threshold for identifying the orientation of the target. Each threshold estimate, corresponding to the contrast resulting in 81.6% correct performance ($d' \approx 1.29$), was based on 100 trials. The contrast thresholds presented in “Results” are the weighted means of at least four individual threshold estimates.

**Results**

Crowding occurs for 180-degree direction identification for Es (Figure 11A, Figure 12, and Figure 15A) and for Cs (Figure 11B). As for the standard 4AFC task (green squares in Figure 11A), crowding occurs when the flanks are less than about 3 standard deviations from the target. Interestingly, although there are slight asymmetries for L/R versus U/D discriminations, for the Es, crowding for 180-degree discriminations is as strong or stronger than in the 4AFC experiment. We also note that there are strong individual differences in the strength of crowding. For example, S.H. (Figure 12A) shows about only a 50% threshold elevation (at 2.5 SDUs), whereas D.L. and R.J. show a factor of 3 or more threshold elevation at the comparable flank distance.
Figure 11. Crowding occurs for 180-degree direction identification for Es (A) and for Cs (B). For comparison, the standard 4AFC task is shown by the squares in Figure 11A. The bow ties and hourglasses show threshold elevation for detecting a single patch with a pair of flanks in a collinear and noncollinear arrangement, respectively. Data are for observer D.L. The blue dotted line (upper panel) is a sum of two Gaussians fit to the single patch data (the fit is not shown in the lower panel because the data are the same as in the upper panel).

Figure 12. Crowding occurs for 180-degree direction identification for Es, similar to Figure 11A, but for observers S.H. and R.J. Note the strong individual differences in the strength of crowding.
Crowding in a 180-degree discrimination task cannot be explained simply on the basis of the two-stage critical spatial frequency (or Fourier) model (Bondarko & Danilova, 1997; Hess, Dakin, & Kapoor, 2000) because there is no orientation cue, and as shown in the “Appendix,” the critical spatial frequency model has limited utility. The Fourier model predicts dips in the threshold versus flank separation function that are not evident in the human data (see Figure 18).

So how can we account for foveal crowding? We propose a simple test-pedestal model. The test-pedestal model for the 180-degree (left vs. right) task is illustrated in Figure 13, and the Fourier representation of the 180-degree task is described in detail in the “Appendix.” The pedestal (Figure 13, top) is represented by the 15 circular patches, each with a strength of 1, and four circular patches at the locations corresponding to the possible gap positions, with a strength of 0.5. The test (Figure 13, center) consists of two pairs of patches, one pair with a strength of –0.5 at the locations corresponding to the gap, the other with a strength of +0.5 at the locations opposite the gap. The pedestal plus test (Figure 13, bottom) corresponds to an E pointing to the right. The pedestal minus test is an E pointing to the left. An ideal observer would perform the task by discriminating the pedestal plus the test from the pedestal alone. If flanks reduce the visibility (strength) of the test, then crowding would occur. This very simplistic model predicts that crowding is essentially masking. To test this prediction, we measured masking in a detection experiment with no pedestal (Experiment 3) and compared the results to the crowding obtained in a direction-identification experiment (Experiment 2).

Figure 13. Schematic illustration of the test-pedestal model for the 180-degree (left vs. right) task. The pedestal (top) is represented by the 15 circular patches, each with a strength of 1, and four patches at the locations corresponding to the possible gap positions, with a strength of 0.5. The test (center) consists of two pairs of patches, one pair with a strength of –0.5 at the locations corresponding to the gap, the other with a strength of +0.5 at the locations opposite the gap. The test plus pedestal (bottom) corresponds to an E pointing to the right. An ideal observer would perform the task by discriminating the pedestal plus the test from the pedestal alone.
Experiment 3: Masking and Facilitation of Detection by Remote Flanks

We are interested in comparing crowding (which is not well understood) with masking (which is). Specifically, we asked whether crowding is simply masking. Masking experiments usually involve detection of a target in the presence of a masking pattern. Although masking typically involves a completely overlapping target and mask, recent studies suggest that detection of a target may be influenced by adjacent flanks (Polat & Sagi, 1993, 1994). In this experiment, we measured contrast thresholds for detecting a single Gabor patch in the presence of surrounding flanks consisting of Gabor patches (similar to studies by Polat and Sagi, 1993, 1994).

Methods

The target in this experiment was a single Gabor patch with a horizontal carrier, identical to the patches used to form the E and C targets in Experiments 1 and 2 (Figure 14, top), and the flanks were a pair of high-contrast (90%) Gabor patches, usually with the same size, spatial frequency, and orientation as the targets. The flanks were either collinear with the target (one on either side; Figure 14, left column) or noncollinear (above and below; Figure 14, right column).

To make the detection experiment comparable to the crowding experiments (i.e., a single temporal presentation), we measured contrast thresholds for the briefly presented (195 msec) target using a rating scale method of constant stimuli (Levi & Klein, 1990). Briefly, on each trial the target was presented at one of four near-threshold contrast levels (including a blank, or 0 contrast level). Following each trial, the observer rated the magnitude of the contrast (from 0 to 3), and was given auditory feedback corresponding to the actual magnitude. A criterion-free estimate of the contrast detection threshold (specified at d' = 1) was obtained from the rating scale data. The thresholds reported here represent the average of at least four blocks of 100 trials/block, weighted by the inverse error. The error bars shown in the figures represent ± 1 SEM, and include both within and between run variation.

