PERGAMON

Vision Research

Rapid communication

The foveal 'crowding' effect: physics or physiology?

Robert F. Hess *, Steven C. Dakin, Neil Kapoor

Department of Ophthalmology, McGill Vision Research Unit, McGill University, 687 Pine Avenue West, Room H4-14, Montreal, Que., Canada H3A 1A1

Abstract

It has been known for some time that both foveal and peripheral visual acuity is higher for single letters than for letters in a row. Early work showed that this was due to the destructive interaction of adjacent contours (termed 'crowding' or contour interaction). It has been assumed to have a neural basis and a number of competing explanations have been advanced which implicate either high-level or low-level stages of visual processing. Our results suggest a much simpler explanation, one primarily determined by the physics of the stimulus rather than the physiology of the visual system. We show that, under conditions of contour interaction or 'crowding', the most relevant physical spatial frequency band of the letter is displaced to higher spatial frequencies and that foveal vision tracks this change in spatial scale. © 1999 Elsevier Science Ltd. All rights reserved.

1. Introduction

It is well known that best acuity is obtained with single letters (Muller, 1951; Flom, Weymouth & Kahneman, 1963; Loomis, 1978; Strasburger, Harvey & Rentschler, 1991). This is true for the fovea (Flom et al., 1963), the periphery (Flom et al., 1963; Jacobs, 1979) and especially in cases of amblyopia (Flom et al., 1963; Hess & Jacobs, 1979; Levi & Klein, 1985). The detrimental effect of the proximity of nearby letters or contours (Loomis, 1978) is referred to as contour interaction or 'crowding' in the visual literature. A number of authors have speculated that it is due to either limitations at a low-level of visual processing (Flom et al., 1963; Estes, Allmeyer & Reder, 1976) or to attentional influences at a high-level of visual processing (Wagner, 1918; Strasburger et al., 1991). The former can be thought of, as either the lateral inhibition within

a single detector or inhibitory influences from distant neurones for which there is anatomical (Gilbert & Wiesel, 1989) and physiological (T'so & Gilbert, 1988) support. We show that there is a much simpler explanation, one based primarily on the physics of the stimulus not the physiology of the visual system.

2. Methods

2.1. Apparatus

An Apple Macintosh computer controlled stimulus presentation and recorded subjects responses. Programs for running the experiment were written in the Matlab programming environment (Mathworks Ltd.) using Psychtoolbox code (Brainard, 1997). Stimuli were displayed on a 21 inch Nanao FlexScan monochrome monitor, with a frame refresh rate of 75 Hz. Pseudo 12-bit contrast accuracy was achieved by electronically combining the RGB outputs from the computer using a video attenuator (Pelli & Zhang, 1991).

---

* Corresponding author. Tel.: +1-514-842-1231; fax: +1-514-843-1691.

E-mail address: rhess@bradman.vision.mcgill.ca (R.F. Hess).

2.2. Stimuli

Landolt C stimuli were based on an annulus with a strokewidth of 30 pixels and a total diameter or 150 pixels. A 30 pixel-wide gap was inserted into the annulus at either the top, bottom, left or right position on the annulus. Outline edges of the figure were not anti-aliased. In the flanked conditions, two horizontal (150×30 pixel) bars were positioned above and below the C, and two vertical (30×150 pixel) bars positioned to its left, and right. Flank distance was defined as the distance from the edge of the bar closest to the C, to the outer edge of the annulus defining the C. The standard stimulus appeared on a mid-grey (45 cd/m^2) background. Flanks either appeared black (same contrast polarity condition) or white (90 cd/m^2; opposite contrast polarity condition).

In the filtering condition, patterns were spatially band-limited by filtering them (having positioned them centrally within a 256 pixel square window) with an isotropic bandlimited, dc balanced filter (i.e. Laplacian-of-Gaussian filter):

$$\nabla^2 G(\sigma) = \left(1 - \frac{x^2 + y^2}{2\sigma^2}\right)\exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \qquad (1)$$

Images were normalised (0–255 grey levels), to maximize Michelson contrast, prior to presentation. The range of filtering tested was $\sigma = 4.78 - 18$ pixels.

