Pergamon

PII: S0042-6989(97)00036-9

Vision Res., Vol. 37, No. 15, pp. 2143–2152, 1997

Printed in Great Britain
0042-6989/97 $17.00 + 0.00

Computing Stereo Channels from Masking Data

A. GLENNERSTER,*† A. J. PARKER*

Received 7 August 1996; in revised form 18 December 1996

The detection of stereoscopic depth in random-dot patterns that have been spatially band-pass filtered is adversely affected by the addition of noise at spatial frequencies in the neighbourhood of the frequencies present in the stereogram. This elevation of threshold is generally termed *masking* and recent data have been interpreted as evidence for a pair of spatial-frequency-tuned stereo "channels" whose peak spatial frequencies are either at 3 and 5 c/deg or 2.5 and 7 c/deg. This interpretation was re-examined. In particular, we have studied how the characteristics of masking interactions might be affected by taking account of the presence of an initial modulation transfer function (including the optical m.t.f. of the eye) that precedes the stage at which signal and mask interact. Using this approach, we reach the conclusion that the peak of the internal masking function for stereo detection coincides with the signal spatial frequency over the whole range tested (1.7–11.6 c/deg). We conclude that the recent data of Yang and Blake [(1991). *Vision Research, 31,* 1177–1189] are consistent with a multiple channel model in much the form proposed by Julesz and Miller [(1975). *Perception, 4,* 125–143]. The analysis presented in this paper has general implications for the interpretation of masking studies in spatial contrast vision. © 1997 Elsevier Science Ltd.

Masking Binocular stereopsis Modulation transfer function

INTRODUCTION

The way in which one visual stimulus may elevate the threshold for another is often called "masking". This typically refers to a psychophysical experiment in which the threshold for a stimulus is measured with and without the presence of a second fixed stimulus. If the second fixed stimulus is well above its own threshold, it often interferes with the detection of the first stimulus. This interference is referred to as *masking* and the second fixed stimulus is called a *mask.* Note that this is a procedural definition of masking that does not include any hypotheses about why, where or how the interactions between the two stimuli take place.

This paper is mainly about the interpretation of masking experiments in stereoscopic vision, although we will attempt to show that the conclusions reached here about stereo vision have much wider implications. The reason for pursuing this question is that the presence of a masking interaction between two visual stimuli has been widely interpreted in terms of the two stimuli interacting within a single visual channel. Thus a strong interaction (large elevation of detection threshold) is taken to imply that the masking stimulus is strongly exciting the visual channel most suited to the detection of the test stimulus, and is thereby preventing the detection of the test stimulus. Weaker interactions imply only a partial overlap between the masking stimulus and the channel used to detect the test stimulus. Thus, the pattern of threshold changes that is generated as the mask is altered along a particular stimulus dimension probes the qualities of the visual channel in that dimension.

In early visual processing, most dimensions are not served by a single visual channel. Stimulus wavelength, spatial frequency and orientation are all examples of this point. In general, the number and shapes of visual channels are unknown. Indeed, this information is often sought from masking experiments. Interpreting the patterns of interactions between test and masking stimuli is therefore not straightforward. In this paper, we examine data on the spatial frequency selectivity of binocular stereoscopic mechanisms. Our original interest in this question was to examine whether the evidence from masking experiments leads inevitably to the conclusion that there is a specialised set of spatial-frequency-selective detectors exclusive to binocular stereopsis. Specifically, we examine the shapes of masking functions for binocular stereoscopic vision after taking into account the possibility that the stage of binocular interaction is preceded by a more general attenuation or transfer function, which is tuned for spatial frequency like the human contrast sensitivity function. The reasoning has wider implications for the interpretation of masking results in the spatial frequency domain.

SUMMARY OF PREVIOUS WORK

Some of the best evidence that binocular processing may occur independently at different spatial frequencies

*University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, U.K.

†To whom all correspondence should be addressed [*Fax*: 01865-272488; *Email*: ag@physiol.ox.ac.uk].

has come from psychophysical masking studies. Originally, Julesz & Miller (1975) used a masking paradigm to suggest that there must be many independent spatial-frequency-tuned channels for binocular stereopsis and rivalry. This conclusion fitted in with the view developed 30 years ago that early visual processing might, in general, be served by a set of independent spatial frequency channels. In this view, the channels used for stereo vision would simply be suitable binocular combinations of generalised early visual channels that might underly all forms of our pattern vision.

More recently, Yang & Blake (1991) have obtained a much more extensive set of measurements of the masking interactions as a function of spatial frequency. By using a detection criterion based on the visibility of stereoscopic depth, they hoped to isolate the channels involved in binocular stereopsis. It should be mentioned at once that the experimental procedure used by Yang and Blake does not guarantee to isolate the channels exclusive for stereoscopic vision. In their procedure (and that of Julesz & Miller, 1975), the test stimulus, whose depth is to be detected, was presented binocularly, but the masking stimulus, which is intended to interfere with the depth detection task, was presented only to one eye. Thus, the interaction that determines the visibility of the test stimulus in depth may actually depend upon the interactions between test and mask patterns at a completely monocular stage of processing. Such interactions at a monocular stage may determine the detectability of the relevant test stimulus in the neural signal from one eye. The measurements in the binocular task may reflect no more than the logical point that binocular stereo demands that two visual stimuli must be detectable, one from each eye. The use of the binocular task clearly means that the relevant signals must reach a binocular stage somewhere in the chain of neural processing but, in this particular form, the experimental procedure does not guarantee that the interaction between the test and the mask takes place at a binocular stage of processing.

From their results, Yang & Blake (1991) reached the conclusion that there may be only two channels, with peak sensitivities at 3 and 5 c/deg. As noted above, this would involve a considerable revision of earlier views. Tyler *et al.* (1994) analysed Yang and Blake's data and also concluded that there may only be two channels, although their estimate of the peak frequencies was slightly different (2.5 and 7 c/deg). As Tyler *et al.* went to some lengths to exploit mathematical modelling techniques to give an objective description of Yang and Blake's data, the two papers seem to make a strong case. Perhaps, therefore, there is a limited set of channels exclusive to stereo and having a markedly different range of spatial frequency sensitivity than most other channels in early visual processing. In this paper, we re-analyse the data of Yang & Blake (1991) and show here that they can be brought into agreement with the type of multi-channel model proposed by Julesz & Miller (1975).