Results

Like crowding, thresholds for detecting a single patch are elevated by collinear flanks when the target-to-flank separation is less than about 3 SDUs (Figure 15A and 15B) (i.e., when target and flanks begin to overlap). However, the detection data differ from the discrimination data of Experiments 1 and 2 in that they show facilitation (thresholds are lowered) by flanks more than about 3 SDUs. The facilitation regime has been the main focus of the work by Polat and Sagi (1993, 1994); however, our main interest is the masking regime. Note Polat and Sagi’s argument that the critical metric is the flank distance in λ (spatial wavelength) units; however, note that the threshold elevation (and facilitation) are quite similar for stimuli with the same standard deviation (4 minutes, red symbols) but different spatial frequencies in Figure 15B. If the spatial wavelength was critical, these two curves should have been quite different because their wavelengths differ by a factor of 2. The lines, fit to all of the solo data, represent a difference of two Gaussians, a positive Gaussian with a small standard deviation (representing the suppressive effect of the flanks) and a negative Gaussian with a large standard deviation (representing the facilitary effect). Figure 15B also illustrates two additional points: increasing the number of patches in the flanks to five (to match the barlike flanks in Experiment 2) has no influence on the amount of threshold elevation; and, changing the orientation of the flank carrier from iso (i.e., horizontal, the same as the target) to cross (i.e., vertical or orthogonal to the target) eliminates the masking at small separations, and, interestingly, results in facilitation at larger separations. This cross-orientation facilitation has been previously reported in contrast-discrimination experiments (Yu & Levi, 2000).

Our main interest is comparing the effects of flanks on detection (Experiment 3) with the effects of flanks on direction discrimination (Experiment 2). Figures 11, 12, and 15A show the results of both experiments. The bow tie and hourglass symbols in Figures 11, 12, and 15A show the solo detection results for both collinear and noncollinear flanks, for comparison with the left versus right and up versus down E, respectively. As with the Es there are strong individual differences in the effects of flanks. S.H., who showed very little threshold elevation for Es, also shows little threshold elevation for the single patch, whereas D.L., who shows the strongest threshold elevation for E, similarly shows strong threshold elevation for the single patch. Importantly, for a given observer, at small flank distances, the threshold elevation for the two tasks is similar. As noted above, however, at larger separations, the effect of flanks is different in the two tasks: there is strong facilitation of detection, but little or no facilitation for the E direction discrimination (we speculate about this difference in “Discussion”). We find facilitation for both collinear and noncollinear flanks, and for observer S.H., contrary to Polat and Sagi (1993, 1994), the facilitation is actually stronger in the noncollinear case. In “Discussion”, we will consider the shape of the psychometric function (Table 1) and its implications for understanding facilitation.
Figure 14. Examples of the Gabor target and flanks used in Experiment 3 ('solo' detection). Contrast thresholds were measured for detecting a single Gabor patch in the presence of flanking Gabor patches. The target was a single Gabor patch with a horizontal carrier (top), identical to the patches used to form the E and C targets in Experiments 1 and 2, and the flanks were a pair of high (90%) contrast Gabor patches, typically with the same size, spatial frequency and orientation as the targets. The flanks were either collinear with the target (left column) or noncollinear (right column).
To further compare and contrast crowding and masking, we measured both solo detection and E
discrimination as a function of the flank contrast (Figure 16). At small flank separations (e.g., 2 SDUs, Figure 16,
top panel), where target and flanks overlap considerably,
for contrasts above about 5 times the flank detection
threshold, the effect of contrast on the two tasks is quite
similar, and thresholds for both tasks increase with an
exponent of ≈ 0.6, consistent with sine-on-sine contrast
masking (Legge & Foley, 1980). Note that at lower
contrast levels the two functions diverge, with solo
detection showing slight facilitation (below the red line,
as expected from the dipper form for masking) and
crowding showing slight threshold elevation (above the
red line). With more remote flanks (3 SDUs, Figure 16,
bottom panel), facilitation of solo detection spans a much
larger contrast range and thresholds only begin to rise
above 10 times detection threshold. Over this entire
range (up to about 10 times threshold), crowding shows a
small threshold elevation, becoming similar to solo
detection at the highest contrast levels.

To directly compare the effects of flanks in masking
and crowding, we plotted threshold elevation for
crowding (E or C direction discrimination) against
threshold elevation for masking (solo detection) for
paired conditions (e.g., L/R E vs. collinear solo at the
same flank distance; or U/D E vs. noncollinear solo
detection at the same flank distance; Figure 17). Each
symbol in Figure 17 represents a paired measure. Data
inside the red box show facilitation for solo detection and
data inside the green box show facilitation for E direction
discrimination. Clearly, there are many more points
inside the red box than inside the green, showing that
there is considerably more facilitation of detection.
However, for values above about 1.3 on the abscissa,
threshold elevation for the two tasks is quite similar, and
the data cluster around the 1:1 line. Thus we suggest that
in the normal fovea, the threshold elevation of crowding
and masking follows similar rules and shares a common
mechanism. In the fovea, crowding simply seems to be
masking.