2.3. Procedure

Subjects — two of the authors who had extensive practice on the task — performed a single-interval four-alternative forced task. They were presented with a Landolt C stimulus for 500 ms, and required to judge if the gap was in the top, bottom, left or right position. We first measured subjects ability to perform this task with unflanked Cs at a variety of viewing distances in order to determine the minimum angle of resolution (MAR) of the gap for each subject for this task. The viewing distance that produced 85–95% correct gap-position discrimination was then used in all subsequent conditions (this performance range was selected to avoid ceiling and floor effects). For subject RFH the viewing distance used was 8.6 m and for subject NK was 10.8 m. This corresponds to an MAR (i.e. gap width) of 0.024° for RFH and 0.019° for NK. Acuities were a factor of two better for the more conventional equivalent black/white stimuli.

Runs consisted of 100 trials but breaks were taken within runs to alleviate the effects of fatigue. Graphs show percent correct performance (from at least 200 presentations), with error bars denoting ± 1 S.E.

2.4. Modeling

The procedure for predicting the spatial frequency is based on the idea that performing the position discrimination task involves two sub-tasks. First, one determines the *orientation* of the gap (either horizontal or vertical) and then one determines its *position* (either left/right or above/below). The first stage is assumed to be the point at which spatial frequency selection takes place. If that is true Bondarko and Danilova (1997) reason that the most sensible spatial frequency for subjects to use when performing the position discrimination task is one that maximizes the difference in Fourier power at horizontal and vertical orientations. This can be determined by simply computing the discrete Fourier power spectrum of a stimulus, plotting the absolute difference between the horizontal and the vertical components (as a function of spatial frequency), and selecting the frequency that maximizes this function.

3. Results

Before entertaining the previous neural proposals for the 'crowding' effect, it is worth considering a much simpler explanation, one based on the physics of the stimulus. According to such an explanation, visual performance should be degraded because nearby contours interfere with the Landolt C in such a way that the energy in the frequency band most relevant to detection is increased at higher spatial frequencies for which the fovea has reduced sensitivity. The classical method for investigating the effects of adjacent contours is that originally used by Flom et al. (1963). Subjects are asked to identify the orientation of a Landolt C in a four alternate-forced choice task (up, down, right or left) in the presence of adjacent contour bars (see Fig. 1b,c) at various separations. Our results for stimuli modulated about a mean light level are shown in Fig. 1e,f by filled symbols; performance starts to deteriorate when the adjacent contours are two times the gap size (or barwidth) of the C, reaching a trough when abutting. Previous studies (see for review, Flom et al., 1963) have used black on white stimuli and shown even stronger effects with a partial release from interaction in the abutting condition. Modulating the stimuli about a mean light level (i.e. black on grey, see Fig. 1a–d) affords one the opportunity of having the letter and adjacent flanking contours of opposite contrast polarity (see Fig. 1d). Performance under these conditions (unfilled triangles) is very different, there is no longer any adverse effect of the adjacent contour, the foveal 'crowding effect' is abolished. A similar though less dramatic effect has been reported for peripheral vision (Kooi, Toet, Tripathy & Levi, 1994).

A previous study by Bondarko and Danilova (1997) showed that from a physical standpoint the most relevant spatial frequency band for detecting the orientation of an unflanked Landolt C (i.e. Fig. 1a) is around 1.15–1.30 cycles per letter, a factor of two lower than that of the gap itself. We wondered whether this would change in the flanking condition normally associated with contour interaction or 'crowding' (i.e. Fig. 1b). The results in Fig. 2 show the difference in the Fourier spectrum for orientations aligned and orthogonal to the gap in the C. The amplitude of the difference spectrum is plotted against the letter spatial frequency in cycles per letter; 2.5 cycles per letter representing a frequency with a halfperiod equal to the gap size. The corresponding absolute spatial frequency in cycles/degree is also indicated. This is different for each subject because of individual differences in absolute visual acuity (see Section 2). In the case of the unflanked C (Fig. 2, top left) the peak is see to be located at around 1.25 cycles per letter (Bondarko & Danilova, 1997) which corresponds to 9.36 c/deg for RFH and 11.7 c/deg for NK. Notice that when flanks are 1 barwidth or gap size away (Fig. 2, top right), the relevant spatial frequency band shifts higher by half an octave to 1.75 cycles per letter which corresponds to 12.5 c/deg for RFH and 15.6 c/deg for NK. Furthermore, when the flanks are of opposite polarity no significant shift occurs in the difference spectrum.