Both Julesz & Miller (1975) and Yang & Blake (1991) used random-dot stereograms with 50% dark and light dots. The stereograms were passed through a frequency-domain filter that was rectangular and band-pass in spatial frequency but was isotropic and had no selectivity for orientation. A masking stimulus was prepared by filtering a sample of the random dots in the same way. This was added to one eye's image. For the purposes of explanation and for consistency with previous work, we will refer to this masking stimulus as "noise" at various points in the paper, even though this pattern contains no random elements over time.

Julesz and Miller used a series of demonstration stereograms to show that, if the masking noise had a centre frequency two octaves away from that of the signal, it was much less effective in disrupting stereopsis than if the noise had the same spectrum as the signal. They concluded that stereopsis (and binocular rivalry) must occur independently in separate spatial-frequency-tuned channels. An important feature of early masking studies (e.g. Stromeyer & Julesz, 1972) is that they considered the effect of the modulation transfer function (m.t.f.) of the early visual system on the mask at different spatial frequencies. They took the view that the threshold contrast sensitivity function reflects a general linear attenuation function that reduces the signal in the low and high spatial frequency regions of all visible targets, regardless of whether they are at threshold or well above threshold. As will be discussed later, this view is distinctly different from the conclusions reached from experiments where observers have been asked to equate the perceived contrast of two gratings of different spatial frequencies (Georgeson & Sullivan, 1975).

Yang & Blake (1991) studied a much wider range of signal and mask spatial frequencies than Julesz & Miller (1975). For each signal/noise combination, they measured the threshold level of noise that was required to disrupt stereopsis.* An example of their data is shown in Fig. 1. There are two important features to note. First, the peaks of the masking functions do not line up with the signal spatial frequency. Rather, the most effective mask tends always to be close to 3–5 c/deg. Second, the position of the curves on the *y*-axis depends upon the spatial frequency of the signal. The signal is most effective relative to the masking noise when the signal spatial frequency is close to 3–5 c/deg, regardless of the variations due to mask spatial frequency. These two

*In the Yang and Blake experiments, the signal contrast was nominally kept constant at the rather high value of 0.27 (r.m.s. contrast) for all test stimuli. The addition of noise inevitably demands that even higher r.m.s. contrasts should be accurately represented. This must have introduced distortions into the images, particularly at high noise contrasts, owing to the fact that local excursions in the image must be truncated at the minimum and maximum possible luminance levels. These truncations have the overall effect of reducing the energy in both the signal and the noise bands and introducing energy at other spatial frequencies. We have measured the extent of these effects for 8-bit stimuli similar to those used by Yang and Blake and find that the change in the ratio of energy in the signal and noise bands caused by truncation is minimal over most of the range in which thresholds were collected. On this basis, we ignored the effects of truncation in the subsequent analysis.

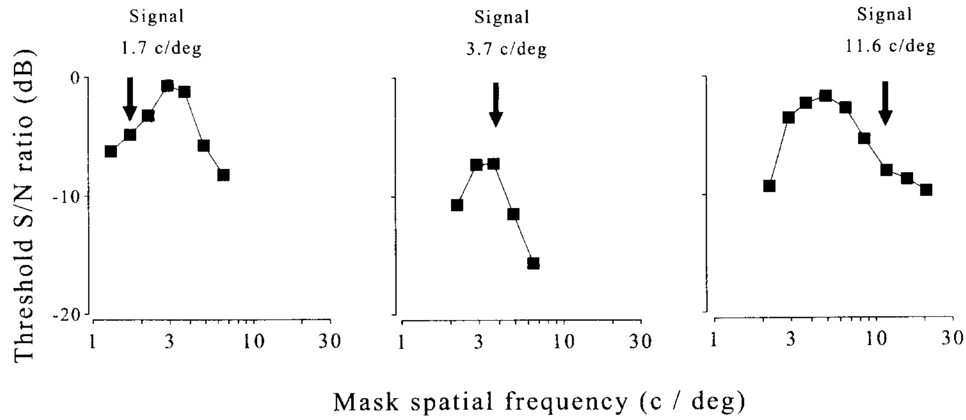

FIGURE 1. Three representative masking functions are shown from the study by Yang & Blake (1991). Threshold signal-to-noise ratios are plotted against the centre spatial frequency of the mask for three different signal frequencies. Signal r.m.s. contrast was fixed at 0.27. The data are from Fig. 3 (subject YY of Yang & Blake (1991)). Target disparity was 8 arcmin.

effects might have the same cause, i.e., a greater sensitivity of the visual system to spatial contrast patterns of about 3–5 c/deg.

IMPORTANCE OF THE INITIAL TRANSFER FUNCTION (I.T.F.)

In essence, the approach we adopt here asserts that those features of the stereoscopic masking data that imply a greater sensitivity to patterns of around 3–5 c/deg really reflect the operation of the well-known modulation transfer function for spatial contrast. The rest of the paper re-examines the modelling of Yang and Blake's data. Julesz & Miller (1975) also emphasised the importance of considering the modulation transfer function of the eye when deducing the shape of any putative "stereo channels". For a strictly linear description of early vision, it should be possible to describe a transfer function (i.e., the relative attenuation of each spatial frequency independent of whether it carries signal or noise) for the optics of the eye, the retina and so on up to the stage at which signal and noise interact. The assumption of linearity is in general wrong, but the aim here is simply to provide a coherent description of masking data.

We have chosen to treat the various stages of early vision by means of a single, overall "initial transfer function" (i.t.f.). In reality, this function depends on spatial position on the retina, the size of the visual stimuli and a number of other factors (Graham, 1989). But for a single class of stimuli presented at the same retinal location, a single estimate of the i.t.f. may be sufficient to predict the effective contrast of the masking stimuli. Apart from broadening the definition of the m.t.f., our model in this paper is much as Julesz & Miller (1975) originally described.

Ideally, then, we need an estimate of the i.t.f. for band-pass filtered random-dot patterns before we can assess their relative effectiveness as a signal or mask at different spatial frequencies. Estimates of the contrast sensitivity function for 2-D patterns vary (e.g. Koenderink & van Doorn, 1974; Mitchell, 1976; Smallman & MacLeod, 1994), although the shape is generally more low-pass than for sinewave gratings. The most relevant measure of the initial transfer function for modelling Yang and Blake's data would be an experimental measure of the c.s.f. for their stimuli collected from the same apparatus and using the same experimental observers. Unfortunately, this is not available (Yang and Blake, personal communication). In the next section, we adopt the next best alternative of deriving an estimate of the initial transfer function from Yang and Blake's published masking data. This estimate is then compared with previous data (Fig. 6).