Discussion

This work had four goals. The first goal was to test
the hypothesis that foveal crowding depends on target
size. To test this hypothesis, we measured the extent of
crowding for targets that were limited in their spatial
frequency content, over a large range of target sizes and
spatial frequencies. Our results show that foveal
crowding is proportional to feature size over the more
than fifty-fold range of target sizes that we examined. Over
this large range, foveal crowding is scale invariant. Our
results also show it is the size of the envelope (SD) rather
than the carrier (SF) that determines the extent of
crowding in the fovea.
The second goal was to ask whether the Hess, Dakin, & Kapoor (2000) critical spatial frequency model provides a general model for foveal crowding. The main effect of crowding for our E-like patterns in Experiment 1 was a loss of mirror image (180 degree) discrimination (Figure 8). This pattern of loss would not be predicted if the main effect of flanks is to shift the critical spatial frequency band for determining the orientation of the gap (i.e., a 90-degree discrimination). To test the Hess et al. model more directly, in Experiment 2, we measured crowding for a task that did not require judging the orientation of the gap. Our results show crowding in the absence of orientation information (the 180-degree task) that is quite similar to crowding where orientation information is available (the 4AFC task). Targets such as our E-like pattern, where the global orientation signal is strong, may represent a special case; however, it is worth noting that the local orientation (i.e., the orientation of the carrier) is also critically important. Foveal crowding is strong when the target and flanks have the same carrier orientation, and is absent when they are orthogonal (Figure 5B). Moreover, we found similar results with
C-like targets, which do not have a strong global orientation cue. Thus, we conclude that the Hess et al. model does not provide a general explanation for foveal crowding.

Our third goal was to test the hypothesis that foveal crowding is simply masking by remote flanks (Chung et al., 2001). To test this hypothesis, in Experiment 3, we measured and compared crowding in a direction-identification experiment with masking by remote flanks in a detection experiment (similar to the experiments of Polat & Sagi, 1993, 1994). Our main result is that for high contrast nearby flanks, threshold elevation for crowding is similar to threshold elevation for masking (e.g., Figure 17) as predicted by our test-pedestal model. Thus, we conclude that crowding is simply masking. This experiment also enabled us to address our fourth goal, which was to assess the relationship between suppressive and facilitatory spatial interactions in foveal vision. To assess the relationship, we measured both crowding and masking with stimuli composed of Gabor patches. Based on our comparison of threshold elevation in crowding and masking, we conclude that in foveal vision, the suppressive spatial interactions due to nearby flanks are similar in the two tasks. On the other hand, we note that the facilitatory interactions are quite different. In the crowding task, we found very little evidence for facilitation by flanks, whereas in the detection task we found strong facilitation with flanks as remote as 15 wavelengths, and at contrasts less than 3 times the flank detection threshold. Polat and Sagi (1993, 1994) have argued for a role for the long-range horizontal neural connections in facilitatory interactions. Long-range horizontal neural connections are known to exist in area V1 (Gilbert, 1998; Fitzpatrick, 2000). Below we elaborate on these results, and discuss several possible alternative explanations for the facilitation observed here.

Relation to Previous Studies

**Foveal crowding depends on target size**

Previous estimates of the extent of foveal crowding vary from none (Strasburger et al., 1991) up to 0.5 degrees (Chung et al., 2001). Strasburger et al. used low-contrast letters and, more importantly, low-contrast flankers, and reported little or no crowding in fovea. As noted in Figure 16, low-contrast flanks have little effect; suppressive interactions in the fovea evidently require high-contrast flankers. Studies using small high-contrast letters or optotypes have suggested that foveal crowding extends over a very small area (Flom et al., 1963; Woldorf & Chambers, 1984; Toet & Levi, 1992), usually only a few minutes of arc. These studies have typically used letters near the acuity limit. Similar estimates of the extent of suppressive foveal spatial interactions are obtained from studies of the effects of high-contrast flanks on Vernier acuity (Westheimer & Hauske, 1975; Levi et al., 1985), orientation acuity (Westheimer et al., 1976), and stereo acuity (Butler & Westheimer, 1978). Chung et al. obtained a much larger extent of crowding (≈ 30 minutes) using bandpass filtered letters that subtended an angle of ≈ 20 minutes. Our results suggest that the extent of spatial interactions in foveal vision should not be thought of as having a fixed retinal distance, but rather as being proportional to the size of the target (or a critical feature of the target). Thus, foveal crowding for letter-like targets is scale invariant. Interestingly, similar scale invariance is evident in spatial interval discrimination in the fovea (Levi, Jiang, & Klein, 1990). Scale invariance, and the way in which the extent of crowding is specified, can explain the large extent of crowding in the study of Chung et al. Recall that we specify the distance of the flank from the target as the distance from the center of the flank to the center of the adjacent limb of the target, whereas Chung et al. specify the distance between the center of the target letter and the center of a flanking letter. Based on Snellen construction (letter 5 times the limb size), Chung et al. specify abutting flanks at 5 times the distance that we specify. When specified relative to target size, our critical distance is equal to about 1/6 of the overall target size (or ≈ 0.8 times the limb or gap size).

Thus, the flank separation of Chung et al. should be divided by 5 to be comparable to our specification (i.e., 6'). Moreover, Chung et al. specified the critical distance as the distance at which threshold elevation falls to 0, whereas we specify our critical distance as the distance at which threshold is elevated by a factor of 2. The Chung et al. specification (threshold elevation = 0) increases their estimate of critical distance by about a factor of 2 relative to ours (threshold elevation = 2; see their Figure 7). If their separation and critical distance were specified like ours, their foveal crowding would extend approximately 3'—roughly one sixth of their foveal letter size. Thus, their estimate of a foveal critical distance of about 30' is fully consistent with our results.