This led us to ask the following questions. First, does foveal vision in the unflanked condition (Fig. 1a) operate at the scale which the difference spectrum suggests is most pertinent (Fig. 2, top left)? Second, in the flanking condition (Fig. 1b) where contour interaction is evident (Fig. 1e,f), does the visual system shift its scale of analysis to the more relevant higher frequencies as indicated in the difference spectrum (Fig. 2, top right)? Third, in the flanking condition with opposite polarity bars (Fig. 1d), does the scale of analysis shift back to that seen in the unflanked (Fig. 2 bottom) condition? All of these predictions follow from an analysis of the power spectra alone. To answer this we measured performance for the same task but this time
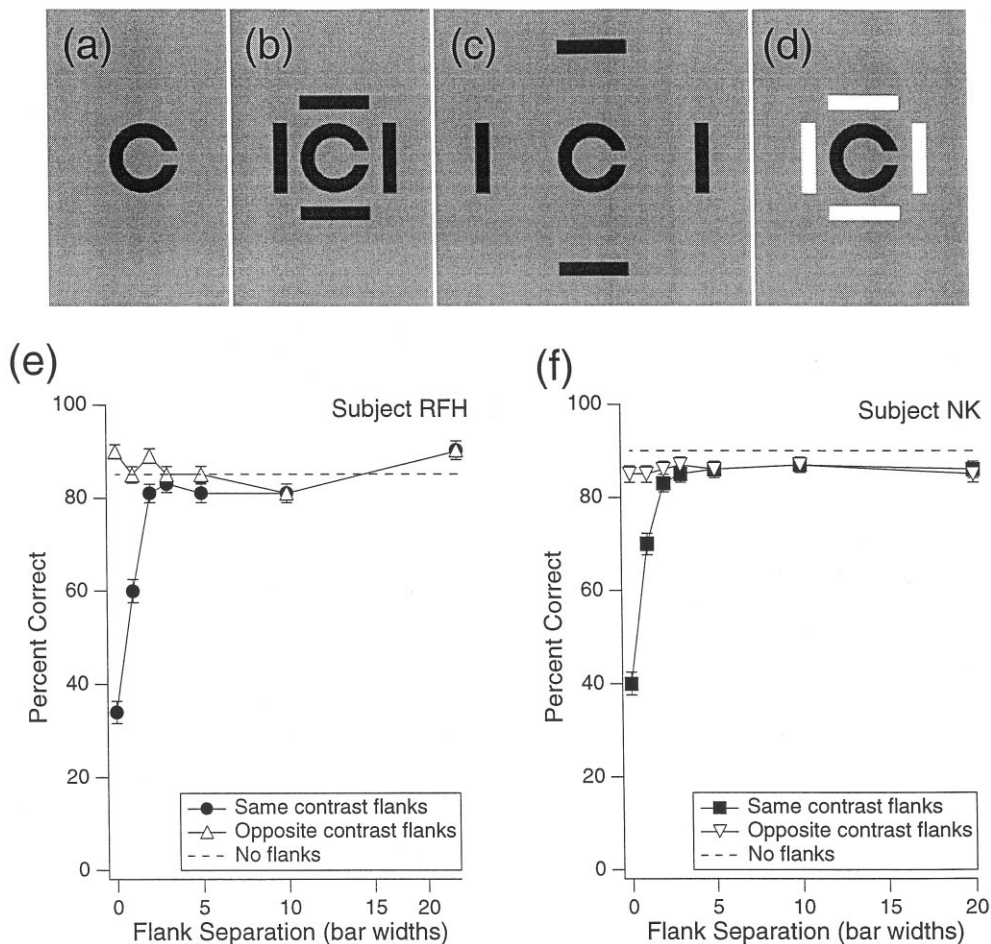

