Vision Res. Vol. 32, No. 2, pp. 347–363, 1992
Printed in Great Britain.

0042-6989/92 $5.00 + 0.00

A Visual Nonlinearity Fed by Single Cones

DONALD I. A. MACLEOD,* DAVID R. WILLIAMS,† WALTER MAKOUS†

Received 2 February 1991; in revised form 14 June 1991

An intensive nonlinearity in the visual system can produce distortion products, or difference frequency gratings, when observers view two high contrast, high spatial frequency interference fringes of slightly different frequency or orientation added together at the retina. These distortion products are visible even when the two fringes imaged on the retina are above the resolution limit. Our experiments take advantage of this nonlinearity to measure the spatial filtering in the visual system following the formation of the retinal image, but preceding the site of the nonlinearity. The point spread function corresponding to this spatial filter is so small that it can be entirely explained by light integration within the apertures of foveal and parafoveal cones. The small size of this point spread function implies that (1) laser interferometry avoids contrast losses inherent in the eye's optics at spatial frequencies as high as 130 c/deg, (2) retinal scatter causes negligible image degradation in the fovea and parafoveal retina, (3) eye movements have little or no effect on contrast sensitivity to the distortion product and (4) that there is no neural spatial summation in the visual system prior to the site of the nonlinearity. Distortion products could also be observed when a bright interference fringe was briefly flashed on the fovea and a test interference fringe was viewed through the resulting afterimage. Measurements of the point spread function at stages in the visual system that precede the generation of this distortion product were similar to those obtained with simultaneous presentation of the two fringes, implying that the aftereffect of light adaptation is extremely local, no larger than the dimensions of single cones.

Adaptation Resolution Optics Cones Photoreceptors

INTRODUCTION

Visual resolution requires independent processing of signals representing different points in the visual field; yet both psychophysical and neurophysiological observations show interactions over some distance from which losses of contrast sensitivity might be expected to ensue. In this paper we ask how these losses are distributed over successive stages of the visual system. We use psychophysical observations to dissect the visual system, separating the losses that precede a nonlinear process from the losses that follow it. Our results demonstrate spatial independence (at those stages preceding the nonlinear interaction) on so fine a scale that they have implications for the effective aperture of cone photoreceptors, mechanism of light adaptation, the optical properties of the eye, and the visual effects of eye movements during fixation. A summary of these results was presented previously (MacLeod, Williams & Makous, 1985).

GENERAL METHODS

Logic of the technique

In Fig. 1 the visual system is depicted as a cascade of optical and neural processes, with an interposed nonlinear stage. The resolution of each stage of processing may be characterized by its spatial frequency response which (neglecting phase shifts and transformations of scale) is simply the factor by which contrast (or its neural representation) is changed during transmission through each stage, as a function of the spatial frequency, f, of a sinusoidal stimulus. Our aim is to exploit the nonlinearity to estimate the spatial frequency response, $N_1(f)$ of the neural spatial filter that precedes the nonlinearity but follows the formation of the retinal image. Measuring this requires that we disentangle $N_1(f)$ from $N_2(f)$, the spatial filtering that occurs following the nonlinear stage.

The experimental study of the spatial frequency responses of these stages as well as the spatial spread of adaptation, has been hampered by two persistent problems. First, control of the spatial distribution of light on the retina is limited by diffraction and by the effects of ocular aberrations. Results suggesting that sensitivity is not locally controlled could in principle be due to optical rather than neural spread. In our work we have minimized optical spread by using laser interferometry; the stimuli were sinusoidal gratings formed at the retina by the interference of two coherent beams of light entering the eye at different points within the pupil (Campbell & Green, 1965; Williams, 1985a). It can be argued on theoretical grounds that this technique prevents degradation of the contrast of the sinusoidal fringe at the retina both by diffraction at the pupil and by

*Department of Psychology, C-009, University of California, San Diego, La Jolla, CA 92093, U.S.A.
†Center for Visual Science, University of Rochester, Rochester, NY 14627, U.S.A.

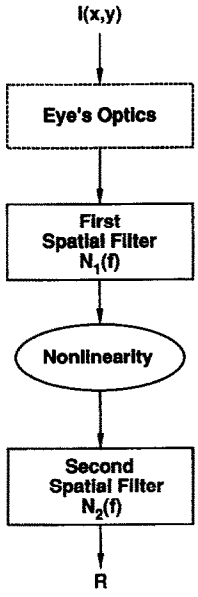

FIGURE 1. Cascade of optical and neural stages through which signals from interference fringe stimuli must pass. We assume initially that the optics of the eye are not involved because of the use of laser interferometry in all our experiments, an assumption ultimately supported by our data. The goal of these experiments is to characterize the spatial filter preceding the compressive nonlinearity, which has a modulation transfer function, $N_1(f)$, avoiding the effects of any spatial filtering following the nonlinearity, $N_2(f)$.

aberrations in the optics of the eye. We will return to this point later, providing empirical support for this claim.

The second problem is posed by the limited resolution of neural stages following the nonlinear site. These later stages may obliterate the signals from high frequency fringes that reach the site of the nonlinearity, making it impossible to use psychophysical observations to characterize the resolving power of the earlier stages. What is needed is a means of avoiding the later neural contrast sensitivity losses just as laser interferometry avoids the preceding loss attributable to the optics. We next describe a procedure that meets this need.

Burton (1973) showed that when two high frequency gratings are added at the retina, a nonlinearity in the visual system can produce a grating of low spatial frequency, or distortion product, visible to the observer but not present in the original stimulus. In our experiments we use such distortion products to separate resolution losses that precede the nonlinear site from those that follow it. Specifically we have investigated how well high frequency fringes are resolved at the nonlinear site, that is how well their modulation is preserved during transmission through the stages preceding the nonlinear site. We can do this without any influence from the spatial frequency response of later neural stages, by monitoring the detectability of a constant distortion product generated by the nonlinearity. In this situation the visual system up to and including the nonlinear process has to resolve the fringes themselves, but the neural processes following the nonlinear stage need only resolve the relatively low-frequency distortion product. By keeping the spatial frequency of the distortion product not only relatively low but constant (at 10 c/deg) we ensure that the frequency dependence of transmission through later neural stages could have no influence on our data.

Figure 2(a) shows an example of the stimulus we used in our experiments, which consists of the sum of two laser interference fringes. In all but one of our experiments these two fringes, which we will call "primaries," were presented simultaneously. In most experiments, it was convenient to use primaries identical in space-average illuminance (I_m) and in spatial frequency (f c/deg), but differing slightly in orientation. One fringe was rotated counterclockwise from the horizontal through an angle θ, and the other clockwise by the same amount. The intensity distribution thus had the form

$$I(x,y) = I_m\{2 + A_p\cos[2\pi f(x\sin\theta + y\cos\theta)] + \cos[2\pi f(-x\sin\theta + y\cos\theta)]\}, \quad (1)$$

where x and y are the horizontal and vertical spatial coordinates in degrees within the stimulus field. A_p is the contrast of one primary while the other primary always had unity contrast.

Note that although the image in Fig. 2(a) contains vertically oriented stripes, defined by the waxing and waning of local contrast, the pattern has no vertical spatial frequency components. This can be seen by inspecting the Fourier transform of the pattern as shown in Fig. 2(b). In the two-dimensional spatial frequency plane, there is no energy along the horizontal axis which would correspond to vertical spatial frequency components. Figure 2(c) shows the results of passing the two fringes through a nonlinearity. An arbitrarily chosen nonlinearity, a squaring of the intensity profile, was used as an example. Figure 2(d) shows the Fourier transform of this pattern, which has distortion products at orientations and spatial frequencies other than those of the primary fringes. (The image shows only the central portion of the transform; very high frequency components are also produced by the nonlinearity that are not shown.) The distortion product of particular interest for the present experiments is the component at the lowest spatial frequency. As one increases one's viewing distance from the plate, the (vertical) low frequency distortion product remains clearly visible in Fig. 2(c) even when the (nearly horizontal) primaries themselves cannot be resolved. Nonlinearities caused by printing this image can also produce visible distortion products in Fig. 2(a), but without nonlinearities the figure would appear as a uniform field if the primaries themselves were not resolved.

The distortion product lowest in spatial frequency is a vertical grating with a horizontal spatial frequency equal to $2f \sin\theta$. This is the vector difference between the primary spatial frequencies, and it is substantially lower than f provided that θ is small. We refer to this distortion product as the *difference frequency grating*. By varying θ along with f, the difference frequency $2f \sin\theta$ can be kept constant. In this way it is possible to test the effect of varying the primary spatial frequency while maintaining