Polat and Sagi (1993) have also argued that the critical distance in spatial interactions is not a fixed retinal distance. They argued that the critical distance is based on the spatial frequency (λ) of their Gabor targets. However, their data are not convincing on this point. Close inspection of their data (their Figure 3) shows that threshold elevation for the highest spatial frequency (λ = 0.075 degree) has its peak and minimum shifted by about a factor of 2 to the right. This condition had an envelope size (σ = 0.15 degree) that was half the size of that used for the other spatial frequencies (σ = 0.3 degree). Had Polat and Sagi (1993) plotted their data with the test-mask distance specified in σ rather than in λ units, their data would have superimposed rather nicely.

**Crowding in a direction-identification task**

Contrary to the model proposed by Hess, Dakin, & Kapoor (2000), we found that crowding can occur in the absence of an orientation cue. It should be noted that
crowding can occur under dichoptic conditions (target-to-one eye and flanks to the other; Flom, Heath, & Takahashi, 1963; Westheimer & Hauske, 1975; Kooi et al., 1994; Tripathy & Levi, 1994), making explanations based solely on retinal information unlikely. As shown in the “Appendix,” the critical spatial frequency model is not very useful. Specifically, the Fourier or critical spatial frequency model produces minima in the threshold versus flank distance function under conditions where human observers show strong crowding (Figure 18). Thus, the explanation we prefer is that foveal crowding is simply contrast masking.

**Figure 18.** The predicted masking strength at the optimal test frequency $f_{opt}$. The small dots on the plot are for mask locations at 2/3 and 4/3 separation units corresponding to the lower two panels of Figures A2 and A3 in the Appendix.

**Masking and facilitation of detection by remote flanks**

Remote flanks can produce both suppressive and facilitative interactions. In the crowding experiments, we found strong masking by nearby flanks, but little evidence for facilitation. Chung et al. (2001) also were unable to find evidence for facilitation in crowding. In contrast, our detection experiments, like those of Polat and colleagues (Polat & Sagi, 1993; Polat, Sagi, & Norcia, 1997; Polat & Tyler, 1999; Polat, 1999), we found facilitation to be as strong or stronger with the noncollinear as with the collinear configuration (compare bow ties and hourglasses in Figures 11 and 12). Moreover, we note that orthogonally oriented flanks may actually facilitate detection (Figure 15, blue hourglasses) as was also found by Yu, Klein, and Levi (2002). Polat and colleagues (1993, 1994, 1997, 1999) have argued that collinear facilitation reflects long-range intrinsic connections that interconnect like-orientation columns along their preferred orientation (Fitzpatrick, 2000; Gilbert, 1998). We hypothesize that no facilitation is seen for the letters because the pedestal provides a constant facilitation; thus, the flanks are unable to add more facilitation. We will revisit the issue of facilitation as a reflection of long-range intrinsic neural connections in “Mechanisms of spatial interactions.”

A surprising result in this study is that facilitation in the detection experiment was not limited to collinear flank configurations. In contrast to studies by Polat and colleagues (Polat & Sagi, 1993; Polat, Sagi, & Norcia, 1997; Polat & Tyler, 1999; Polat, 1999), we found facilitation to be as strong or stronger with the noncollinear as with the collinear configuration (compare bow ties and hourglasses in Figures 11 and 12). Moreover, we note that orthogonally oriented flanks may actually facilitate detection (Figure 15, blue hourglasses) as was also found by Yu, Klein, and Levi (2002). Polat and colleagues (1993, 1994, 1997, 1999) argued that collinear facilitation reflects long-range intrinsic connections that interconnect like-orientation columns along their preferred orientation (Fitzpatrick, 2000; Gilbert, 1998). We hypothesize that no facilitation is seen for the letters because the pedestal provides a constant facilitation; thus, the flanks are unable to add more facilitation. We will revisit the issue of facilitation as a reflection of long-range intrinsic neural connections in “Mechanisms of spatial interactions.”

**Mechanisms of spatial interaction**

**Suppressive interactions**

A number of quite different mechanisms have been proposed for the suppressive interactions that are commonly known as crowding or lateral masking. As noted in the “Introduction,” these vary from very low-level to high-level explanations. At one extreme is a retinal explanation where foveal crowding has been explained on the basis of the optics (Liu & Arditi, 2000) or the physics of the stimuli (Hess, Dakin, & Kapoor, 2000). This model accounts for foveal crowding on the basis that nearby flanks displace the critical spatial frequency band to higher spatial frequencies, thereby reducing the visibility of the orientation cue in the Fourier representation of the stimulus. For several reasons, we argue that this orientation cue explanation
cannot be complete. In the first experiment, we found that the main effect of crowding for our E-like patterns is a loss of mirror image (180 degree) discrimination, whereas the Hess et al. model predicts a loss of orientation (90 degree) discrimination. Targets such as our E-like pattern, where the global orientation signal is strong, may represent a special case; however, we found similar foveal crowding when the judgment is limited to a 180-degree discrimination (Experiment 2) even for C-like patterns. It is worth noting that the local orientation (i.e., the orientation of the carrier) is also critically important. Foveal crowding is strong when the target and flanks have the same carrier orientation, and is absent when they are orthogonal (Figure 5B). It should be noted that crowding can occur under dichoptic conditions (target to one eye and flanks to the other; Flom, Heath, & Takahashi, 1963; Westheimer & Hauske, 1975; Kooi et al., 1994; Tripathy & Levi, 1994), and taken together with the orientation specificity, this result makes explanations based on solely retinal information unlikely.