Fig. 1. Illustration of the stimuli used to measure contour interaction or 'crowding' (a–d). In a four AFC task, subjects (RFH & NK) had to detect the orientation of the gap in the C. Performance (percent correct) is plotted as a function of the separation of nearby contours (e,f). Performance was reduced when adjacent contours were within 2 barwidth separations (filled symbols). No measurable effect was seen when the adjacent contour was of opposite polarity (d and unfilled triangles in e–f).
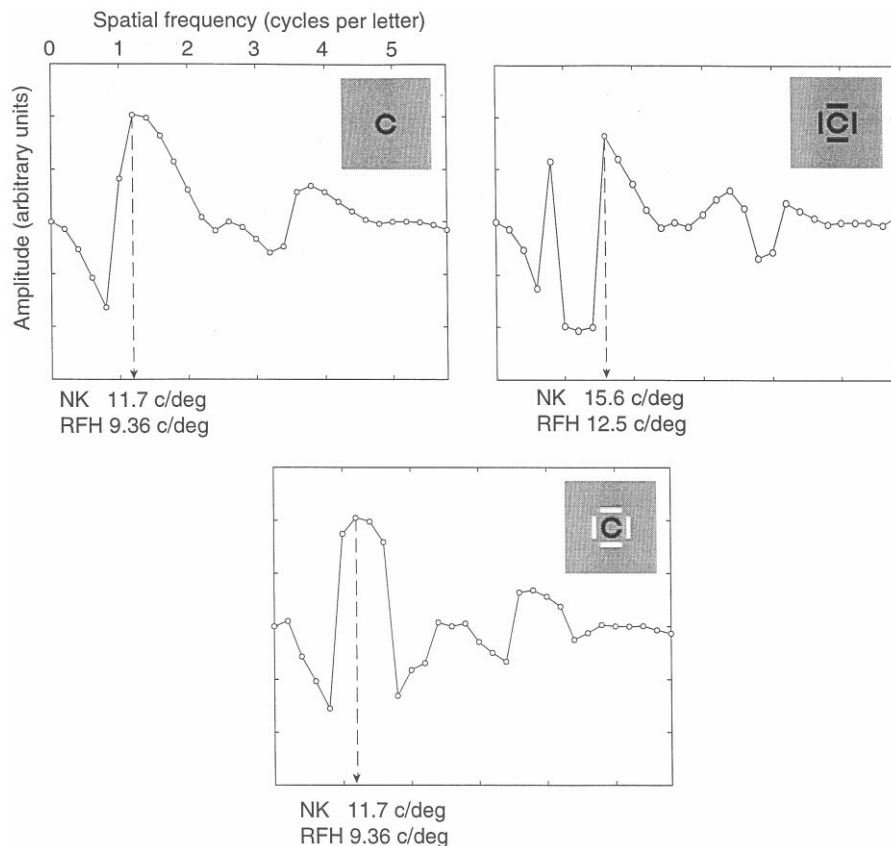
Fig. 2. Amplitude difference spectra taken aligned and orthogonal to the gap in the Landolt C for the stimuli seen in the upper right insets are plotted in terms of cycles/letter. The peak positions are indicated (vertical dashed line) in terms of the corresponding absolute spatial frequency for each subject in cycles/degree (same scale for all panels). These differ because the absolute visual acuity differs for the two subjects. Note the change in the position of the peak between the no flanking case (top left) and the 1 barwidth separation, flanking case (top right). This shift which is approximately half an octave does not occur when the flanking contour is of opposite polarity.

under conditions where subjects could use only a subset of the available spatial frequency information contained in the stimuli. This was achieved by filtering the stimulus such that the spatial phases of all spatial frequencies on either side of a specified passband were randomized (idealized notch-filtered noise thereby not altering the original amplitude spectrum). By moving the peak position of the notch-filter we were able to gauge the influence of specified spatial frequency bands in our letter acuity task, with the aim of determining what letter spatial frequency band underlay performance for different flanking conditions. These results are shown in Fig. 3 for two subjects. In the upper three panels, spatial frequency tuning functions are shown for each subject for the unflanked condition (Fig. 3a), for flanks at 1 barwidth separation (Fig. 3b) and for flanks at 1 barwidth separation for opposite polarity bars (Fig. 3c). The horizontal dashed line depicts performance levels for unfiltered stimuli and the vertical dashed line is the predicted position of the peak from the difference spectra of Fig. 2. These results demonstrate firstly that in the unflanked condition, the peak of the masking function is positioned at spatial frequen-

cies which closely match the peak in the difference spectrum (solid and dashed vertical lines in Fig. 3a) which is at about 1.25 cycles per letter. This provides the first proof that the peak in the different spectrum, first identified by Bondarko and Danilova (1997), is actually used by human vision. Second, it can be seen that when the adjacent contours are 1 barwidth away, performance utilizes higher stimulus spatial frequencies (Fig. 3b). The shift is about half an octave, matching that previously seen in the difference spectrum (solid and dashed vertical lines in Fig. 3b). This shift in the spatial scale subserving detection does not occur when the adjacent contours are of opposite contrast polarity (Fig. 3c) also paralleling that seen in the difference spectrum (solid and dashed vertical lines in Fig. 3c).