ANALYTIC PROCEDURES

Eq. (1)–(4) set out a simple description of masking and its effects on visual thresholds. Suppose that the threshold for detection of a signal in the presence of a mask is reached when the "internal" signal-to-noise ratio reaches some critical value, k, which is constant across all conditions:

$$k = \frac{c'_s}{c'_n}, \tag{1}$$

where c'_s and c'_n are the effective contrasts of the signal and masking noise after they have passed through some linear attenuation function, G. In general, we can write:

$$c'_s = c_s G_i(f_s) \tag{2}$$

$$c'_n = c_n G_i(f_n) \tag{3}$$

hence,

$$k = \frac{c_s}{c_n}\frac{G_i(f_s)}{G_i(f_n)} \tag{4}$$

where c_s and c_n are the physical contrasts of the signal and mask respectively, f_s and f_n are their centre spatial frequencies. There may be several attenuation functions, or masking functions, $(G_1, G_2, \ldots)$ with peak sensitivities at different spatial frequencies. With these assumptions, the threshold signal-to-noise ratio (c_s/c_n) gives a direct

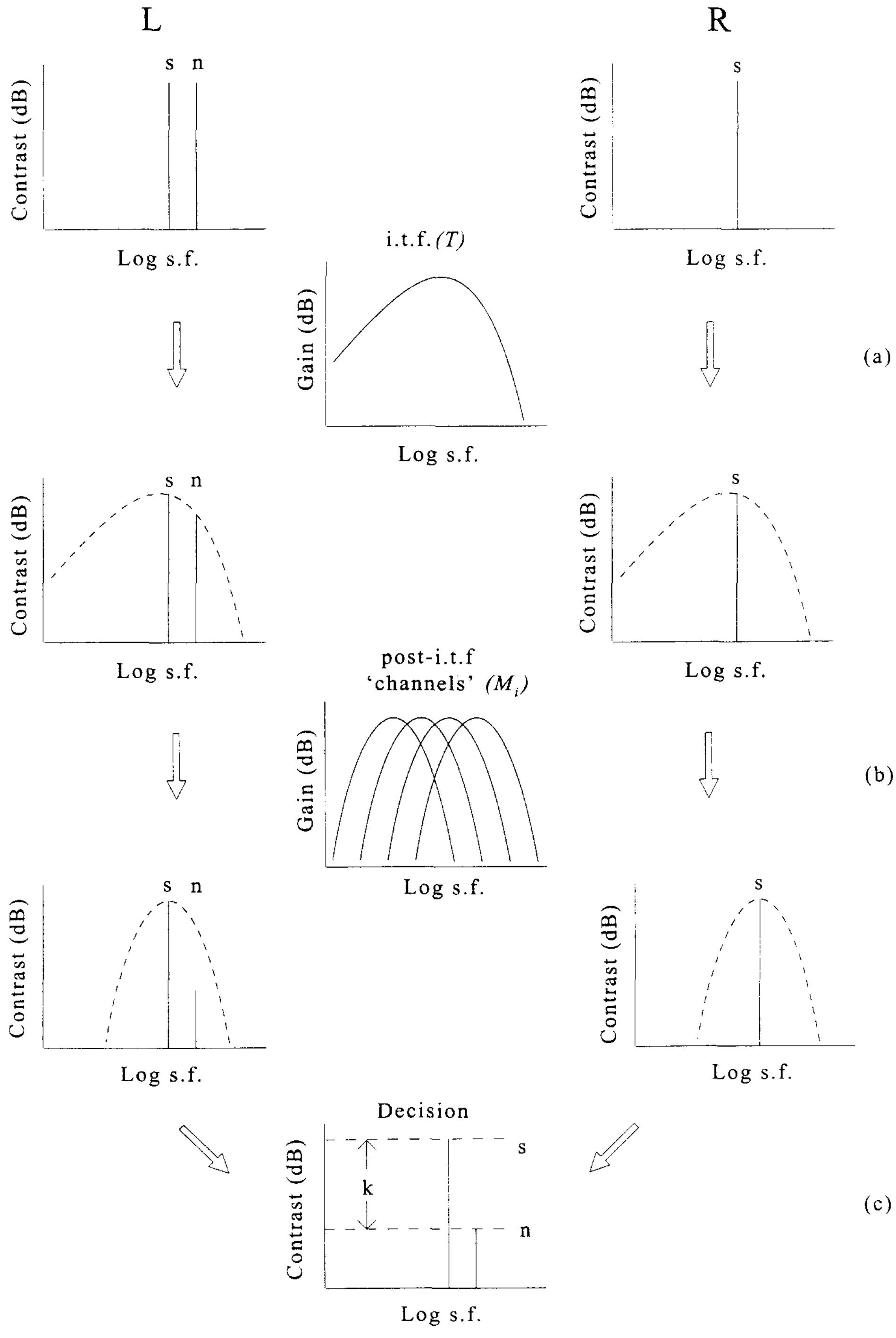

FIGURE 2. According to the model, the stimulus passes through two stages in series prior to a decision rule. (a) First, the stimulus in each eye passes through its own initial transfer function (i.t.f.) which is assumed to be the same in both eyes. (b) The stimulus is then processed by a set of parallel "channels" or, more generally, "post-i.t.f. masking functions" whose shape describes the attenuation of the mask relative to that of the signal at this stage. These may be monocular (as indicated here), binocular or both. (c) Finally, the amplitude of the signal and noise are compared and the signal detected if its amplitude exceeds that of the noise by a certain factor, k.

measure of the ratio of sensitivities of the attenuation function G_i at the mask and signal frequencies:

$$\frac{c_s}{c_n} = k\frac{G_i(f_n)}{G_i(f_s)} \tag{5}$$

and the set of thresholds obtained when signal frequency is kept constant and mask frequency is varied (often called a "masking function" or "tuning curve") will trace out the shape of G_i. This function does not necessarily reflect the shape of a "channel", where a channel is understood to be a single independent mechanism internal to the nervous system. For example, Tyler *et al.* (1994) point out that, if there is more than one "channel" operating in parallel (and overlapping in the