A not quite so low-level explanation for foveal crowding is that crowding occurs when there is overlap between the target and flank within the same neural unit (e.g., cortical receptive field; Flom et al., 1963; and/or hypercolumn; Levi et al., 1985). This notion predicts that crowding would occur over a range of target sizes, rather than just at the acuity limit, and that the flanking distance would be proportional to the target size. It also predicts that crowding would be orientation dependent. Thus, in this model, crowding is essentially contrast masking by nearby flanks (rather than by a superimposed mask), and will occur when there is overlap between the target and flank (either physically, or within the same neural unit) that obscures the cue. This is the basis of our test-pedestal model. Our results show that the extent of foveal crowding is strongly dependent on the size of the global features (the envelope SD) rather than the local features (carrier SF) and that it dependent on the local orientation information. Moreover, the strength and extent of suppressive interactions in crowding are similar to the extent and strength of interactions as measured in a detection task, as predicted by our simple test-pedestal model.

Our working hypothesis is that flanks and target combine at a second stage of visual processing (Pelli & Palomeres, 2000). In the fovea, the spatial extent of integration is determined primarily by the size of the features. As noted elsewhere (Levi, Hariharan, & Klein, 2002), and in previous work, in peripheral vision (Bouma, 1970; Jacobs, 1979; Levi et al., 1985; Toet & Levi, 1992; Kooi et al., 1994; Wilkinson, Wilson, & Ellemberg, 1997; Hess, Dakin, Kapoor, & Tewfik, 2000) and in strabismic amblyopia (Flom et al., 1963; Jacobs, 1979; Levi & Klein, 1985; Levi, Hariharan & Klein, in press) crowding occurs over much larger distances, where there is no physical overlap between target and flanks. If one adopts the prescription of Parkes et al. (2001) that crowding should be distinguished from the irretrievable loss of information that occurs through masking, then we would say, in agreement with them, that crowding does not occur for foveally centered targets. In our companion paper (Levi et al., 2002) we show that in peripheral vision crowding is distinct from masking, consistent with the recent study of Parkes et al. (2001).

Facilitative interactions

Facilitative interactions in spatial vision have been the subject of a great deal of recent experimentation and modeling. Remote flanks can facilitate detection (Polat & Sagi, 1993, 1994; Dresp, 1993; Kapadia, Ito, Gilbert, & Westheimer, 1995; Yu & Levi, 1997, 2000; Williams & Hess, 1998; Solomon, Watson, & Morgan, 1999; Chen & Tyler, 2001; this study) but not suprathreshold contrast appearance (Williams & Hess, 1998). Facilitation is strongest when the flanks are at a distance of about 3 to 6λ (when λ = σ), i.e., when the target and flanks are abutting, but may extend over long distances (more than 10λ). There are several possible explanations for facilitation by remote flanks:

1. Facilitation by remote flanks might be simply a consequence of the standard linear filter model that is the generally accepted model for facilitation by a low-contrast superimposed mask (Morgan & Dresp, 1995; Solomon et al., 1999). This model had been rejected because facilitation occurs for opposite sign (or phase) flanks. Thus, Zenger and Sagi (1996) argued for a two-stage filter model, with facilitation due to an accelerating nonlinearity applied at the second filtering stage. However, this objection (and the need for two filtering stages) has been called into question on two counts. First, Williams and Hess (1998) were not able to replicate facilitation with opposite phase flankers, and second, Solomon et al. showed that the standard linear filter model does in fact predict weak facilitation by opposite-sign flanks (consistent with their own measurements). It is possible that a single-contrast gain model might account for both the suppressive and facilitative effects on perceived contrast (Xing & Heeger, 2001); however, it is not clear that it could account for the effects on threshold reported here. Although this model could account for separations less than about 4σ, it would have trouble with larger separations.

2. Facilitative interactions in spatial vision have been the subject of a great deal of recent experimentation and modeling. Remote flanks can facilitate detection (Polat & Sagi, 1993, 1994; Dresp, 1993; Kapadia, Ito, Gilbert, & Westheimer, 1995; Yu & Levi, 1997, 2000; Williams & Hess, 1998; Solomon, Watson, & Morgan, 1999; Chen & Tyler, 2001; this study) but not suprathreshold contrast appearance (Williams & Hess, 1998). Facilitation is strongest when the flanks are at a distance of about 3 to 6λ (when λ = σ), i.e., when the target and flanks are abutting, but may extend over long distances (more than 10λ). There are several possible explanations for facilitation by remote flanks:

1. Facilitation by remote flanks might be simply a consequence of the standard linear filter model that is the generally accepted model for facilitation by a low-contrast superimposed mask (Morgan & Dresp, 1995; Solomon et al., 1999). This model had been rejected because facilitation occurs for opposite sign (or phase) flanks. Thus, Zenger and Sagi (1996) argued for a two-stage filter model, with facilitation due to an accelerating nonlinearity applied at the second filtering stage. However, this objection (and the need for two filtering stages) has been called into question on two counts. First, Williams and Hess (1998) were not able to replicate facilitation with opposite phase flankers, and second, Solomon et al. showed that the standard linear filter model does in fact predict weak facilitation by opposite-sign flanks (consistent with their own measurements). It is possible that a single-contrast gain model might account for both the suppressive and facilitative effects on perceived contrast (Xing & Heeger, 2001); however, it is not clear that it could account for the effects on threshold reported here. Although this model could account for separations less than about 4σ, it would have trouble with larger separations.