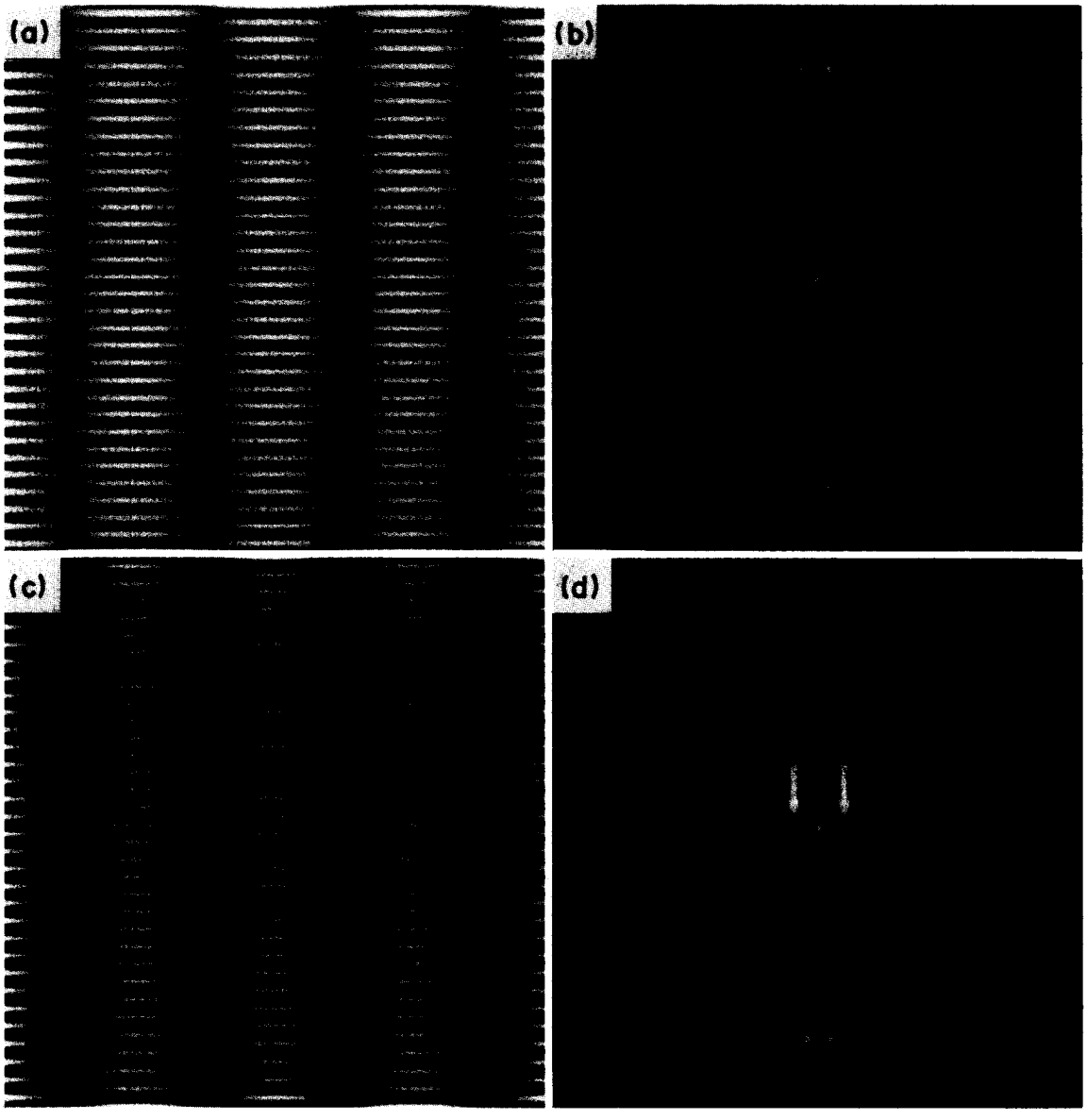

FIGURE 2. An example of the stimulus used in our experiments, which consisted of the sum of two laser interference fringes. These "primaries" have the same spatial frequency but differ slightly in orientation. Though the image contains vertically oriented stripes, defined by the waxing and waning of local contrast, the pattern has no vertical spatial frequency components. This can be confirmed by placing the pattern at a distance at which the high frequency primaries are no longer resolved. Nonlinearities inherent in the reproduction process can produce visible distortion products in (a), but without them (as on the retina) it blurs to a uniform field. However, the high contrasts available with laser interferometry make the nonlinear distortion easily visible. (b) The two-dimensional Fourier transform of the pattern in (a). The bright spot at the center of the image corresponds to zero spatial frequency. There is no energy along the horizontal axis, which would correspond to vertical spatial frequency components. (c) The result of passing the two fringes through a nonlinearity. An arbitrarily chosen nonlinearity, a squaring of the local intensity profile, was used. Note that this produces energy at a low spatial frequency corresponding to the contrast modulation in (a). As one increases one's viewing distance from the plate, the (vertical) low frequency distortion product remains clearly visible in (c) even when the (nearly horizontal) primaries themselves cannot be resolved. (d) The Fourier transform of (c), which has distortion products at orientations and spatial frequencies other than those of the primary fringes. The distortion product of particular interest for the present experiments is the component at the lowest spatial frequency, indicated by the arrows.

a distortion product of constant spatial frequency and orientation. This feature is critical for our goal of avoiding, or at least keeping constant, the effects of neural filtering following the nonlinear site.

The technique provides an estimate of the spatial frequency response preceding the nonlinearity because the amplitude of the distortion product depends on the amplitudes of the linear signals representing the fringes themselves at the input to the nonlinear site. Our procedure was to measure the amplitude (fringe contrast) of one of the primaries that was required to produce a constant (threshold) distortion product, keeping the amplitude of the other primary fringe fixed. Even assuming that laser interferometry does effectively avoid the optical losses, these amplitudes still undergo attenuation by the factor, $N_1(f)$, in transmission through the first spatial filter preceding the nonlinearity. The intensity distribution of equation (1) therefore generates, at the linear input to the nonlinear stage, an excitation profile given (in arbitrary units) by

$$I(x,y) = I_m\{2 + A_p N_1(f)\cos[2\pi f(x\sin\theta + y\cos\theta)] + N_1(f)\cos[2\pi f(-x\sin\theta + y\cos\theta)]\}. \quad (2)$$

It is convenient to consider the deviation of this linear signal from its space-average level. We denote this by $\xi(x,y)$:

$$\xi(x,y) = I_m\{A_p N_1(f)\cos[2\pi f(x\sin\theta + y\cos\theta)] + N_1(f)\cos[2\pi f(-x\sin\theta + y\cos\theta)]\}. \quad (3)$$

The amplitude of the distortion product, then, depends on $N_1(f)$, but it also depends on the form of the nonlinearity. For present purposes this can be represented simply by a power series in ξ, and we only consider here the linear and quadratic terms. (For a more detailed discussion, see Makous, Williams & MacLeod, 1991.) The nonlinear response R then becomes

$$R = c_1\xi + c_2\xi^2. \quad (4)$$

Substituting equation (3) into equation (4) yields the sum of a number of linear and quadratic terms. The critical term for the generation of the low frequency distortion product, which depends upon both primaries, is:

$$2c_2 I_m^2 A_p N_1^2(f)\cos[2\pi f(x\sin\theta + y\cos\theta)] \times \cos[2\pi f(-x\sin\theta + y\cos\theta)]. \quad (5)$$

Use of the trigonometric identity for the product of two cosines allows us to simplify this expression as follows:

$$c_2 I_m^2 A_p N_1^2(f)\{\cos[2\pi f(2y\cos\theta)] + \cos[2\pi f(2x\sin\theta)]\}. \quad (6)$$

It is the second term in this expression that generates the low frequency distortion product. We ignore all other terms in what follows and assume that, in our experimental conditions, variations in sensitivity to the stimulus that depend on the spatial frequency of the primaries are determined solely by the amplitude of this particular distortion product. (We present empirical support of this assumption below.) At the threshold for detection of this distortion product, the spatial frequency of which was fixed at 10 c/deg in our experiments, the amplitude of the distortion product is presumably constant at a value which we denote by A_d^*. Experimentally this threshold is reached by setting the contrast, A_p, of one of the primary fringes to a threshold value, A_p^*, that varies with the primary spatial frequency f. We thus have

$$A_d^* = (1/2)c_2(A_p^*)^{-1}N_1^2(f). \quad (7)$$

If we define contrast sensitivity, $S(f)$, as the reciprocal of the primary contrast, A_p^*, that brings the distortion product to contrast threshold, we have

$$S(f) = (1/2)c_2(A_d^*)N_1^2(f). \quad (8)$$

The unknown neural frequency response or transmission factor for the part of the system preceding the nonlinearity, $N_1(f)$, is therefore proportional to the square root of the contrast sensitivity as defined in equation (8). The square root relationship arises because the distortion product varies in proportion to *both* fringe amplitudes, each of which is attenuated by the factor $N_1(f)$.

Apparatus

The two interference fringe patterns required for these experiments were produced by separate interferometers, each fed by a different Helium–Neon (632.8 nm) laser. One interferometer was as described elsewhere (Williams, 1985a). In this interferometer, the contrast of the sinusoidal fringe was varied by gating the two interfering beams independently, with acousto–optic modulators, to form 1 msec rectangular pulses at a pulse rate of 400 Hz. Since the two beams interfere only when simultaneously admitted to the eye, the contrast of the fringe is proportional to the temporal overlap between the two independent pulse trains, which was controlled by computer. This variable-contrast interferometer will be referred to as interferometer A. In this paper we use the term "contrast" to refer to the Michelson contrast of one or other of the separate fringe patterns (unless we explicitly indicate otherwise). The *retinal* contrast due to each component will be only half of the contrast of that component so defined, for its contrast is reduced on the retina by equal illumination from the other interferometer.

The other interferometer (B) was designed to provide very high retinal illumination for bleaching experiments (more than 10^8 td, enough to bleach most of the visual pigment in about 10 msec). It did not incorporate modulators for controlling fringe contrast, which except where noted below was always maximal, being reduced below unity contrast only by optical losses. A triangular design for the interferometer, shown in Fig. 3, was adopted in the interests of mechanical stability and economy of components. Light from a 10 mW He–Ne laser forms a point source at S after transmission through the beam-expanding lens L_1. A pinhole at S serves as a spatial filter. Light from this point source is split by beamsplitter B_1 in two components that make either clockwise or

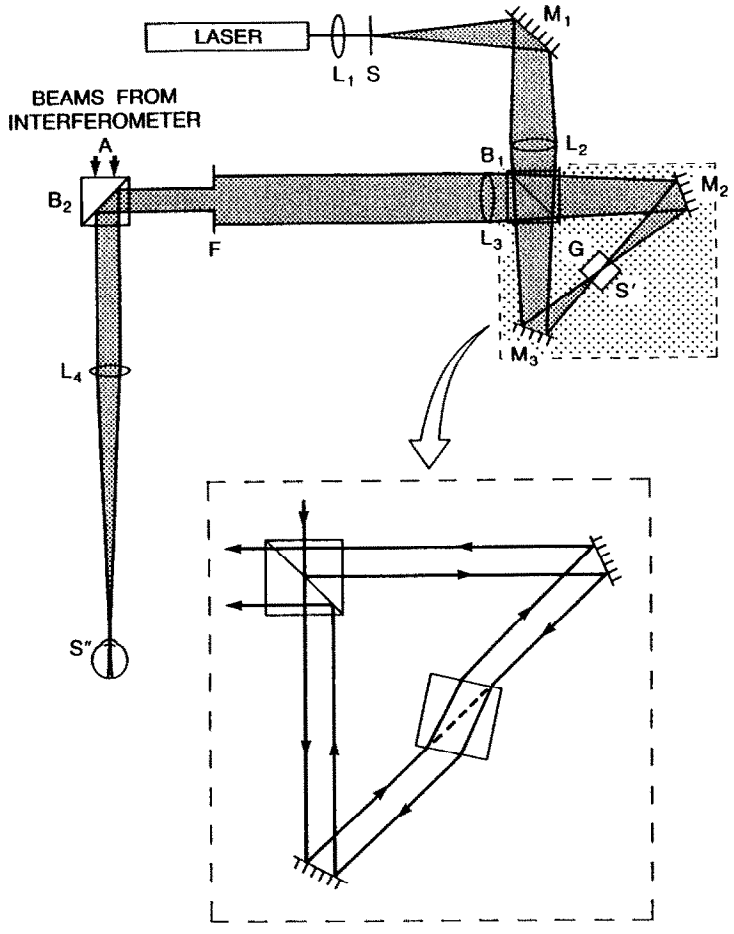

FIGURE 3. The design of the second interferometer (B) used in these experiments. The inset shows the triangular interferometer and the method for controlling fringe spatial frequency and orientation. See text for details. The design for interferometer A can be found in Williams (1985a).