4. Discussion

Our results demonstrate that the classical 'crowding effect' for foveal vision is explicable in terms of the physics of the stimulus. It is not the result of inhibitory neural interactions within low or high level stages of

visual processing as previously thought (Flom et al., 1963; Estes et al., 1976; Loomis, 1978; Strasburger et al., 1991).These changes in the spatial scale utilized by the visual system are wholly predicted by the stimulus power spectra, and are contingent on conditions of contour interaction or 'crowding'. This suggests that the physics of the stimulus forces the visual system to use a spatial scale which while containing the most relevant information for the task is too high to support optimum performance (Campbell & Green, 1965).

4.1. Assumptions

We assume that the limiting operation on subjects determining the location of the gap in the C (u, d, l or r) is an initial processing stage to identify the orientation of the gap relative to the stimulus centre (u/d vs. l/r). following this stage the gap would be localised within the identified orientation band. For foveal viewing the assumption of high positional accuracy (i.e. relative to acuity) is justified. The validity of our analysis rests on the visual system having some representation of amplitude within a particular orientation/spatial frequency band, similar to assumptions based on any 'channel' based model.

To illustrate this scale specific loss of information due to the flanking contour, consider the images in Fig. 3d–f. All of these images (unflanked, flanked at 1 barwidth, flanked at 1 barwidth, opposite polarity) are filtered by a LoG with a standard deviation of 12.5 pixels. This corresponds to the spatial frequency tuning data indicated by crosses superimposed on the data points in Fig. 3a–c. This is the critical scale for the unflanked case and where performance falls in the same polarity flanked case. Note the 'C' in Fig. 3e shows a reduction in contrast energy and thinner strokes compared with either the unflanked (Fig. 3d) or opposite polarity flank condition (Fig. 3f). This physical change in the stimulus brought about by the flanks adequately explains the contour interaction or 'crowding' for our stimuli in foveal vision.

4.2. Caveats

We anticipate that there will be conditions for which a purely physical explanation will not suffice. For ex-
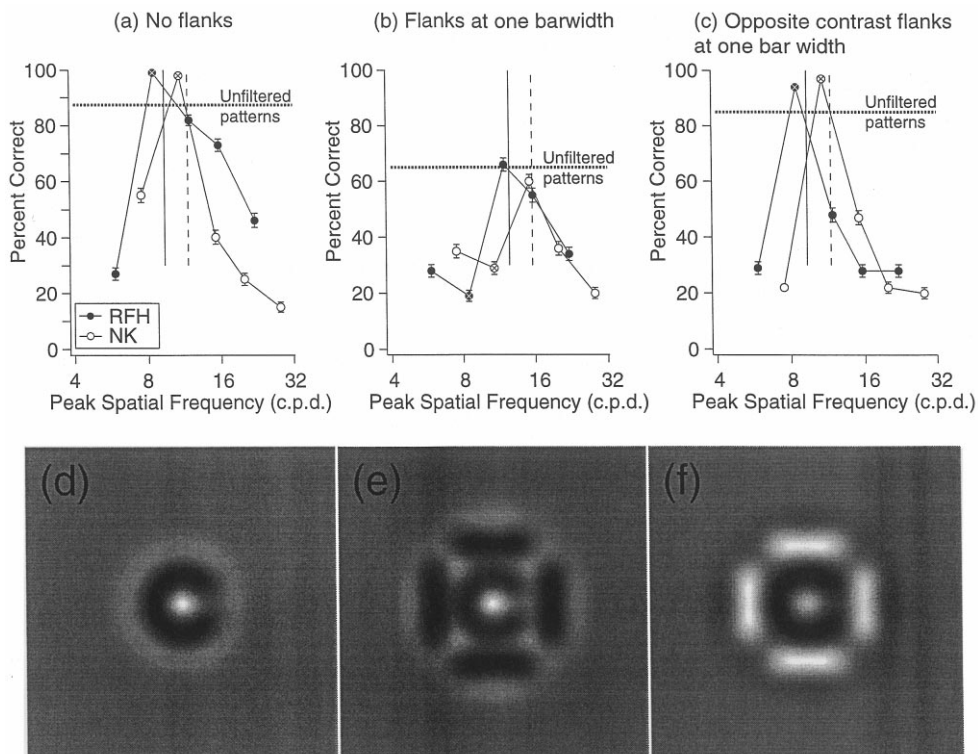