2. Facilitative interactions in spatial vision have been the subject of a great deal of recent experimentation and modeling. Remote flanks can facilitate detection (Polat & Sagi, 1993, 1994; Dresp, 1993; Kapadia, Ito, Gilbert, & Westheimer, 1995; Yu & Levi, 1997, 2000; Williams & Hess, 1998; Solomon, Watson, & Morgan, 1999; Chen & Tyler, 2001; this study) but not suprathreshold contrast appearance (Williams & Hess, 1998). Facilitation is strongest when the flanks are at a distance of about 3 to 6λ (when λ = σ), i.e., when the target and flanks are abutting, but may extend over long distances (more than 10λ). There are several possible explanations for facilitation by remote flanks:

1. Facilitation by remote flanks might be simply a consequence of the standard linear filter model that is the generally accepted model for facilitation by a low-contrast superimposed mask (Morgan & Dresp, 1995; Solomon et al., 1999). This model had been rejected because facilitation occurs for opposite sign (or phase) flanks. Thus, Zenger and Sagi (1996) argued for a two-stage filter model, with facilitation due to an accelerating nonlinearity applied at the second filtering stage. However, this objection (and the need for two filtering stages) has been called into question on two counts. First, Williams and Hess (1998) were not able to replicate facilitation with opposite phase flankers, and second, Solomon et al. showed that the standard linear filter model does in fact predict weak facilitation by opposite-sign flanks (consistent with their own measurements). It is possible that a single-contrast gain model might account for both the suppressive and facilitative effects on perceived contrast (Xing & Heeger, 2001); however, it is not clear that it could account for the effects on threshold reported here. Although this model could account for separations less than about 4σ, it would have trouble with larger separations.
fixed distance (in mm of cortex). In primate area V1, they extend only up to about 1 to 2 mm (Rockland & Lund, 1983; Blasdel, Lund, & Fitzpatrick, 1985; Fitzpatrick, Lund, & Blasdel, 1985; Lund, Yoshioka, & Levitt, 1993; Amir, Harel, & Malach, 1993). We do not know how long these connections are in human visual cortex, but if they are also around 1 to 2 mm, then they would extend over a distance of only about 3 to 6 minutes of arc in the fovea (Levi, 1999), far too short to account for the effects seen psychophysically. Moreover, the fixed cortical distance predicts interactions over a fixed retinal distance in the fovea, rather than interactions that are related in extent to either target size or spatial frequency. Although it is possible that large receptive fields are connected by long intrinsic connections, and small receptive fields by short ones, to our knowledge, the detailed physiology required to reveal this has not been done.

3. Facilitation by remote flanks is a consequence of uncertainty reduction. In a detection experiment, intrinsic uncertainty can elevate threshold (Pelli, 1985; Graham, 1989). Intrinsic uncertainty can take many forms: inability of the decision neural network to know the precise location, spatial frequency, orientation, or phase of the optimally stimulated sensory neurons (for review, see Pelli, 1985; Graham, 1989). Uncertainty models (Pelli, 1985) can predict a number of effects, including facilitation of contrast detection with near-threshold pedestals, and steepening of the psychometric function slope when uncertainty is high. Because the target in our detection experiment is a small, high spatial frequency Gabor patch, it is reasonable to consider the possibility that nearby high contrast flanks might reduce uncertainty about the stimulus, and, therefore, facilitate detection. This is essentially the explanation suggested by Williams and Hess (1998). Although we do not have any direct support for this account, we can evaluate one prediction of the uncertainty model: if optimally placed flanks reduce uncertainty, then the slope of the psychometric function (or transducer function) with flanks should be flatter than in the absence of flanks. Table 1 shows the mean exponent (slope) of the psychometric function relating d’ to stimulus contrast for each observer and condition in the absence of flanks, and with flanks at 4.5 SDUs (where facilitation is strong, and there is no overlap between target and flanks). The mean exponent with no flanks is ≈ 1.7, consistent with our previous detection studies. Interestingly, the mean exponent with flanks is lower: with collinear flanks ≈ 1.1 and with noncollinear flanks ≈ 1.4. It may be that noncollinear flanks are not quite as effective because they are not useful for reducing phase uncertainty. Uncertainty would also explain why we find facilitation in the masking experiment, but not in the crowding experiment. In the crowding case there is uncertainty reduction by the pedestal, so the flanks are irrelevant in reducing uncertainty. In the masking experiment, there is no pedestal.