counterclockwise transits around the triangle $B_1M_2M_3$. Two images of the point source at S are formed by lens L_2 at S′, with the light forming each image traveling in opposite directions. These two beams are recombined by the same beamsplitter B_1. On exiting from the triangle the two beams are each collimated by lens L_3. The circular field stop, F, subtending 1° of visual angle at the retina, is placed roughly one focal length beyond L_3, and one focal length from L_4. This final lens focuses the two collimated beams to form twin images of S in the plane of the pupil at S″, symmetrically displaced from the observer's Stiles–Crawford maximum, or point of greatest sensitivity. The separation between these images in the pupil plane determines the fringe spatial frequency on the retina through the equation

$$f = \pi d/180\lambda, \tag{9}$$

where d is the pupil separation measured in the same units as the wavelength λ. This relationship was verified by placing a Ronchi ruling in the plane of the field stop at F so that the fringe and the ruling generated a Moiré pattern. By adjusting the orientation of the glass plate G we could reduce the Moiré frequency to zero by bringing the dark stripes of the fringe into register with those of the ruling. This calibration procedure, used by Williams (1985), has the advantage that it requires no assumptions about the eye's optics so long as spatial frequency is defined in terms of the external angle subtended at the eye. The separation of the point sources at S′ and therefore at S″ is controlled by the tilt of a glass plate G at S′. The inset in Fig. 3 shows an expanded view of the transit of the chief rays from the two beams around the triangle and through the glass plate. Adjusting the tilt of the plate causes symmetric displacements of the two beams about the optic axis of the instrument and the observer's Stiles–Crawford maximum, allowing adjustments of spatial frequency without the need to realign the observer. A graded neutral filter was interposed in one of the beams near S′ to equate the intensities of the interfering beams so as to produce a fringe of unity contrast. The beamsplitter B_2 combines the beams from this interferometer with those from interferometer A, which is independently controllable in spatial frequency, orientation, contrast and space-average illuminance. Since the two fringes are mutually incoherent, their sinusoidal intensity distributions simply add at the retina.

Since the pupillary eccentricity of the light beams increases in proportion to fringe frequency, a compensation for the directional sensitivity of the retina (Stiles–Crawford effect) is required to keep the effective space-average illuminance the same for the different fringe frequencies. This compensation was made on the basis of measurements of the Stiles–Crawford effect for

each subject. The measure of directional sensitivity used was the threshold for 20 Hz flicker of a centrally fixated 1° field uniformly lit with laser light.

The three authors served as subjects, but the main phenomena we discuss have also been demonstrated to naive observers. Correcting lenses were worn, and they produce a small change in the spatial frequency of the interference fringes, which was taken into account.

RESULTS

Simultaneously Presented Primary Fringes

Procedure

In most of these experiments we measured the visibility of the distortion product by determining the lowest value of A_p, the fringe contrast from interferometer A, that could reliably be detected in the presence of the full contrast fringe from interferometer B. Thresholds were generally measured with a two alternative temporal forced-choice technique with feedback. The full contrast fringe from interferometer B was always present throughout each block of trials. On each trial, two 500 msec intervals (separated by an interval of 1 sec) were defined by tones. During one of these intervals a fringe from interferometer A was introduced without changing space-average illuminance. After each trial the observer had to indicate which interval contained the fringe presentation. Contrast varied from trial to trial in an adaptive search routine (Watson & Pelli, 1983), and contrast threshold is the value estimated, from a Weibull function fit to the results of 50 trials, to yield 75% correct responses. Standard errors of the threshold are based on variation between the individual estimates from different blocks of 50 trials each.

Results

As in the original observations of Burton (1973), all observers could see distinct distortion products, even when the spatial frequencies of the fringes themselves were too high to be resolved. These distortion products subjectively resembled physical sine wave gratings, in many cases of high contrast, with a chromatic component that made the dark bars look to most observers desaturated or even greenish. Under conditions where the primaries were visible, the dark bars always occurred in the *in-register* or high contrast regions of the stimulus. This polarity of the distortion product is consistent with a compressive or saturating nonlinearity; this is further discussed in another paper (Makous *et al.*, 1991). Most observers also reported the desaturation, rapid flicker, and "zebra stripes" characteristically observed (Williams, 1985a, 1988) when high contrast high frequency patterns interact with the receptor mosaic, and these effects were visible in the dark bars of the distortion product. One other distinction between the distortion product and conventional gratings of the same frequency is that the light and dark bars of the distortion product appeared almost equally wide, whereas other gratings have thicker bright bars than dark bars (Pelli, 1986), a point we will return to later.

Distortion products were also visible for stimuli that differed slightly in spatial frequency instead of orientation. This was the way Burton (1973) originally produced the effect. Moreover, distortion products at spatial frequencies $2f_1 - f_2$ were also visible, when the fringe frequencies from the two interferometers deviated slightly from a 2:1 ratio. These distortion products, which we term "cubic difference frequency gratings" because of their theoretical connection with a cubic term in the power series expansion of the nonlinearity, were also discovered by Burton (1973). Their significance is discussed in another paper (Makous *et al.*, 1991).

Contrast sensitivity for the distortion product. Figure 4 shows contrast sensitivity at a space-average retinal illuminance of 1000 td from each fringe for three observers. Here and elsewhere error bars subtend ±1 standard error based on variation among sessions. Generally, set blocks of 50 trials were run at each frequency, randomly interleaved. The results for the three observers (especially those for DM and DW) are in good agreement. As expected, visualization of the difference frequency grating required progressively higher contrasts in the variable contrast fringe as fringe frequency increased. But the increase was remarkably gradual, and contrast thresholds remained measurable at fringe frequencies as high as 130 c/deg of visual angle, roughly twice the normal limit of visual resolution. Thus primary spatial frequencies this high must be available at the nonlinear site. For fringes in the range 160–200 c/deg, we were unable to visualize any difference frequency gratings. Gaussian functions were fit to the data.

Transmission to the nonlinear site. The preservation of sensitivity for the fixed-frequency distortion product at high fringe frequencies in Fig. 4 is especially remarkable because the two interference fringes *both* undergo attenuation en route to the nonlinear site, and the

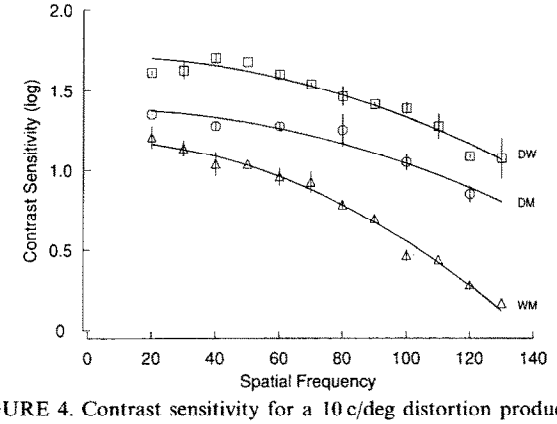

FIGURE 4. Contrast sensitivity for a 10 c/deg distortion product as a function of the spatial frequency of the two primary fringes presented simultaneously to the retina, for three observers. Curves for DW and DM are displaced vertically 0.3 and 0.6 log units, respectively, for clarity. Vertical error bars represent ±1 times the standard error based on variation among sessions, and are generally smaller than the symbols. The curves show the Gaussian functions fit by a least squares procedure.

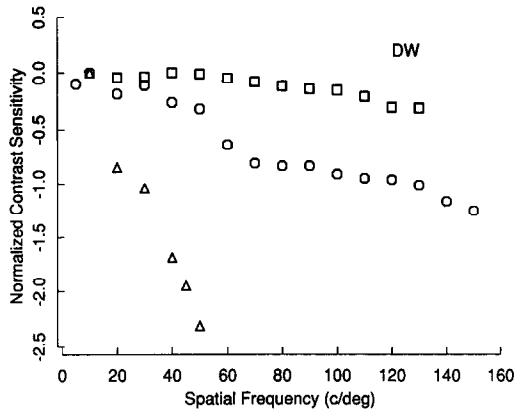

FIGURE 5. Squares show the transmission factor, $N_1(f)$ as a function of fringe frequency for observer DW. It was computed by taking the square root of the raw data for this observer shown in Fig. 4. The circles show contrast sensitivity for a single interference fringe measured on the same observer (from Williams, 1985a). The triangles show contrast sensitivity measured in the conventional manner with incoherent white light from a CRT display.

detectability of the difference-frequency grating depends on the square of the first attenuation factor, $N_1^2(f)$, in equation (8). The neural transmission factor for the part of the visual system preceding the nonlinear site is simply $N_1(f)$, and varies as the square root of the observed contrast sensitivity.

The squares in Fig. 5 show this transmission factor as a function of fringe frequency for observer DW. The relative reduction in amplitude drops very slowly, exceeding a factor of two (0.3 log units) only for spatial frequencies of 100 c/deg or more. The circles show, for comparison, contrast sensitivity for a single interference fringe measured for the same observer (Williams, 1985a). This function also extends to very high frequencies because of aliasing associated with sampling by the receptor mosaic (Williams, 1988). Yet even these data show a steeper decline than the difference-frequency sensitivities do, especially in the range below 60 c/deg where aliasing is unimportant. Clearly only a small part of the spatial integration of such interference fringes precedes the nonlinear distortion that produces the difference-frequency grating. The nonlinearity is strictly local; just how local is discussed below. By the same token, nearly all the neural attenuation implicit in the direct, fringe detection measurements (circles, Fig. 5) must occur in the later part of the system and be embodied in the transmission factor $N_2(f)$. Finally, the triangles in Fig. 5 show contrast sensitivity measured in the conventional manner with incoherent white light. The sensitivity loss at high frequencies is even greater, demonstrating the importance of optical relative to neural losses in vision.