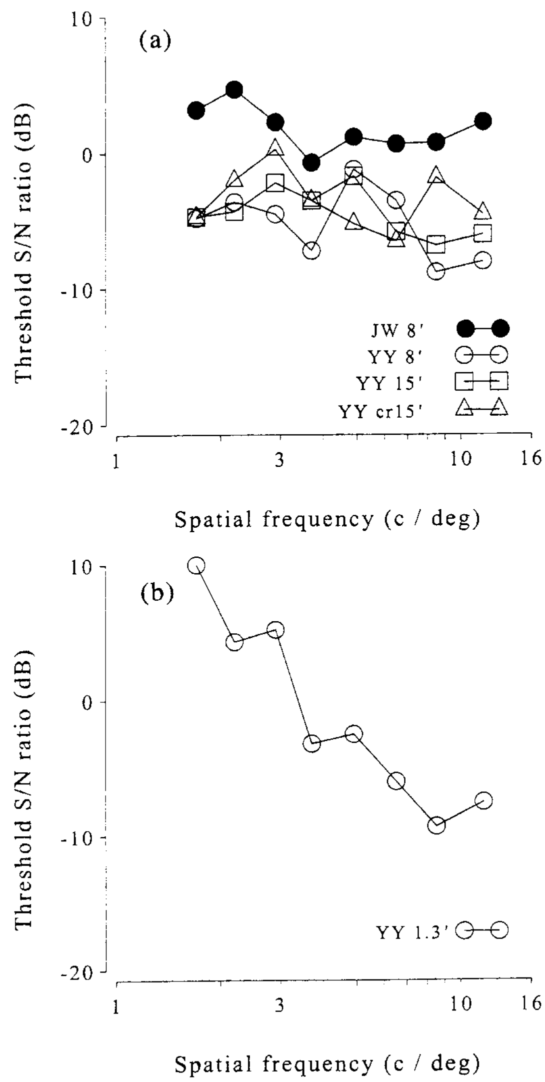

FIGURE 3. Data, again from Yang & Blake (1991), showing threshold signal-to-noise ratios for stimuli in which the centre frequency of the signal and mask were the same. (a) Open symbols show data for one subject for targets with a disparity of 8 and 15 arcmin (crossed and uncrossed). Closed symbols show data for a different subject (target disparity 8 arcmin). There is no consistent trend with spatial frequency. Subject JW requires a higher signal-to-noise ratio to detect the target than subject YY. (b) When the target disparity is relatively small (1.3 arcmin), on the other hand, there is a clear trend with spatial frequency: low spatial frequency targets are more easily masked.

frequency domain), different parts of the tuning curve may arise from different channels.

Here we address the possibility that, if there is a set of parallel channels (whose shape and tuning are unknown and need eventually to be deduced), this set of channels is preceded by a pre-filter, whose sensitivity to contrast varies as a function of spatial frequency. We call this pre-filter the i.t.f.. Once the shape of the i.t.f. is established, we can then proceed to the derivation of a set of "internal" or "post-i.t.f." masking functions.

How these internal masking functions should be interpreted remains an open question. It is uncertain whether the shapes of the internal masking functions are taken to correspond directly to the shape of "channels". It also remains unclear whether the masking interactions are taking place at a monocular or a binocular stage in visual processing, or possibly at more than one level of processing.

Suppose, then, that the attenuation function G_i is composed of two separate attenuation functions in series: (i) an overall initial transfer function (i.t.f.) illustrated in Fig. 2(a); and (ii) a set of parallel internal or "post-i.t.f." masking functions [see Fig. 2(b)]. Thus, for a given signal and mask centre spatial frequency:

$$c_s' = c_s T(f_s) M_i(f_s) \qquad (6)$$

$$c_n' = c_n T(f_n) M_i(f_n) \qquad (7)$$

where T is the initial transfer function (i.t.f.) and M_i describes the appropriate internal masking function. Again the ratio c_s'/c_n' is assumed to be constant when the signal is at detection threshold [see Fig. 2(c)], i.e.:

$$k = \frac{c_s'}{c_n'} = \frac{c_s}{c_n} \frac{T(f_s)}{T(f_n)} \frac{M_i(f_s)}{M_i(f_n)}. \qquad (8)$$

Both Eq. (4) and Eq. (8) predict that, when the signal and mask have the same centre spatial frequency, the contrast threshold for the detection of the stereo signal should be constant. Figure 3(a) shows that, within Yang and Blake's data, this prediction holds for stereo targets of 8 and 15 arcmin disparity: for one subject, thresholds are approximately constant at a signal-to-noise ratio of about 2 dB; for the other subject, the ratio is about −4 dB.

Yang and Blake's data for targets with 1.3 arcmin disparity do not follow the same pattern: thresholds for low frequency stimuli tend to be higher [Fig. 3(b)]. In other words, small disparities are more difficult to detect (more easily masked) when the signal spatial frequency is low. This result is not unexpected in the light of previous data on stereoacuity, which is poorer at low spatial frequencies (Legge & Gu, 1989) and contrast sensitivity for stereopsis, which is also poorer at low spatial frequencies when the target disparity is small (Smallman & MacLeod, 1994).* Initially, therefore, we concentrate on Yang and Blake's data for 8 and 15 arcmin of disparity.

DERIVING AN INITIAL TRANSFER FUNCTION (I.T.F.) FROM THE DATA

According to the approach set out above, thresholds for conditions in which the signal and mask are of the same centre spatial frequency provide no information about the shapes of the "channels" or attenuation functions through which they pass. This is because both signal and noise would be attenuated to the same extent. When the signal and mask have different spatial frequencies, however,

*Eq. (1) would have to be modified to account for the data on small disparity stimuli (so that a larger internal signal-to-noise ratio, k, is required to detect low spatial frequency targets). With this modification, a similar analysis can be applied (see Results and Fig. 9).

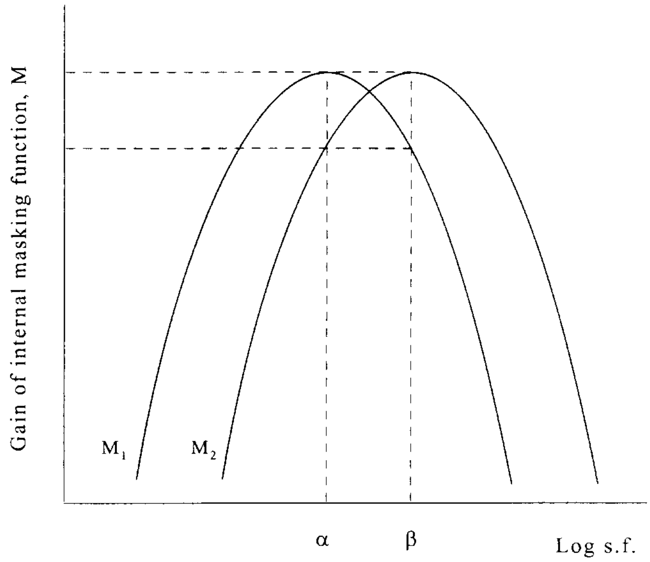

FIGURE 4. Two internal masking functions are shown, M_1 and M_2, whose peaks lie at spatial frequencies α and β. The dashed lines illustrate Eq. (11) and Eq. (12) in the text, which are used in the derivation of the the i.t.f.

they may be attenuated by different amounts and, in this case, thresholds are predicted to vary. The expected variation of thresholds depends on the assumptions of the underlying "channels" or attenuation functions.