We do not wish to imply that there is no neural basis for facilitation by flanks. Indeed, there are several findings that are not easily explained on the basis of uncertainty. For example, Solomon and Morgan (2000) and Adini, Sagi, and Tsodyks (1997) showed that facilitation by flanks was reduced when additional flanks were added. Similarly, experiments using noise show that facilitation of detection by orthogonally oriented (cross) surrounds is a result of excitatory interactions between orthogonal spatial filters (Yu et al., 2002). Our position is that both neural facilitation and uncertainty effects may contribute to the threshold-lowering effects of flanks. The change in detection exponents evident in Table 1 suggests that for our stimuli (like those of Pola & Sagi, 1993, 1994), uncertainty plays an important role.

Table 1. Psychometric Function Exponents

<table>
<thead>
<tr>
<th>Observer</th>
<th>SD (min)</th>
<th>SF (c/degree)</th>
<th>Exponent -No Flank</th>
<th>SE</th>
<th>Exponent Collinear Flank (4.5 SDUs)</th>
<th>SE</th>
<th>Exponent Non-Collinear (4.5 SDUs)</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>S.H.</td>
<td>4</td>
<td>10</td>
<td>1.66</td>
<td>0.31</td>
<td>1.02</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.H.</td>
<td>4</td>
<td>10</td>
<td>1.93</td>
<td>0.36</td>
<td></td>
<td></td>
<td>1.62</td>
<td>0.26</td>
</tr>
<tr>
<td>D.L.</td>
<td>4</td>
<td>10</td>
<td>1.36</td>
<td>0.25</td>
<td>1.03</td>
<td>0.26</td>
<td>1.21</td>
<td>0.20</td>
</tr>
<tr>
<td>D.L.</td>
<td>4</td>
<td>5</td>
<td>1.99</td>
<td>0.30</td>
<td>1.24</td>
<td>0.25</td>
<td>1.27</td>
<td>0.26</td>
</tr>
<tr>
<td>D.L.</td>
<td>8</td>
<td>5</td>
<td>1.76</td>
<td>0.27</td>
<td>1.19</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.L.</td>
<td>12</td>
<td>3.33</td>
<td>1.37</td>
<td>0.23</td>
<td>1.17</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J.T.</td>
<td>4</td>
<td>10</td>
<td>1.67</td>
<td>0.60</td>
<td>0.93</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td>1.68</td>
<td>0.09</td>
<td>1.10</td>
<td>0.05</td>
<td>1.37</td>
<td>0.13</td>
</tr>
</tbody>
</table>

SF = Carrier Spatial Frequency.
Conclusions

1. Foveal crowding is proportional to feature size over the more than 50-fold range of target sizes that we examined. Over this large range, foveal crowding is scale invariant.

2. Crowding occurs in the absence of orientation information. Thus, we conclude that the Hess, Dakin, & Kapoor (2000) model does not provide a general explanation for foveal crowding.

3. Threshold elevation for crowding is similar to threshold elevation for masking as predicted by our test-pedestal model.

4. Based on our comparison of crowding and masking, we conclude that in foveal vision, the suppressive spatial interactions due to nearby flanks are similar in the two tasks, and that foveal crowding is simple contrast masking; however, the facilitative interactions are quite different. Facilitation is evident only in the detection experiments. We suggest that uncertainty reduction may contribute to the facilitation produced by remote flanks.

Appendix: Fourier Information for Detecting 180-Degree Rotations of an E

This “Appendix” has two purposes: to consider the task of discriminating an E from a backward E based on the Fourier representation of the stimulus and to describe a formalism for writing analytic expressions for the Fourier representation of a class of letters.

The E to be considered consists of 17 Gaussian patches. The Fourier transform will be done in two steps. First we will calculate the Fourier transform of an E consisting of 17 points, and then we will convolve that E with the Gaussian pattern. In the Fourier domain, the spatial convolution becomes a simple multiplication by a Gaussian. The situation for Gabor patches is almost the same as the Gaussian case because the Fourier transform of the Gabor is simply a sum of an upward and a downward shifted (by an amount specified by the carrier frequency) version of the Gaussian pattern. Otherwise the analysis and conclusions are unchanged. The Fourier transform of a pattern, $P$, of discrete points is given by

$$F(f, g) = \sum_x \sum_y \exp(i(fx+gy)) P(x, y)$$

(3)

where $f$ and $g$ are the spatial frequencies (in units of radians/degree) in the x and y directions. The patterns that we used have equally spaced samples, with a sample spacing of $s$ degree. In the plots to appear in this “Appendix,” we will take the separation to be 0.2 degree. In these units, the Fourier transform of the middle horizontal bar of the pedestal (top panel, Figure 13) is

$$F_{mid}(fs) = \exp(-i2fs) + \exp(-ifs) + 1 + \exp(if) = 1 + 2 \cos(fs) + 2 \cos(2fs).$$

(4)

The factor of $i = (-1)^{1/2}$ occurs in Equation 6 because of the antisymmetry of the test pattern. The Fourier transform of the four bar mask is

$$F_{mask} = \cos(mgs)F_{mid}(fs) + \cos(mfs)F_{mid}(gs)$$

(7)

where $m$ is the spatial distance of the five mask samples from the central axis.

To take into account that each sample was actually a Gaussian rather than a point, the Fourier amplitudes must be multiplied by the function:

$$G(f, g) = \exp(-(f^2 + g^2) / 2\sigma^2)$$

(8)

where $\sigma = 1$/spatial sigma = 3/s because the spatial sigma is 1/3 of the separation.