Detectability of the primary fringes themselves. There are several lines of evidence that these thresholds represent detection of the distortion product rather than the high frequency fringe required to generate it. First, the pattern seen at small multiples of the threshold value of A_p was of the spatial frequency and orientation appropriate to the distortion product, rather than of the fringe itself; only in a few conditions indicated below was the test fringe resolvable as such. We kept the distortion product at 10 c/deg, which is near the peak of the contrast sensitivity function measured with interference fringes, while the primaries were invariably of high spatial frequency and poorly resolved by later neural stages. This maximized the visibility of the distortion product relative to that of the fringes required to generate it.

Our procedure of fixing the steady fringe at full contrast and setting the threshold by varying the contrast of the briefly pulsed fringe is also helpful in this regard: for correct identification of the test interval, detection of the steady fringe is useless, and the observer's judgement must be based either on detection of the pulsed fringe as such or on detection of the pulsed distortion product. By minimizing the contrast of the pulsed fringe we minimize the chance that its detection can form the basis of a correct judgement.

Moreover, the invisibility of the pulsed test fringe was demonstrated by an experiment in which the test fringe was exposed without the opportunity to generate a visible difference frequency. To do so, we simply rotated the unity contrast fringe of interferometer B from an orientation nearly parallel with the test fringe to the orthogonal orientation. In this way the threshold for the test fringe was measured under conditions as close as possible to those used for generating the difference frequency grating, that is, in the presence of a steady, unity contrast grating of the same spatial frequency as the test fringe.

Figure 6 shows the threshold for detecting introduction of an interference fringe in the presence of a unity contrast fringe of the same spatial frequency. The triangles represent thresholds when the two fringes are perpendicular to one another, and the octagons represent thresholds when the two fringes are so oriented as to produce a 10 c/deg distortion product. These results show that the observer's sensitivity to the

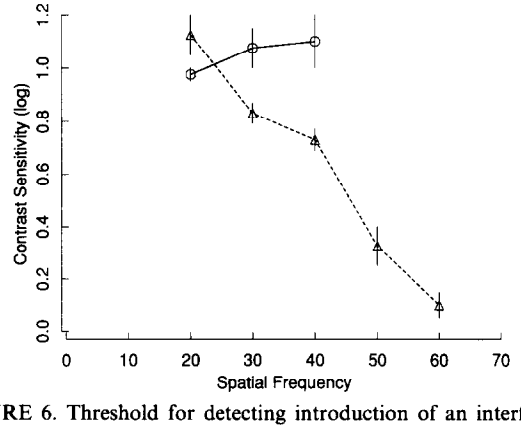

FIGURE 6. Threshold for detecting introduction of an interference fringe in the presence of a unity contrast fringe of the same spatial frequency. The triangles represent thresholds when the two fringes are perpendicular to one another, and the octagons represent thresholds when the two fringes are so oriented as to produce a 10 c/deg distortion product. These results show that the observer's sensitivity to the distortion product was greater than that to the interference fringe producing it whenever the spatial frequency of the interference fringe was 30 c/deg or more.

distortion product was greater than that to the interference fringe producing it whenever the spatial frequency of the interference fringe was 30 c/deg or more.

Masking of the distortion product. We chose to produce our distortion gratings by introducing slight differences of orientation instead of slight differences of spatial frequency because it allows both primary fringes to have the same spatial frequency and because the distortion product is roughly perpendicular to the fringes that create it. An advantage of the nearly orthogonal orientation of the distortion product relative to the primary fringes is that this arrangement minimizes masking of the distortion product by the primary fringes. We measured the spatial frequency dependence of such masking by measuring sensitivity to an interference fringe of 10 c/deg against a perpendicular grating of unity contrast and variable frequency. Observer WM, with checks on DM, showed a sensitivity loss of 0.2 log unit, when the masking fringe was 20 c/deg, that decreased gradually to nil at 50 c/deg. The three lowest frequency points for WM (triangles) in Fig. 4 have been corrected for this effect by increasing the plotted sensitivities by the amount of the estimated loss due to masking. This small correction has no appreciable influence on the curve that best fits the set of points.

Is the distortion product really a product? In the analysis of equations (4–8), the distortion product arises from a term representing the product of the intensity profiles of the two fringes. If this is correct, the amplitude of the distortion product should be proportional to the product of the fringe amplitudes. We wished to check this prediction as a test of the general theoretical framework.

Although the contrast of the fringe from interferometer B could not be reduced, its amplitude could be reduced simply by placing a graded neutral filter in the path common to its two beams. A second graded filter placed in the path common to all beams was used to keep the total space-average intensity of all beams at 2000 td. For three amplitudes of fringe B, we determined the amplitude (contrast, in this case) needed in fringe A to generate a threshold distortion product. Forced-choice methodology was not appropriate in this case, for fringe A now had to be set at a higher contrast than in the usual situation where it was accompanied by a unity contrast fringe of equal intensity, and at the fringe frequency of 40 c/deg used, the threshold for the fringe itself often lay below the threshold for the distortion product. Therefore, both fringes were steadily exposed, with fringe A at a value randomly chosen by the experimenter. After prolonged observation of the stimulus the subject (DM) gave a judgement as to whether a difference frequency grating was visible. Four of five presentations at each contrast defined a crude threshold with an uncertainty of about 0.1 log unit. These results (Fig. 7) conform well to the predicted product-of-contrasts relationship in this log–log plot, a line of negative unity slope, along which the product of the retinal contrasts of the two fringes is constant.

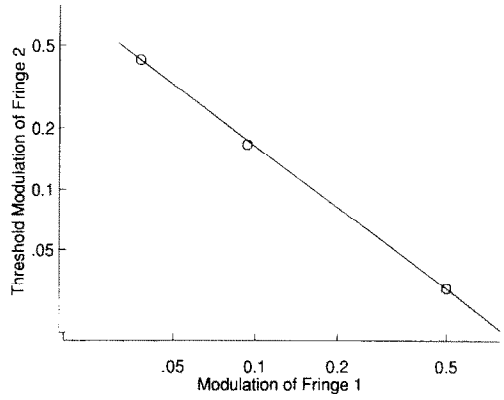

FIGURE 7. Threshold modulation of one interference fringe as a function of the modulation of a second simultaneously presented fringe for observer DM. The theoretical line with slope −1 is the prediction based on the notion that the difference frequency grating amplitude is proportional to the product of the amplitudes of the primary fringes.

Apparent mark–space ratio: one nonlinear site or many? As we have noted, the difference-frequency gratings generated by the visual nonlinearity appear to be made up of equally wide light and dark bars, unlike real gratings, in which the light bars appear wider than the dark ones (Pelli, 1986). This difference is of theoretical interest. The asymmetry between the bright and dark bars of a real grating has been plausibly attributed to a compressive nonlinearity like the one that generates the difference-frequency grating (Pelli, 1986). A real grating's representation may undergo nonlinear transformation at any stage of processing, whereas the distortion product is processed only by those stages following the initial nonlinear stage that generates it. The equality of the light and dark bars in the difference-frequency grating suggests that it does not encounter the nonlinearity or nonlinearities responsible for making the light bars of real gratings wider than the dark ones. The latter nonlinearity must therefore be imposed no later than the stage at which the difference-frequency grating itself is generated; presumably the same early local nonlinearity underlies both. The difference in appearance was documented by matching both light and dark bars in turn to a black standard bar of variable width in an adjacent, otherwise uniform companion field; with this procedure, any under- or over-estimate of the standard bar should cancel out when we consider the ratio of the widths obtained for the light and dark bars of the grating. At 2 c/deg and 12 times threshold contrasts, the ratio of the perceived widths of light and dark bars for DRW was 1.6 ± 0.04 SEM for a real grating and 0.8 ± 0.05 for a distortion product formed with 40 c/deg fringes.

Parafoveal Observations

Rationale and procedure

Difference frequency gratings are not generally seen over the entire stimulus field, but only within a roughly circular region centered on the line of sight. This region of visibility shrinks as the primary fringes are made finer. For example, for one observer (DM) viewing 40 c/deg

primaries both at unity contrast, the region had a radius of about 2.5 deg, but with 120 c/deg primaries, the radius was reduced to only about about 40′. Evidently the spatial resolution at the site of the nonlinearity decreases with increasing retinal eccentricity. To provide information about the spatial filtering preceding the nonlinearity at an eccentricity other than the fovea, we measured contrast sensitivity for a difference frequency grating at an eccentricity of 3.8° nasal retina. The conditions were similar to those used with central fixation in Fig. 4, except that the field size was increased from 1 to 2 deg and the spatial frequency of the distortion product was reduced from 10 to 3 c/deg. Both these changes were introduced to favor detectability of the distortion product by the more poorly resolving extrafoveal retina.

Results

The results from two observers are shown in Fig. 8 with best-fitting curves which were the squares of Gaussian functions. These results differ from those with central fixation (Fig. 4), both in over-all sensitivity and scaling factor for spatial frequency. The ratio in spatial scaling, estimated from the Gaussian fits, is a factor of 2.6 for DW and 3.0 for DM. Despite this, the loss of sensitivity with increasing fringe frequency remains small by ordinary standards.

Successively Presented Primary Fringes

Rationale and procedure

In the previous experiments, both primary fringes were presented simultaneously. Here we describe experiments in which a bright fringe primary was briefly flashed on the retina and a second primary was subsequently viewed at the same retinal location. By varying the spatial frequency of these successive primaries, one can measure the spatial spread of the light adaptation produced by the flashed grating. When this technique is applied to rod vision (MacLeod, Chen & Crognale, 1989), the results imply an absence of strictly local sensitivity regulation: no difference frequency grating could be generated by unresolved targets. For cone vision, we show that difference frequency gratings formed by the nonlinear interaction of a flashed adapting fringe with a subsequently presented test fringe are easily obtained at frequencies far above the resolution limit, just as they are in the case of simultaneously presented fringes discussed above.