Consider a pair of stimuli, both of which contain frequency bands centred on spatial frequencies α and β, but in one stimulus α is the centre frequency of the signal band and β is the centre frequency of the noise mask and for the other stimulus the situation is reversed. Using Eq. (8), when $f_s = \alpha$ and $f_n = \beta$:

$$k = \frac{c'_{s\alpha}}{c'_{n\beta}} = \frac{c_{s\alpha}}{c_{n\beta}} \frac{T(\alpha)}{T(\beta)} \frac{M_1(\alpha)}{M_1(\beta)}; \tag{9}$$

where $c_{s\alpha}/c_{n\beta}$ is the signal-to-noise threshold and M_1 describes the appropriate masking function for $f_s = \alpha$. On the other hand, when $f_s = \beta$ and $f_n = \alpha$:

$$k = \frac{c'_{s\beta}}{c'_{n\alpha}} = \frac{c_{s\beta}}{c_{n\alpha}} \frac{T(\beta)}{T(\alpha)} \frac{M_2(\beta)}{M_2(\alpha)}, \tag{10}$$

where, in this case, the signal-to-noise threshold is $c_{s\beta}/c_{n\alpha}$ and M_2 describes the appropriate masking function for $f_s = \beta$.

There is a way to estimate T, independently of the exact shape and spatial bandwidth of any subsequent internal masking functions, M_i. If one is prepared to assume that the shape of the internal masking functions, M_i, are symmetrical and of equal bandwidth when plotted on an abscissa of log spatial frequency, then some estimates of the T can be derived. The assumption that M_i is roughly symmetrical on a log frequency axis agrees with psychophysical estimates of the shapes of monocular spatial frequency channels (e.g. Blakemore & Campbell, 1969; Stromeyer & Julesz, 1972).

We also assume that there are two internal masking functions, M_1 and M_2, that have their peaks at α and β respectively (see Fig. 4), and that the gain of the functions at their peak is the same, i.e.

$$M_1(\alpha) = M_2(\beta). \tag{11}$$

The assumption that the internal masking functions, M_i, are symmetrical (when plotted on log axes) means that $M_1(\beta)$ and $M_2(\alpha)$ are reduced compared with the peak by an equal amount:

$$M_1(\beta) = M_2(\alpha). \tag{12}$$

So, for the case in which the i.t.f.(T) varies with spatial frequency, and with the assumptions outlined above [Eq. (11) and Eq. (12)], a comparison of the thresholds $c_{s\alpha}/c_{n\beta}$ and $c_{s\beta}/c_{n\alpha}$ gives information solely about the i.t.f.(T). From Eq. (9), Eq. (10), Eq. (11) and Eq. (12):

$$\frac{c_{s\alpha}/c_{n\beta}}{c_{s\beta}/c_{n\alpha}} = \frac{T(\beta)^2}{T(\alpha)^2} \tag{13}$$

where $c_{s\alpha}/c_{n\beta}$ is the signal-to-noise threshold when the signal has centre spatial frequency α and the mask has centre spatial frequency β, $c_{s\beta}/c_{n\alpha}$ is the threshold for the reverse case. Under this set of assumptions, the difference between the pair of thresholds $c_{s\alpha}/c_{n\beta}$ and $c_{s\beta}/c_{n\alpha}$ expressed in dB is twice the difference in the gain of the i.t.f.(T), also expressed in dB, at the frequencies α and β. Note that in the Yang and Blake experiments, $c_{s\alpha} = c_{s\beta}$, while $c_{n\beta}$ and $c_{n\alpha}$ are varied to determine threshold.

Figure 5(a) shows a plot of all the thresholds from one data set (subject YY, 8 arcmin disparity) in which the centre spatial frequency of the signal and mask differed by 0.8 octaves, i.e. α is 0.8 octaves lower than β, and where data are available to estimate both $c_{s\alpha}/c_{n\beta}$ and $c_{s\beta}/c_{n\alpha}$. Thresholds have been plotted against the lower spatial frequency band in the stimulus (α). Thus, at each position along the x-axis, each pair of data points corresponds to stimuli made up of the same spatial frequency bands, but in one case α is the signal spatial frequency and in the other case α is the mask spatial frequency. If the assumptions underlying Eq. (13) are correct, the difference between a pair of data points at a particular spatial frequency reflects a difference in the gain of the i.t.f.(T) at α and β.

A systematic pattern to the data is evident when they are plotted in this way. At higher spatial frequencies, thresholds are consistently lower for the $c_{s\alpha}/c_{n\beta}$ condition (closed symbols), i.e. when the signal is 0.8 octaves lower than the mask. According to Eq. (13), this implies that higher spatial frequencies are attenuated more than lower spatial frequencies, i.e. the i.t.f. falls off at high spatial frequencies. At low spatial frequencies, the opposite is true: thresholds are consistently lower for the $c_{s\beta}/c_{n\alpha}$ condition (open symbols), suggesting that lower spatial frequencies are relatively more attenuated in this range.

By considering a series of pairs of spatial frequencies separated by 0.4 and 0.8 octaves [see Fig. 5(a) and Fig. 5(b)], we were able to reconstruct two separate estimates of the initial transfer function, T. The data are all from Fig. 3 of Yang & Blake (1991), but different data points were used to derive each of the curves. Both estimates are