The mask and pedestal are both real functions (symmetric) so that they can be directly added. The test pattern is purely imaginary, so it must be kept separate. The Pythagorean sum of the real and imaginary parts is of interest because its square gives the stimulus Fourier energy density that is relevant to how that stimulus excites a bank of mechanisms of different phases or positions.

This “Appendix” shows that the analytic Fourier representation of a somewhat complex object can be relatively simple. It is possible to work with numerically obtained Fourier transforms of letters, such as was done by Anderson and Thibos (1999). However, having the analytic representation allows a clearer and cleaner analysis.
Figure A1. The two-dimensional Fourier transforms of the E patterns for sample separation equal to 0.2 deg. (as in our experiments). Panel A is the pedestal, panel B is the test pattern, panel C is the pedestal plus mask, and panel D is the Pythagorean sum of panels A and C.

Figure A1 shows the two-dimensional Fourier transforms of the pedestal (A1A), the pedestal plus mask (A1C), the test (A1B), and the pedestal plus mask root energy (A1D), where root energy is the Pythagorean sum of the even and odd harmonics. For these and all subsequent figures, we have taken the base separation of the Gaussian patches to be $s = 0.2$ degrees, and the mask contrast is 20 times the pedestal contrast. The two-dimensional nature of the patterns makes it difficult to scrutinize the effects of the pedestal and mask on the test pattern. For this reason, we will also display a horizontal cut through the patterns with the rationale that because we are looking for right-left asymmetries, the horizontal cut at $g = 0$ would be the most reasonable place to look. The Fourier transforms become

$$F_{\text{ped}} = 3 + 6 \cos(fs) + 8 \cos(2fs)$$  \hspace{1cm} (9)$$

$$F_{\text{test}} = i \, 2 \sin(2fs)$$  \hspace{1cm} (10)$$

$$F_{\text{mask}} = 1 + 2 \cos(fs) + 2 \cos(2fs) + 5 \cos(mfs).$$  \hspace{1cm} (11)$$

The pedestal and mask can be combined as

$$F_{\text{ped plus mask}} = 3 + 6 \cos(fs) + 8 \cos(2fs) + c_{\text{mask}} (1 + 2 \cos(fs) + 2 \cos(2fs) + 5 \cos(mfs))$$  \hspace{1cm} (12)$$

where $c_{\text{mask}}$ is the contrast of the mask in units of the contrast of the E.
Figure A2. A horizontal cut at $g = 0$ of the Fourier amplitude for a rightward facing E composed of Gaussian patches. The top two panels are the test and pedestal alone. The bottom two panels are for mask locations of $2/3$ and $4/3$ in separation units (2 and 4 standard deviation units [SDUs]) from the outer patches of the letter E. The dashed red line in the second panel is the test pattern from the top panel to show the relative scaling. Similarly, the dashed line in the third panel is the test pattern from the second panel.

Figure A2 shows the one-dimensional Fourier amplitudes as specified by Equations 8, 10, and 12. The top panel is the test alone; the second panel is the pedestal alone. The third and fourth panels are for masks located at $2/3$ and $4/3$ separation units (2 and 4 SDUs) from the outer patches of the letter E. Note the difference in scale of the pedestal, test, and pedestal plus mask. The vertical lines in Figure A2 are placed at the frequency of the peak of the test pattern. For a separation of $s = 0.2$ degrees, the test pattern becomes

$$\text{test}(f_d) = \sin(2\pi f_d 0.4) \exp\left(-\frac{(2\pi f_d)^2}{2\sigma^2}\right)$$  \hspace{1cm} (13)$$

where $\sigma$ is the spatial sigma of 3/0.4 (Equation 8) and $f_d = f/2\pi$ is the spatial frequency in c/degree. The peak of the test pattern (ignoring the Gaussian envelope) occurs at a spatial frequency of $2.5/4 = 0.625$ c/degree. The vertical marks are placed at this frequency in all the panels because this is the optimal frequency for detecting the test pattern. We are interested in the Fourier amplitude of the mask at this frequency in order to estimate the expected amount of masking.

It is useful to view Figure A2 in terms of the Fourier energy by squaring the Fourier amplitudes of Figure A2, and expanding the abscissa and ordinate to facilitate examination of the prediction in Figure A3.

Figure A3. A replot of Figure A2 in terms of the Fourier energy (the square of the amplitudes in Figure A2). Note that for ease of viewing, the abscissa and ordinate have been modified relative to Figure A2. Although the dashed lines are present in the middle two panels as in Figure A2, they are barely visible because of the squaring operation.

A rough estimate of the threshold elevation of the test pattern due to the pedestal plus masker is given by the square root of the energy of the masker plus pedestal (the absolute value of the amplitude) at the optimal frequencies of the test. Figure 18 gives the value of the masking strength at the optimal test frequency $f_{opt}$. The small circles on the plot are for mask locations at $2/3$ and $4/3$ separation units corresponding to the lower two panels of Figures A2 and A3. The masking strength in Figure 18 has minima when the masker is 0.7 and 3.3 separation units from the center of the E. The actual data show no resemblance to this pattern. We conclude that Fourier analysis has limited usefulness in predicting the crowding effect.

Acknowledgments

We are grateful to Hope Queener for programming these experiments. This research was supported by Research grants R01EY01728 and R01EY04776, a Core Center Grant P30EY07551, and short-term training grant T35EY07088. All grants were from the National Eye Institute. Commercial Relationships: None.

References