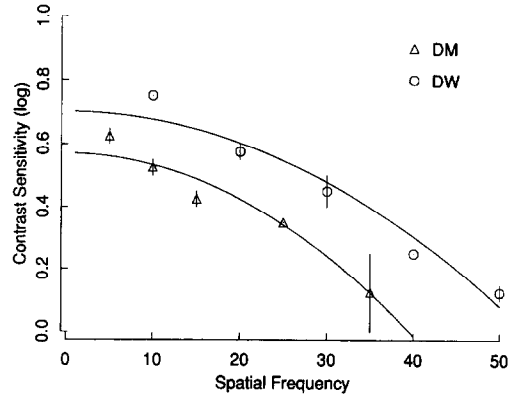

FIGURE 8. Contrast sensitivity for a difference frequency grating at an eccentricity of 3.8°, for two observers. The conditions were similar to those used with central fixation in Fig. 4, except that the field size was increased from 1 to 2 deg and the spatial frequency of the distortion product was reduced from 10 to 3 c/deg. Both these changes were introduced to favor detectability of the distortion product by the more poorly resolving extrafoveal retina. The curves represent least squares fits to the data with the square of a Gaussian. Vertical error bars represent ±1 times the standard error based on variation among sessions.

The large losses of visual sensitivity associated with visual pigment bleaching are not ordinarily due to reduction in the probability of absorbing light in the bleached pigment, since the fraction of pigment bleached is typically too small to account directly for the associated loss of sensitivity (Rushton, 1965). To discover whether strictly local processes implement the regulation of sensitivity associated with pigment bleaching, the fraction of pigment bleached should not be large enough to effect a significant local variation in the probability of photon absorption. We illuminated a 1 deg field for 10 msec with a grating of unity contrast and space-average power of 24 μW/deg^2, which we estimate bleached less than 7% of the cone visual pigment at the maxima of the fringe intensity. After each adapting flash, the adapted retina was steadily illuminated with a field that alternated between a homogeneous field and a test grating at fixed contrast. As before, the primary (adapting and test) gratings were identical in frequency but slightly different in orientation, so that the difference frequency grating was a vertical one of 10 c/deg. The observer indicated whether he could observe a difference-frequency pattern, and if so for how long. The procedure was repeated for different test contrasts and frequencies, randomly interleaved, until in total about 6 adapting flash presentations had been made for each frequency and test contrast. The space-average retinal illuminance was set at 100 td, because preliminary trials suggested that this was a good choice for maximizing the duration for which the difference-frequency remained visible. Nevertheless it was seldom visible for more than 10 sec; the threshold test contrast was the estimated contrast at which the difference-frequency persisted, on the average, for 2 sec after the adapting flash. A recovery period of 90 sec was allowed between adapting flashes.

Results

Distortion products were readily seen even at frequencies far above the resolution limit. They looked very much like the ones observed with superimposed gratings. That is, to most observers the dark bars appeared desaturated or greenish and displayed the "zebra stripes" (Williams, 1985a, 1988) associated with aliasing of the fine fringes by the receptor mosaic. That aliasing could be seen in the dark stripes of the distortion product instead of the light stripes has implications for the nature of the aftereffect that produces the distortion product in the successive presentation case, a point we will return to later.

The dependence of contrast sensitivity for the distortion product on the frequency of the fringes is shown in Fig. 9. Just as with superimposed fringes (Fig. 4) a visible difference-frequency grating can be elicited by fringes with frequencies as high as 100 c/deg or more.

Even with the intensity of the adapting flash set to three or four times the level normally used, we were not able to produce a difference frequency grating that remained visible for longer than about 10 sec. Since the visual pigment in the cones takes minutes to regenerate, this is evidence that depletion of the pigment was not the direct cause of the difference-frequency phenomenon.

A new complication in the successive presentation experiment is that the difference-frequency grating appeared to jump around continually, especially if the fringe frequency was very high. This is an expected consequence of involuntary eye movements during fixation. If the point of regard drifts over one full cycle of the test fringe, the difference-frequency grating drifts over one cycle of its own much larger spatial period, so that the velocity of the retinal excitation profile is increased in proportion to the ratio of the fringe frequency to the difference-frequency, making the effects of small eye movements quite noticeable to the observer.

The appearance of the afterimages of the unresolvable adapting fringes in total darkness was of some interest, since they offered a unique opportunity to observe a retinally stable excitation pattern from a stimulus of extremely high spatial frequency. They did not display the flicker usually visible in fringes in the aliasing frequency range, but did manifest clearly (and in a retinally stable arrangement) the zebra stripe patterns described by Williams (1985a, 1988), with comparable spatial characteristics at comparable fringe frequencies. Evidently, fixational eye movement has no critical role in the generation of the zebra stripes.

Finally, we note that "cubic" difference frequency gratings ($2f_1$–f_2), with a mistuned second harmonic stimulus, were observable in the bleaching experiment just as they were in the simultaneous presentation experiment.

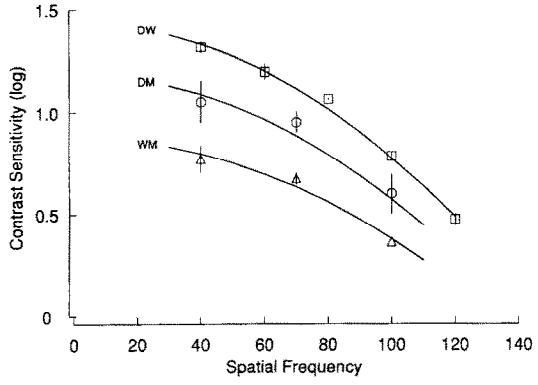

FIGURE 9. Contrast sensitivity for a distortion product observed when viewing a test fringe immediately following exposure to a bright adapting fringe flashed for 10 msec, for three observers. The data for DW is vertically displaced 0.3 log unit for clarity. Vertical error bars represent ±1 times the standard error based on variation among sessions, and are generally smaller than the symbols.

DISCUSSION

Comparison with other studies

A number of previous investigations have examined "beats" produced by superimposed gratings or equivalent procedures (Badcock & Derrington, 1989; Derrington & Badcock, 1985, 1986; Logvinenko, 1990; Nachmias & Rogowitz, 1983). These studies always employed primary fringes with relatively low spatial frequencies so that the "beats" were detected in the presence of these highly visible primaries. Our study differs from these in our use of primary gratings with much higher spatial frequency, which were often above the resolution limit. These studies also used incoherent light so that the primary grating contrast was limited by optical blurring in the eye, unlike the interference fringes we used. The luminance levels we could obtain were also substantially higher, which has the advantage of increasing contrast sensitivity to the distortion product (Makous *et al.*, 1991). These factors may also account for the failure to observe nonlinear distortion in some studies (Badcock & Derrington, 1989; Derrington & Badcock, 1986).

Sekiguchi, Williams and Packer (1991) described a phenomenon which they call "secondary zebra stripes". When viewing interference fringes with a spatial frequency of about 60 c/deg, some observers report a pattern of wavy stripes that resembles the zebra stripe pattern caused by foveal cone aliasing at twice the spatial frequency (about 120 c/deg). They provide a model incorporating sampling by the cone mosaic followed by an intensive nonlinearity that predicts the behavior of these secondary zebra stripes. It is likely that the nonlinearity responsible for this phenomenon is the same as that responsible for the distortion products studied here.

All our experiments give evidence that signals from interference fringes finer than the ordinary limit of visual resolution are remarkably well preserved at the nonlinear site. Figures 4, 8 and 9 show that as the spatial frequency of the fringe primaries increases, somewhat greater contrasts are required to produce a threshold distortion product. We assume that this increase in contrast is necessary to compensate for losses that precede the nonlinear stage producing the distortion grating. In related experiments, Burton (1973) found similar results, indicating little spatial summation for frequencies up to and somewhat beyond the resolution limit. Burton's difference-frequency gratings became substantially less visible than ours just above the resolution limit. We can only speculate on the reason for this difference between these data, but it is noteworthy that Burton's direct contrast sensitivity measurements for interference fringes are lower everywhere and much lower at high spatial frequencies than ours. Any factors that reduce fringe contrast or contrast sensitivity, especially in the high spatial frequencies [e.g. masking by noise in the interferometric field (Williams, 1985b) or mechanical vibration, to which interference fringe contrast can be very sensitive] are possible causes of this difference.

Comparison of the point spread function preceding the nonlinear site with the cone aperture

We now ask what point spread function would be required to account for the extremely shallow contrast sensitivity functions we observed. Transforming these data from the spatial frequency domain to the space domain facilitates their comparison with retinal anatomy, and the diameters of single cones in particular. Such a transformation is possible if we think of the signal arriving at the nonlinear site as integrating the fringe intensity over some region such as the cone cross-section. We will proceed on that basis, although as we note below the implicit model fails in important ways to capture the reality of photoreceptor optics.

If the cone has apertures described by cylinder functions and all had circular apertures of the same diameter, fringe contrast would be a first order Bessel function of spatial frequency. However, the entrance to an elongated structure only a few wavelengths of light in width, such as a cone, is not appropriately treated by geometric optics. Present evidence favors the view that cones are wave-guides (Enoch & Tobey, 1981). The entering light energy is concentrated on the waveguide axis and decreases gradually with distance from the axis according to a function that (for the lowest order mode) is better approximated by a Gaussian curve than by a cylinder function. Moreover, psychophysical results depend on the collective action of a population or ensemble of cones, the members of which cannot be absolutely identical. Then any variations of diameter, shape, or profile of cone apertures tend to change the shape of the transfer function for the ensemble towards that of a Gaussian curve. The modulus of the Fourier transform of a Gaussian curve is itself Gaussian. Therefore, we have fit the squares of Gaussian curves to the data in Figs 4, 8 and 9. The corresponding point spread functions can then be obtained by computing the Fourier transform of the Gaussian functions that fit the square root of these raw data. (As mentioned earlier, since both primary fringes are attenuated by the spatial filter, and the amplitude of the distortion product is proportional to the product of their amplitudes, the raw data reflect the *square* of the contrast attenuation.) Our data do not constrain the phase spectrum of the spatial filter, so we assume even symmetry in the underlying point spread functions.