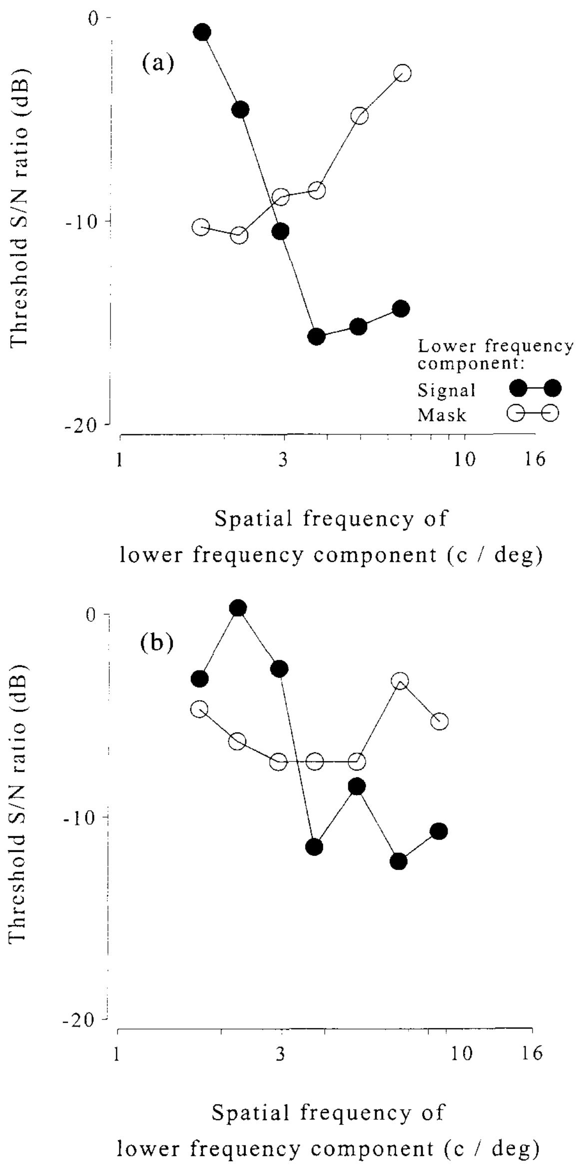

FIGURE 5. Data points are taken from Fig. 3 (subject YY) of Yang & Blake (1991). (a) Closed symbols show data for stimuli in which the centre frequency of the signal was 0.8 octaves lower than the signal. Open symbols plot data for the reverse case (mask centre frequency was 0.8 octaves lower than the signal). Threshold signal-to-noise ratios are plotted against the spatial frequency of the lower frequency component. Thus, any two points plotted at the same position on the x-axis represent data for two different stimuli that contained the same spatial frequency bands: in one case the lower band was the signal (closed symbols) in the other case it was the mask (open symbols). (b) As for (a), except that signal and mask centre frequencies differed by 0.4 octaves.

shown in Fig. 6. The mean of the these two estimates was used as a value for T at each spatial frequency in subsequent calculations. Also shown in Fig. 6 are data from Campbell & Robson (1968) on the contrast sensitivity for sinusoidal gratings in a 2 deg field, and data from Smallman & MacLeod (1994) for detection of 2-D patterns in a 1 deg field. Yang & Blake (1991) used a 4.4×4.4 deg field but the region of the disparate stimulus, which may be more appropriate for comparison with detection studies, was 2×1.1 deg.

The various estimates of the i.t.f. agree quite closely except at the lowest frequencies where Smallman & MacLeod (1994) found relatively greater sensitivity. It may be pointed out that if the assumptions underlying our estimation procedure were violated, then the shape of the resulting i.t.f. would appear unusual. The shape of the i.t.f. is particularly sensitive to the assumption about the symmetry of the internal masking functions.

In a sense, we have therefore achieved a partitioning of Yang and Blake's data into two components. The first component is a general i.t.f. that applies to all visual targets. This function has a broad spatial band-pass form similar to the threshold contrast sensitivity function for spatial frequency. The second component can be interpreted as a set of masking interactions between test and mask patterns as a function of the relative spatial frequency of the two.

Is it appropriate to propose a peaked i.t.f. similar to the shape of the threshold c.s.f. found for detection of low contrast stimuli, when the signal contrast used by Yang and Blake was supra-threshold? Relevant evidence comes from masking studies on the detection of static (Stromeyer & Julesz, 1972) and moving (Anderson & Burr, 1989) gratings. In both cases, detection thresholds increased with mask strength (specifically r.m.s. contrast) at a similar rate for all spatial frequencies tested. Thus, the relative sensitivity to different spatial frequencies remained the same over a wide range of stimulus contrasts. In addition, Yang and Blake report (personal communication) that the pattern of their results remained the same when a lower contrast signal was used.

An entirely different conclusion has been reached from experiments in which the perceived contrast of patterns with very different contrast thresholds have been matched correctly (Georgeson & Sullivan, 1975; Brady & Field, 1995). It has been argued that the effective supra-threshold contrast of high spatial frequency patterns may be "boosted" to compensate for the lack

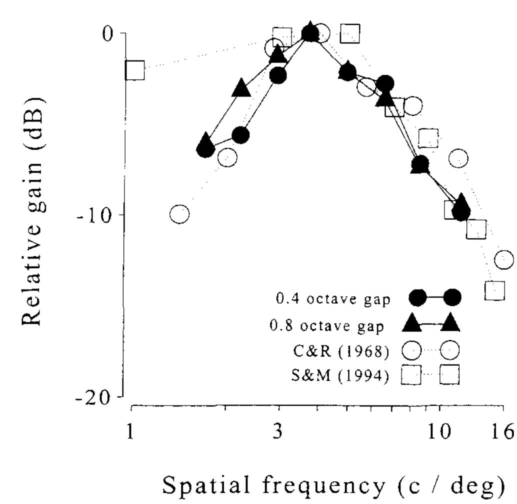

FIGURE 6. Two estimates of the initial transfer function (i.t.f.) derived from the data shown in Fig. 5(a) (triangles) and Fig. 5(b) (closed circles). Relative gain (in dB) is plotted against spatial frequency. The method used to derive the gain functions is described in the text. Data from Campbell & Robson (1968); Smallman & MacLeod (1994) are plotted on the same axes (open circles and squares, respectively).

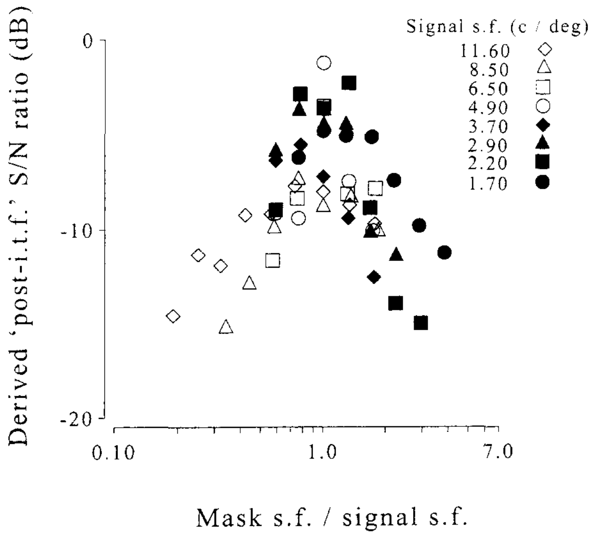

FIGURE 7. Post-i.t.f. signal-to-noise ratios are plotted here, derived from the original thresholds in Fig. 3 (subject YY) of Yang & Blake (1991) by assuming that the image first passes through an initial transfer function (i.t.f.), (see text). The abscissa plots the relative spatial frequency of the mask. All the transformed masking functions peak at about 1, i.e., when the centre frequencies of the signal and mask bands are equal. The peaks of the functions shown here are plotted as open squares in Fig. 8(b).

of sensitivity at threshold. In this view, it would be possible to argue that the masking patterns used by Yang and Blake should be equally effective regardless of spatial frequency, in contradiction to the assumption used in this paper. It should however be understood that these two experimental procedures are measuring very different properties of the visual system and there is no compelling reason why they should be related.