Fig. 3. Notch-filter tuning functions for Landoit C detection for two subjects (○ and ●) for unflanked (a), same (b) and opposite polarity (c) flanks at 1 barwidth separation. Percent correct is plotted against the peak spatial frequency of the passband in cycles/deg. The horizontal dashed line gives the unfiltered performance whereas the vertical solid and dashed lines give the predictions for the peak masking location based solely on the difference spectra (Fig. 2). The predictions differ slightly for each subject because their absolute acuity is not the same. The physical predictions (vertical lines) match the peak locations in the masking functions. d–f shows (d) unflanked, (e) flanked at 1 barwidth, same polarity (f) flanked at one bar width, opposite polarity all filtered with a LoG with a (c), of 12.5 pixels. This corresponds to the condition indicated by the crosses superimposed on data symbols in the graphs above (i.e. a–c). Note the 'C' in (e) shows great reduction in contrast energy with flanks present and the strokes look thinner than in (d) and (f). This is due to the shift shown in Fig. 2b.

ample, it is known that in peripheral vision a more pronounced interaction occurs when larger flanking elements are used (Loomis, 1978; Kooi et al., 1994). Such interactions are reduced but not abolished by reversing the contrast polarity of the flanking elements (Kooi et al., 1994). We expect that under these conditions there may be a significant neural contribution. Similarly, the crowding effect originally reported by Flom et al. for black/white stimuli shows a release from interaction in the abutting condition which is not present for stimuli modulated about a mean level. Although we do not have an explanation for this it may be related to the better acuity (factor of 2) obtained in the black/white condition. However, the extent of any neural contribution can only be known after the purely physical explanation that we have proposed is first taken into account.

Acknowledgements

References

Bondarko, V. M., & Danilova, M. V. (1997). What spatial frequencies do we use to detect the orientation of a Landolt C? *Vision Research*, *37*, 2153–2156.

Brainard, D. H. (1997). Spatial Vision. *The Psychophysics Toolbox*, *10*, 433–436.

Campbell, F. W., & Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *Journal of Physiology* (*London*), *181*, 576–593.

Estes, W. K., Allmeyer, D. H., & Reder, S. M. (1976). Serial position functions for letter identification at brief and extended exposure periods. *Perception and Psychophysics*, *19*, 1–15.

Flom, M. C., Weymouth, F. W., & Kahneman, D. (1963). Visual resolution and contour interaction. *Journal of the Optical Society of America*, *53*, 1026–1032.

Gilbert, C. D., & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, *9*, 2432–2442.

Hess, R. F., & Jacobs, R. J. (1979). A preliminary report of acuity and contour interaction across the amblyope's visual field. *Vision Research*, *19*, 1403–1408.

Jacobs, R. J. (1979). Visual resolution and contour interaction in the fovea and periphery. *Vision Research*, *19*, 1187–1195.

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interactions in peripheral vision. *Spatial Vision*, *8*, 255–279.

Levi, D. M., & Klein, S. A. (1985). Vernier acuity, crowding and amblyopia. *Vision Research*, *25*, 979–991.

Loomis, J. M. (1978). Lateral masking in foveal and eccentric vision. *Vision Research*, *18*, 335–338.

Muller, P. (1951). Uber das schen der amblyopen. *Ophthalmologia*, *121*, 143–149.

Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research*, *31*, 1337–1347.

Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for the identification of numeric charters in direct and eccentric view. *Perception Psychophysics*, *49*, 495–508.

T'so, D. Y., & Gilbert, C. D. (1988). The organization of chromatic and spatial interactions in the primate striate cortex. *Journal of Neuroscience*, *8*, 1712–1727.