TABLE 1

	Width at half height (arc sec)			
	Simultaneous presentation			Successive presentation
Observer	Fovea	Confidence limits	3.8°	Fovea
WM	16.0	13.3–18.5	—	14.4
DW	12.6	10.5–15.2	39.7	16.4
DM	11.9	2.1–16.6	46.3	15.9
Mean	13.5	8.0–19.0	43.0	15.5
Standard deviation	2.19		4.67	1.15
Standard error	1.27		3.3	0.67
Anatomical estimates	27.6		77.6	27.6

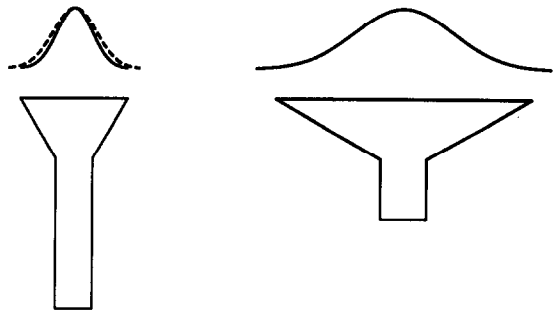

FIGURE 10. The profiles of the estimated Gaussian apertures are shown above a schematic diagram of a foveal cone on the left and a typical cone found at 3.8 deg in temporal retina on the right. The vertical scale is foreshortened greatly relative to the horizontal scale. The dashed profile is for the data obtained with successively presented gratings. Note that the estimated apertures are a constant proportion of the inner segment diameters rather than the outer segment diameters.

Table 1 summarizes the results in the space domain. Estimates are the full width at half height of Gaussian apertures in arc sec. Confidence limits (0.95) for the individual observers are based on the conservative assumption that minimum residual variance (i.e. residual variance from the best fit) is an estimate of random error. There were too few data from the experiments involving successively presented gratings and from the tests at 3.8 deg eccentricity to warrant estimating confidence limits, but the comparability of the two foveal results, and the comparability among individuals in all three conditions bolsters confidence in these estimates.

Both in the fovea and at 3.8 deg, the psychophysical estimates of the spatial filtering attributable to the cones yields Gaussian apertures with widths at half height on the order of half the diameter of cone inner segments. This is comparable to the size of modal patterns observed within cones (Enoch, 1961, 1963). These admit the same amount of light as a circular aperture with a diameter 1.21 times the Gaussian widths at half height listed in the table. Clearly, this result is not compatible with substantial spatial integration beyond the confines of a single cone. Anatomical estimates of the cone inner segment diameters obtained with Nomarski microscopy are shown in the last line of Table 1 (from Curcio, 1990, personal communication). The conversion to visual angle assumes 4.85 μm/min of arc.

Figure 10 illustrates these results graphically. The profile of the estimated Gaussian apertures are shown above a schematic diagram of a foveal cone and a cone such as is found 3.8 deg in temporal retina. The dashed profile is for the data obtained with successively presented gratings. The point spread functions of Table 1 and Fig. 10 suggest that the effective size of cones in these experiments is a compromise between the inner and outer segment diameters. But interpretation of that result, and generalization to other situations (particularly to incoherent illumination) requires an accurate conception of how light propagates through the photoreceptor matrix, and this we currently lack.

If each cone simply accumulated the incident flux across some effective aperture, the point spread

functions of Table 1 and Fig. 10 would specify the size of that aperture and would represent the cone's spatial weighting function for incident energy in any situation. We might think of the spread functions as representing the transmittance of a window in an opaque screen, presumably in the inner segment, with each cone absorbing all the light that passes through its window. In reality, however, the marked directional sensitivity of cones shows that not all light that enters the inner segment is absorbed in the cone that it first enters (Chen & Makous, 1989), and in this situation the relative influence of the inner segment aperture and the size of the outer segment where absorption takes place may depend upon the arrangement of the light stimulus. An extreme case will illustrate this. Consider an array of cones that have large inner segment apertures but that absorb light only in an axial "point sink" of negligible size in the outer segment. The two interfering wavefronts in our experiments are of equal amplitude, and since they are symmetrically arranged about the cone axis, their amplitudes remain equal at the site of absorption. As their relative phase varies continuously across the retina, there will be an alternation between regions where they reinforce each other and regions where they cancel perfectly, producing a fringe of full contrast. The interferometric contrast sensitivity at the cone level will then be maintained up to arbitrarily high spatial frequencies, no matter how large the cones' apertures may be, and our interferometrically derived "point spread functions" could underestimate the cone aperture for resolution in incoherent light. The irrelevance of the inner segment aperture in this case is exceptional, however: the inner segment dimensions are relevant when the outer segment absorbing region is extended. For example in the case where the outer segment absorbs all the energy transmitted through the inner segment aperture, the "opaque screen" model becomes applicable and the generality of the point spread function is guaranteed.

Even if all the transmitted energy is delivered to the outer segment for absorption, the cone aperture for axially incident light can not be quite as narrow as Table 1 and Fig. 10 suggest. Measurements of the increase in fundus reflectance after bleaching the foveal cone pigment (Norren & Kraats, 1989) indicate that at the wavelength of peak absorption 44% of the light axially incident on the retina is absorbed in cone pigment. The fraction of light absorbed during an axial passage through the outer segments is about 65% (Wyszecki & Stiles, 1982), so about $(0.44 \div 0.65)$, or 67% of the incident light must enter the outer segment. This is consistent with efficient collection of axially incident light over the entire area of the inner segments, which take up about two-thirds of the retinal area (Curcio, 1990, personal communication; Miller & Bernard, 1983), but our point spread functions allow for less than half that much absorption.

This discrepancy could be resolved if the difference between the refractive index inside and outside the cone increases gradually from the wide mouth of the inner segment toward the narrower end near the outer segment, as observed by Miller and Bernard (personal communication). In this case, obliquely incident light may leak more from the wider end than the narrower end. This would reduce the effective aperture of the cone for oblique light relative to its aperture for axially incident light. Recall that the obliquity of light incident at the retina increases with the spatial frequency of an interference fringe, simply because fringe frequency is proportional to the separation of two point sources in the entrance pupil. Thus it is possible that the high spatial frequencies required for our experiments revealed a smaller cone aperture, estimated in Table 1, than that available to collect axially-incident light.*

Note that the estimated apertures are a constant proportion of the inner segment diameters rather than the outer segment diameters, which grow with eccentricity at a much slower rate. This correlation has been extended to 30 deg eccentricity by Chen, Makous and Williams (1989), who found that the width at half height of the Gaussian point spread function was 48% of the inner segment diameter on average. Thus these results favor the inner segment as the light collecting organelle of the cone photoreceptor.

The fact that these point spread functions are as small as or smaller than cone inner segments does not require that the nonlinearity that generates these distortion products resides in individual cones. The nonlinearity could reside at some postreceptoral stage that consists of receptive fields fed by single cones.

Implications for additional sources of blurring before the nonlinear site

The fact that our estimates of the point spread function preceding the nonlinearity agree well with expectations about the optical properties of cones, and indeed are comfortably smaller than anatomical estimates of cone inner segments, leaves little or no room for a host of other potential sources of contrast loss, such as optical blurring including retinal scatter, temporal integration coupled with eye movements, and neural summation among cones prior to the nonlinear site. If any of these factors produced significant blurring, then our data would dictate an even smaller estimate of the cone aperture. Thus our estimate of 13.5 sec arc for oblique incidence (Table 1, width at half height) represents an upper bound on the size of the functional cone aperture. An even smaller cone aperture on the other hand is somewhat difficult to accept, because the functional aperture would then more closely resemble the outer segment, which would force us to abandon the notion of the inner segment as a light collector. The quantum efficiency of the cone mosaic would be

*The use of large angles of incidence, corresponding to displacements greater than 2.5 mm in the entrance pupil, is known to slightly reduce the contrast of interference fringes (Chen & Makous, 1989; Makous, 1977; Makous & Schnapf, 1973; Schnapf & Makous, 1974). Efforts to observe any such effects have shown negligible effects under the conditions of the present experiments, mainly because we used displacements that were <2.5 mm.

needlessly reduced, with many photons passing between photoreceptors, unavailable to the photopigment. And cone pigment absorption would be insufficient to account for the results from retinal densitometry discussed above. Our estimated point spread functions are therefore tightly constrained from both sides. We now consider each of the implications of these point spread functions in more detail.

Implications for laser interferometry and the optical quality of the eye. The correspondence between these point spread functions and cone inner segments provides the first empirical support for the theoretical claim that laser interferometry "bypasses" the optics of the eye, even at spatial frequencies as high as 130 c/deg. If interference fringes were progressively blurred with increasing spatial frequency, we would have obtained steeper contrast sensitivity curves for difference-frequency gratings. The contrast of laser interference fringes could in principle be degraded by forward scatter from the inner retina. But the shallow contrast sensitivity curves show that forward scatter in the inner retina is negligible. To the extent that the somewhat steeper curves we obtained in the parafovea reflect the increase in inner segment diameter with eccentricity, our measurements also leave little room for retinal scatter even outside the fovea where the inner retina is thicker. Ohzu and Enoch (1972) measured a modulation transfer function of the retina showing considerably more scattering than our data would allow. However, their measurements were on *post mortem* tissue which clouds rapidly after death. Furthermore, their measurements were made by transmission through the full retina and it is possible that some degradation of the image occurs after the light first has the opportunity to be absorbed by visual pigments. Gorrand (1989a,b) measured the optical quality of an aerial image of interference fringes formed on the retina, showing that it is higher in the fovea than in the extrafovea. The increased thickness of the inner retina in the extrafovea compared with the fovea could potentially scatter more light, explaining the difference. But there are other differences between the fundus reflections from fovea and extrafovea, such as might be caused by the relative numbers of rods and cones which have different waveguide properties, or differences in the amount of diffuse reflection from the choroid, so that scatter in the inner retina is not required by Gorrand's data.

Another way in which the contrast of interference fringes could be reduced is by scatter in the optics or diffuse scatter from the fundus. This would have the effect of casting a uniform veil of light over the whole retinal image (marred by speckle in the case of coherent imaging), and would lower the contrast of fringes at all spatial frequencies. However, measurements of veiling glare (Vos, 1962) suggest that this effect is small. We confirmed this, measuring the true stimulus contrast by casting a steady, vertical interference fringe of 0.5 c/deg within a field of 4 deg, and finding the contrast threshold of a 10 c/deg horizontal grating on the center of the bright stripes of the vertical background fringe. Then we increased the intensity of the background fringe until the 10 c/deg test field was just at threshold in the dark bars of the background field. The ratio of the field intensities necessary to produce the same threshold on the bright bars of one background fringe as on the dark bars of the other is taken as a measure of the amount of entopic stray light, which determines the retinal contrast of the 0.5 c/deg field. The ratios were 2% for one observer, and 2.4% for another. Thus, 0.5 c/deg fringes that were nominally of 100% contrast were actually more than 95% contrast at the site of absorption in the eye, and a portion of this small contrast loss may have been attributable to the apparatus. This small loss of contrast does not affect the conclusion that interference fringes are remarkably successful at avoiding optical blurring in the eye.