RESULTS

Having calculated an overall initial transfer function (T), it is simple to calculate the "effective" or "post-i.t.f." signal-to-noise ratio at the threshold for stereopsis for all combinations of signal and mask spatial frequencies:

$$\text{"post-i.t.f." signal-to-noise ratio} = \frac{c_s}{c_n}\frac{T(f_s)}{T(f_n)} \qquad (14)$$

where c_s and c_n are the contrasts of the signal and mask, f_s and f_n are their centre spatial frequencies. This revised estimate of the "effective" signal-to-noise ratio is plotted on the ordinate in Fig. 7. Thus, the points on this graph indicate the shapes of the internal masking functions, M_i, independently of the influence of the shape of the overall initial transfer function (i.t.f.) of the visual system. The abscissa is plotted as relative spatial frequency. The fact that the functions all peak at about 1 indicates that the most effective mask, after taking into account the i.t.f., is one with approximately the same spatial frequency as the signal, for all signal spatial frequencies tested. Most of the peaks lie within 0.4 octaves of the signal centre spatial frequency and the deviations appear to be randomly distributed.

The same point can be illustrated in a different way by plotting the peak mask spatial frequency against the signal spatial frequency. Figure 8(a) shows original data for six conditions from Yang & Blake (1991) plotted in this way. The peaks have been described by Yang and Blake as clustering around 3 c/deg at low signal frequencies and around 5 c/deg at high signal frequencies. Certainly, the peaks do not lie convincingly along the diagonal as Julesz & Miller (1975) predicted. However, the peaks of the transformed data, i.e., the internal masking functions, M_i, lie much closer to the diagonal as shown in Fig. 8(b).

For all the transformed data shown in Fig. 8(b), the same estimate of the i.t.f., T, has been used (illustrated in Fig. 6). We have also tried calculating a separate i.t.f. for each condition and transforming each data set using its "own" i.t.f. The results are essentially the same.

The heights of the masking functions shown in Fig. 7 are all approximately the same, but this is not always the case, as shown in Fig. 9. The data plotted here are for a

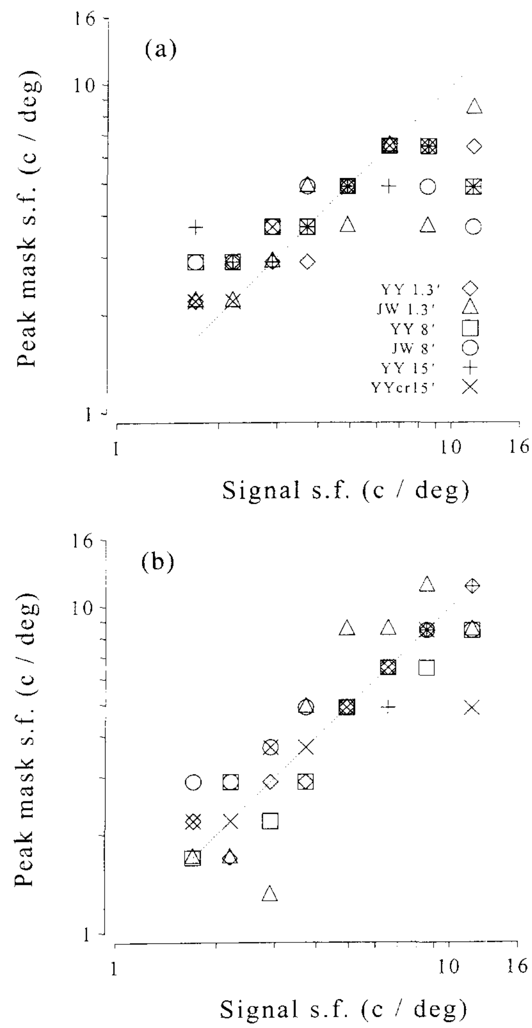

FIGURE 8. (a) Peak mask spatial frequency is plotted against signal spatial frequency for two subjects and three target disparities. Data are from Figs. 3, 5, 6 and 7 of Yang & Blake (1991). (b) When the original data are transformed to take account of the effect of the initial transfer function (see text), the peaks of the post-i.t.f. masking functions shift so that the peak mask spatial frequency lies closer to the signal spatial frequency (dotted line). Examples of complete transformed functions are shown in Fig. 7 and Fig. 9.

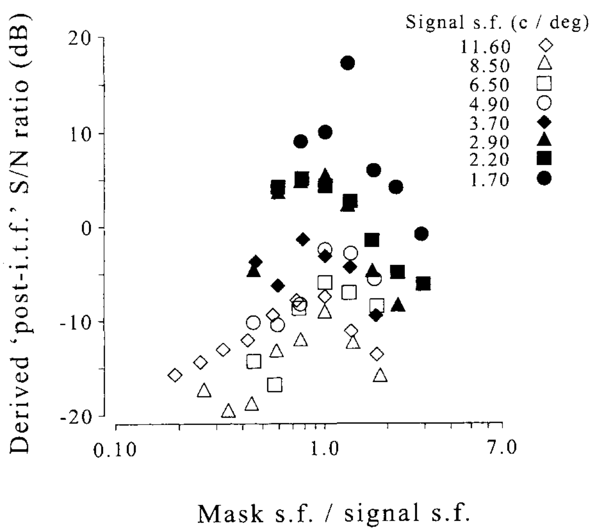

FIGURE 9. Post-i.t.f. signal-to-noise ratios plotted against relative mask spatial frequency for a 1.3 arcmin disparity target (original data from Fig. 5 of Yang & Blake (1991)). Unlike Fig. 7, the transformed masking functions are at different heights on the ordinate: thresholds are higher when the signal spatial frequency is low. The same point is made in Fig. 3(b), and in fact data from that graph are replotted here (all the points for which mask spatial frequency/signal spatial frequency = 1). Note that, despite the vertical shift in the functions, the peak masking frequency remains approximately equal to the signal. The peaks of the functions shown here are plotted as diamonds in Fig. 8(b).

target whose disparity with respect to the background was only 1.3 arcmin. As discussed above (*Analytic procedures*), such targets are poorly detected when the signal spatial frequency is low, which accounts for the fact that the masking functions are higher for low spatial frequency signals. Nevertheless, even in this case, the peak mask tends to be at the same spatial frequency as the signal.