Implications for eye movements and temporal integration. In principle, eye movements coupled with the integration time of the visual system prior to the nonlinear site could smear or blur the primary fringe signals at the nonlinear site. This would affect high spatial frequency primaries more than the low ones, accounting for some of the fall-off in contrast sensitivity we obtained in Figs 4, 8 and 9. If this were a significant factor, then the true cone apertures would be even smaller than our already small estimates. However, recent measurements of the effects of eye movements on contrast sensitivity for single interference fringes show that they are less deleterious under these conditions than might be expected (Chen & Makous, 1990; Packer & Williams, 1990). Packer and Williams (1990) suggest essentially no difference in contrast sensitivity with 500 msec fringe presentation (when eye movements can smear) and with 1 msec flashes (when eye movements can have no effect) for the highest spatial frequencies they could measure (100 c/deg). Apparently, the eyes are frequently stable for periods long enough to allow glimpses of fringes unattenuated by eye movement smear. This is consistent with the appearance of high frequency single fringes (>60 c/deg) as well as the distortion products produced by two simultaneously viewed fringes, the contrast of which flickers or varies rapidly over time. The effects of such eye movements are further reduced by the recently discovered low integration time (10 msec) of the processes preceding the nonlinear stage (Chen & Makous, 1990). This does not preclude eye movement smear at lower intensities, when the integration time of the eye would be longer.

Implications for neural summation prior to the nonlinear site. Cone inner segments at a given eccentricity do not have uniform diameter (Ahnelt, Kolb & Pflug, 1987). Neural receptive fields could potentially vary much more in their size. Our foveal measurements were made with interference fringes of such high spatial frequency that it is difficult to believe that anything but the smallest receptive fields preceding the nonlinearity contributed to the task. If the nonlinear site resides at a stage beyond the receptors, our estimates of the preceding point spread function place an upper limit on the size of the smallest population of receptive fields in the fovea,

namely those fed by single cones. But these data do not exclude the existence of larger receptive fields fed by more than a single cone, which do not contribute in our task.

The strength of the claim that there is no neural summation prior to the nonlinear site rests on the assumption that aliasing does not obscure it. First, we consider an example that illustrates how aliasing could potentially obscure neural pooling of cone signals prior to the nonlinear site, and then we present an argument that rejects this possibility.

Fringes exceeding the foveal Nyquist frequency of 60 c/deg have aliases at a lower frequency. For example, in a perfectly regular 1-dimensional array the alias frequency equals twice the Nyquist frequency minus the input frequency in the input frequency range 0–120 c/deg. Consequently, as the input frequency increases above 60 c/deg, the alias actually declines in spatial frequency. (For a disordered lattice characteristic of the actual fovea, this alias would actually be a band of energy rather than a discrete component, but this does not materially affect the argument.) Signals and their aliases produce exactly the same distribution of photon catches in the photoreceptors so that the visual system beyond the sampling stage where aliasing is produced cannot distinguish between them. This means that as spatial frequency increases from 60 to 120 c/deg, the signal entering the low pass filter produced by neural pooling would actually decline in spatial frequency. In such a system, sensitivity to the distortion product could actually increase with increasing frequency of the external fringe primaries, and our conclusions that retinal scatter, eye movements, and neural pooling play a minimal role in determining the visibility of the distortion product must be evaluated with this in mind. Note, however, that a model in which aliasing is followed by appreciable neural pooling predicts a modulation transfer function that drops initially due to pooling, and then rises again for frequencies above the Nyquist frequency. Yet there is no such dip, nor even any shoulder or inflection in the data. Between 30 and 60 c/deg, aliasing can offer no relief from neural spatial integration, and in this range the data are remarkably flat, supporting the notion that neural pooling is small or nonexistent.

A second related argument is that if difference-frequency gratings for high frequency fringe primaries depended on transmission of the products of aliasing through a low pass spatial filter, they would appear with greatest contrast in an annular region where fringe spacing and receptor spacing are roughly matched (compare Williams, 1985a). Instead, as we have noted they are always most visible at the foveal center.

Even if aliasing has some influence, it can not invalidate our conclusion that the effective size of cones is as small as shown in Fig. 7 for the following reasons. Consider the distortion product sensitivity with the two primary spatial frequencies both either below the Nyquist frequency by some amount or equally far above it. For example if both were 40 or both were 80 c/deg, both pairs of fringes would be expected to produce identical signals at the receptor array (ignoring irregularity in the mosaic), because 40 and 80 c/deg are aliases of each other for a mosaic with a Nyquist frequency of 60 c/deg. Therefore, any differences in sensitivity observed between these two spatial frequencies cannot be due to neural processing, and must reflect instead the spatial filtering preceding generation of the photoreceptor signals. An assessment of the MTF based on sensitivity at 40 and 80 c/deg is clearly little different than that based on all the data.

Size and site of adaptation pools

One pervasive nonlinearity in vision is light adaptation. Our experiments with the successive presentation of fringes were begun with the aim of examining whether this regulation of visual sensitivity proceeds independently at different points in the visual field (as it must if the adaptive mechanism operates within functionally separate photoreceptors), or whether the adaptation mechanism itself has limited spatial resolution. There has been a persisting conflict between evidence from receptor electrophysiology in lower vertebrates on the one hand, and psychophysical observations and recordings from mammalian retinal ganglion cells on the other (MacLeod, 1978; Walraven, Enroth-Cugell, Hood, MacLeod & Schnapf, 1990).

Psychophysical and physiological experiments have mostly dealt with rod vision and have yielded convincing evidence that in primate retina there is no substantial local sensitivity regulation (Baylor, Nunn & Schnapf, 1984; MacLeod *et al.*, 1989; Rushton, 1965). For cones, the picture is less clear. The few psychophysical experiments that have been done (e.g. Cicerone, Hayhoe & MacLeod, 1990) indicate that cone adaptation pools, if they exist at all in man, are small enough to be consistent with an important role for the cone receptors in adaptation. Primate electrophysiology shows much less adaptation than is needed to account for psychophysically observed adjustments of sensitivity (Boynton & Whitten, 1970; Schnapf, Nunn, Meister & Baylor, 1992; Valeton & van Norren, 1983).

To the extent that the aftereffect we observed in our successive presentation experiments was a manifestation of light adaptation, our experiments show that some sensitivity regulation appears to proceed independently in different cone pathways. Cicerone *et al.* (1990) reported a similar result in parafovea, and their data suggested that this may also be true of the fovea. Burr, Ross and Morrone (1985) found that the pooling of adaptive gain changes approximately equalled that attributable to optical spread, and suggested that neural pooling might be much less. However, the use of laser interferometry to avoid losses of optical resolution, and the use of difference-frequency gratings to avoid later neural losses, together provide a uniquely precise assessment of the spatial resolution of the sensitivity-regulating site even in the fovea. The results of adapting with gratings and testing with similar gratings (Fig. 9) suggest that the only resolution losses affecting the cone

sensitivity-regulating mechanism are those due to the integration of absorbed light energy within the individual cone photoreceptor. This strictly local regulation of sensitivity of the cone system contrasts impressively with the result obtained for rod vision (MacLeod *et al.*, 1989), where there is no suggestion of any strictly local regulation of sensitivity at all.

Other psychophysical evidence shows clearly that photopic sensitivity regulation cannot always be strictly local. For instance, the phenomena of chromatic adaptation cannot be accounted for on the basis of independent changes in individual cone classes (see for example the review by Walraven *et al.*, 1990). However, the fine fringe stimuli in our experiments would hardly excite cells that pooled signals from multiple cones, whether synergistically or antagonistically: that they do not is what leads us to conclude that the processes we observe are restricted to single cone pathways. We were not able to extend our observations to low enough spatial frequencies to observe such spatial interactions. More recent work with this technique (Chen, Makous & Williams, 1988) suggests spatially antagonistic interactions both before and after the nonlinear stage.

Nevertheless, as Makous *et al.* (1991) show, the Weber Law change of sensitivity that occurs with exposure to uniform background light appears to be fully reflected in the visibility of difference-frequency gratings. This indicates that the primary sensitivity-regulating mechanisms of photopic vision are at stages no later than where the difference-frequency grating is generated, which in turn precedes spatial summation of signals from neighboring cones. If further adaptation follows neural spatial integration, its contribution to the Weber's Law adjustment must be minor

This is not to say that the sensitivity controlling mechanism must reside within the cone itself. Extracellular recordings reflecting cone receptor behavior (Valeton & van Norren, 1983) and photocurrent recordings from cone outer segments (Schnapf *et al.*, 1992) do show sensitivity changes within primate cones, but only at higher light levels than those at which Weber's Law begins to be observed psychophysically; and psychopysical evidence can be reconciled by invoking an adaptation mechanism at the cone-bipolar synapse (Belgum & Copenhagen, 1988), which would preserve local independence in sensitivity regulation but would not be reflected in the electrical recordings.

The presence of local sensitivity regulation in the cone system, and its absence in the rod system, make functional sense. An early adaptation locus is useful to protect the system from being overloaded. At scotopic levels, quantal fluctuations in the light stimulus prevent individual rod receptors from forming a usefully rapid and precise estimate of the ambient illumination level, but in photopic vision this problem does not arise, and it is feasible to regulate sensitivity independently for each cone.

We next consider the nature of the aftereffect that is generated by exposure to the bright flash of an interference fringe. Brief illumination of the retina could be followed by any of three kinds of aftereffects any one of which could produce a visible distortion product when the observer viewed the test grating: the addition of a signal in the form of a positive afterimage (Rushton, 1965); subtractive effects; and multiplicative gain changes (Adelson, 1982; Geisler, 1981; Hayhoe, Benimoff & Hood, 1987; Werblin, 1974). The additive and subtractive aftereffects are linear phenomena and by themselves would not generate a distortion product. But if a nonlinear stage such as that revealed by the experiments with simultaneous fringe presentation followed the generation of the aftereffects, distortion products could be produced. For example we can describe the sum of an additive aftereffect and the excitation profile produced by a subsequently viewed grating with the equation,

$$\xi(x,y) = I_m\{A_p N_1(f)\cos[2\pi f(x\sin\theta + y\cos\theta)] + B(t)N_1(f)\cos[2\pi f(-x\sin\theta + y\cos\theta)]\}, \quad (10)$$

which is identical to equation (3) except that the grating produced by the adapting flash has a time varying coefficient, $B(t)$, to denote its decaying positive aftereffect. Therefore, if this combined signal were to pass through the compressive nonlinearity revealed in the experiments with simultaneously present gratings, it would produce an identical distortion product except for the time varying coefficient. A subtractive aftereffect would produce a similar result, except with a half period phase shift. The distortion product then would be the negative of that produced by interaction with the positive afterimage, or an identical grating in reverse phase.

Any multiplicative gain changes are inherently nonlinear and will produce distortion products even without a second stage that is nonlinear. A multiplicative effect would correspond to viewing the test grating with a patch of retina that has sensitive (e.g. unadapted) and insensitive (adapted) strips alternating at the spatial frequency of the target but in a slightly different orientation. Thus, the effect of light in the test grating at any retinal locus would be modulated in inverse proportion to the amount of adapting light that has fallen at the locus, and the result contains the following product term:

$$c_2 A_p B(t) N_1^2(f)\cos[2\pi f(x\sin\theta + y\cos\theta)] \times \cos[2\pi f(-x\sin\theta + y\cos\theta)]. \quad (11)$$

This is identical to equation (5), except for the time varying coefficient, $B(t)$, and it produces the same distortion grating.

So, no matter which of these aftereffects is produced by the adapting flash, a distortion product could be produced with the same spatial frequency, but perhaps different amplitude and phase, as that produced by two simultaneously presented gratings. The amplitude and phase of the difference-frequency grating will depend on the kind of sensitivity change, and we will see next that the phase of the distortion product actually observed has implications for which kind of aftereffect was actually involved.

The perceived location of zebra stripes indicates where the contrast of the high frequency primaries is high. The strips where the zebra stripes cannot be seen correspond to locations where the aftereffect and the test grating tend to cancel each other, reducing the signal at the primary fringe frequency. Both subtractive and multiplicative aftereffects should have produced the highest contrast at the primary frequency in the light bars of the distortion product, not the dark bars where they are seen. An additive aftereffect, on the other hand, predicts zebra stripes in the dark bars, consistent with a view of adaptation in which bleached cones produce a persisting signal in the dark (Baylor & Hodgkin, 1974) that could reinforce the test fringe when they are in register (which produces a dark bar in the distortion product due to the compressive nonlinearity) or partly cancel the test fringe when they are out of register (which produces a light bar in the distortion product). Therefore, the evidence favors an additive process, which could produce adaptive effects through response compression.

REFERENCES

Adelson, E. H. (1982). Saturation and adaptation in the rod system. *Vision Research, 22*, 1299–1312.

Ahnelt, P. K., Kolb, H. & Pflug, R. (1987). Identification of a subtype of cone photoreceptor, likely to be blue sensitive, in the human retina. *Journal of Comparative Neurology, 255*, 18–34.

Badcock, D. R. & Derrington, A. M. (1989). Detecting the displacements of spatial beats: No role for distortion products. *Vision Research, 29*, 731–739.

Baylor, D. A. & Hodgkin, A. L. (1974). Changes in time scale and sensitivity in turtle photoreceptors. *Journal of Physiology, 242*, 729–758.

Baylor, D. A., Nunn, B. J. & Schnapf, J. L. (1984). The photocurrent, noise and spectral sensitivity of rods of the monkey *Macaca fascicularis*. *Journal of Physiology, London, 357*, 575–607.

Belgum, J. H. & Copenhagen, D. R. (1988). Synaptic transfer of rod signals to horizontal and bipolar cells in the retina of the toad. *Journal of Physiology, 96*, 225–245.

Boynton, R. M. & Whitten, D. N. (1970). Visual adaptation in monkey cones: Recordings of late receptor potentials. *Science, 170*, 1423–1426.

Burr, D. C., Ross, J. & Morrone, M. C. (1985) Local regulation of luminance gain. *Vision Research, 25*, 717–727.

Burton, G. J. (1973) Evidence for non-linear response processes in the human visual system from measurements on the thresholds of spatial beat frequencies. *Vision Research, 13*, 1211–1225.

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

Chen, B. & Makous, W. (1989). Light capture by human cones. *Journal of Physiology, London, 414*, 89–109.

Chen, B. & Makous, W. (1990). Temporal modulation at high spatial frequencies (abstract). *Investigative Ophthalmology and Visual Science (Suppl.), 31*, 428.

Chen, B., Makous, W. & Williams, D. R. (1988). Serial spatial filters in vision. *Investigative Ophthalmology and Visual Science, 29*, 58.

Chen, B., Makous, W. & Williams, D. R. (1989). A nonlinearity localized in the outer plexiform layer. *Investigative Ophthalmology and Visual Science (Suppl.), 30*, 52.

Cicerone, C. M., Hayhoe, M. M. & MacLeod, D. I. A. (1990). The spread of adaptation in human foveal and parafoveal cone vision. *Vision Research, 30*, 1603–1615.

Derrington, A. M. & Badcock, D. R. (1985). Separate detectors for simple and complex grating patterns? *Vision Research, 25*, 1869–1878.

Derrington, A. M. & Badcock, D. R. (1986). Detection of spatial beats: Non-linearity or contrast increment detection? *Vision Research, 26*, 343–348.

Enoch, J. M. (1961). Wave-guide modes in retinal receptors. *Science, 133*, 1353–1354.

Enoch, J. M. (1963). Optical properties of the retinal receptors. *Journal of the Optical Society of America, 53*, 71–85.

Enoch, J. M. & Tobey, F. L. (1981). *Vertebrate photoreceptor optics*. New York: Springer.

Geisler, W. S. (1981). Effects of bleaching and backgrounds on the flash response of the cone system. *Journal of Physiology, London, 312*, 413–434.

Gorrand, J. M. (1989a). Reflection characteristics of the human fovea assessed by reflectomodulometry. *Ophthalmic and Physiological Optics, 9*, 53–60.

Gorrand, J. M. & Bacin, F. (1989b). Use of reflecto-modulometry to study the optical quality of the inner retina. *Ophthalmic and Physiological Optics, 9*, 198–204.

Hayhoe, M. M., Benimoff, N. I. & Hood, D. C. (1987). The time-course of multiplicative and subtractive adaptation process. *Vision Research, 27*, 1981–1996.

Logvinenko, A. D. (1990). Nonlinear analysis of spatial vision using first-and-second-order Volterra transfer functions measurement. *Vision Research, 30*, 2031–2057.

MacLeod, D. I. A. (1978). Visual sensitivity. *Annual Review of Psychology, 29*, 613–645.

MacLeod, D. I. A., Chen, B. & Crognale, M. (1989). Spatial organization of sensitivity regulation in rod vision. *Vision Research, 29*, 965–978.

MacLeod, D. I. A., Williams, D. R. & Makous, W. (1985). Difference frequency gratings above the resolution limit. *Investigative Ophthalmology and Visual Science (Suppl.), 26*, 11.

Makous, W. (1977). Some functional properties of visual receptors and their optical implications. *Journal of the Optical Society of America, 67*, 1362.

Makous, W. & Schnapf, J. L. (1973). Two components of the Stiles–Crawford effect: Cone aperture and disarray. *Proceedings of the Association for Research in Vision and Ophthalmology*, 88.

Makous, W., Williams, D. R. & MacLeod, D. I. A. (1991). *Early visual nonlinearity and adaptation*. In preparation.

Miller, W. H. & Bernard, G. D. (1983). Averaging over the foveal receptor aperture curtails aliasing. *Vision Research, 23*, 1365 1369.

Nachmias, J. & Rogowitz, B. E. (1983). Masking by spatially modulated gratings. *Vision Research, 23*, 1621–1630.

Norren, D. & Kraats, J. (1989). Retinal densitometer with the size of a fundus camera. *Vision Research, 29*, 369–374.

Ohzu, H. & Enoch, J. M. (1972). Optical modulation by the isolated human fovea. *Vision Research, 12*, 245–251.

Packer, O. & Williams, D. R. (1990). Eye movements and visual resolution. *Investigative Ophthalmology and Visual Science (Suppl.), 31*, 494.

Pelli, D. G. (1986). Bright areas appear magnified: A visual saturation? *Investigative Ophthalmology and Visual Science (Suppl.), 27*, 226.

Rushton, W. A. H. (1965). The Ferrier Lecture. *Proceedings of the Royal Society of London, B162*, 20–46.

Schnapf, J. L. & Makous, W. (1974). Individually adaptive optical channels in human retina. *Proceedings of the Association for Research in Vision and Ophthalmology*, 26.

Schnapf, J. L., Nunn, B. J., Meister, M. & Baylor, D. A. (1992). Visual transduction in cones of the monkey, *Macaca fascicularis*. *Journal of Physiology, London*. In press.

Sekiguchi, N., Williams, D. R. & Packer, O. (1991). Nonlinear distortion of gratings at the foveal resolution limit. *Vision Research, 31*, 815–831.

Valeton, J. M. & van Norren, D. (1983). Light adaptation of primate cones: An analysis based on extracellular data. *Vision Research, 23*, 1539–1547.

Vos, J. J. (1962). *On mechanisms of glare*. Soesterberg, Netherlands: Institute for Perception.

Walraven, J., Enroth-Cugell, C., Hood, D. C., MacLeod, D. I. A. & Schnapf, J. L. (1990). The control of visual sensitivity: Receptoral

and postreceptoral processes. In Spillman, L. & Werner, J. S. (Eds), *Visual perception: The neurophysiological foundations* (pp. 53–101). San Diego: Academic Press.

Watson, A. B. & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics, 33*, 113–120.

Werblin, F. S. (1974). Control of retinal sensitivity. II. Lateral interactions at the outer plexiform layer. *Journal of General Physiology, 63*, 62–87.

Williams, D. R. (1985a). Aliasing in human foveal vision. *Vision Research, 25*, 195–205.

Williams, D. R. (1985b). Visibility of interference fringes near the resolution limit. *Journal of the Optical Society of America, A2*, 1087–1093.

Williams, D. R. (1988). Topography of the foveal cone mosaic in the living human eye. *Vision Research, 28*, 433–454.

Wyszecki, G. & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae*. New York: Wiley.

Acknowledgements—We wish to thank Bing Chen and Bill Haake for their help in the production of the Figures for this paper. Supported by AFOSR 85-0019, EY01319, EY04367, EY01711 and EY04885.