DISCUSSION

We have shown that the masking data of Yang & Blake (1991) do not necessarily imply that there are stereo "channels" tuned to 3 and 5 c/deg. Instead, after taking into account the effect of an initial transfer function (including the m.t.f. of the eye), the data are consistent with the conclusion reached by Julesz & Miller (1975): the most effective mask is one with the same spatial frequency as the signal.

It may be objected that the logic we have applied appears to be circular: we have derived an estimate of the initial transfer function (i.t.f.) from Yang and Blake's original data and then used this in turn to transform the original data. The sceptic might argue that it is inevitable that the assumptions we used in deriving the i.t.f. (in particular, symmetry of the internal masking functions) will be borne out in the transformed data.

First, our main result (that the most effective mask is one of the same spatial frequency as the signal) is not critically dependent on the exact shape of the initial transfer function used to transform the data. A completely different estimate of the i.t.f. (independent of Yang and Blake's data) could have been used to arrive at exactly the same conclusion [e.g. the measurements of Campbell & Robson (1968) as, in fact, we used initially when beginning this analysis]. All that is required is for the i.t.f. to have a peak somewhere in the region of 3 c/deg. Furthermore, we used a single estimate of the i.t.f. [derived from Fig. 3 of Yang & Blake (1991)] to transform all of Yang and Blake's data. The issue of circularity applies only to the transformation of the data from which the i.t.f. was derived. As a separate exercise, we also derived estimates of the i.t.f. for other data sets and, although some were noisier, the shapes of all the i.t.f. estimates were approximately the same.

Second, even for the data of Yang and Blake's Fig. 3, only a small proportion of the data was used to derive the i.t.f. (those data points for which the signal and mask centre spatial frequencies differed by 0.4 or 0.8 octaves). It is true that a pair of data points used in deriving the i.t.f. are constrained to have the same height in the transformed graph (Fig. 7) but the points lie on different post-i.t.f. masking functions so, even for these points, the functions are not necessarily symmetrical about the signal spatial frequency. The positions of all the other data points remain unconstrained.

A somewhat different sceptical comment about our analysis might be that we have simply treated the data as a linear combination of two functions. There is no *a priori* reason why our chosen partition of the data should be preferred over any other. The main case for this choice has to be (i) that it provides a parsimonious account of this data set: the fact that the estimates of the i.t.f. and of the internal masking functions tend to agree across conditions; and (ii) a parsimonious agreement with other data: the fact that the shape of the i.t.f. that we have inferred from Yang and Blake's data is similar to many previous estimates.

Some of the earlier masking studies did consider the potential effects of an initial transfer function (usually the m.t.f. of the eye) on estimates of the shape of "channels", while some of the later studies have not. Stromeyer & Julesz (1972) noted that the shape of the threshold elevation curves from one of their experiments (their Figs 4 and 5), which are plotted against the cut-off frequency of the mask, would be changed if they had taken account of the greater sensitivity of the eye to gratings in the range of 2.5–10 c/deg. "For this reason," they say, "in the next experiment the noise band was maintained constant, and the frequency of the test grating was varied relative to the noise." By plotting threshold elevation against *signal* spatial frequency (e.g. their Figs 6–9) the effects of the i.t.f. are taken into account.*

It turns out that there are several studies, not just in stereoscopic vision, that have ignored the possible effects of an initial transfer function on the interpretation of the shapes of channels derived from masking data (e.g. Legge, 1978; Wilson *et al.*, 1983; Anderson & Burr, 1989). This is not the place to pursue this question in

*The relative heights of different curves may be affected by a reanalysis, but not the shape of the curves.

detail, but a re-analysis of such data would be of considerable interest. Specifically, in the case of stereo, Shiori *et al.* (1994) have carried out a masking study very similar to that of Yang & Blake (1991). They find that the peaks of the masking functions tend to cluster around a relatively narrow range of frequencies, just as Yang and Blake did, and a similar re-analysis may be worthwhile.

We have transformed raw signal-to-noise ratios into "post-i.t.f." masking functions but we have not claimed that these necessarily represent the shape of "channels" tuned to particular spatial frequencies. (Indeed, it is only when the peak of a masking function differs from the signal spatial frequency that there is good evidence for discrete "channels" at all.) Tyler *et al.* (1994) point out that, even if there was a flat i.t.f., the shape of a masking function does not, in general, correspond to the shape of the sensitivity profile of any of the underlying channels. This is a quite separate issue from the effect of a peaked i.t.f., which we have emphasised. The type of analysis Tyler *et al.* propose could, in theory, be applied to the post-i.t.f. masking functions we have derived (e.g., Fig. 7). If this were done, it is likely that a wider range of spatial-frequency-tuned channels would be required to fit the data than the two at about 2.5 and 7 cpd that Tyler *et al.* (1994) deduced from their analysis of the raw threshold signal-to-noise ratios in Yang and Blake's paper.

CONCLUSION

Julesz & Miller (1975), among others, have emphasised the effect of the modulation transfer function of the early visual system on both the signal and the mask contrast and the importance of taking this into account when interpreting the results of masking studies. We have re-analysed the data of Yang & Blake (1991) and argue that the results are consistent with the following conclusion: at the stage at which the signal and mask are combined, the most effective mask is one of the same spatial frequency as the signal.

REFERENCES

Anderson, S. J. & Burr, D. C. (1989). Receptive field properties of human motion detector units inferred from spatial frequency masking. *Vision Research, 29*, 1343–1358.

Blakemore, C. & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology, London, 203*, 237–260.

Brady, N. & Field, D. J. (1995). What's constant in contrast constancy? The effects of scaling on the perceived contrast of bandpass patterns. *Vision Research, 35*, 739–756.

Campbell, F. W. & Robson, J. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, London, 197*, 551–566.

Georgeson, M. A. & Sullivan, G. D. (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *Journal of Physiology, London, 252*, 627–656.

Graham, N. V. S. (1989). *Visual pattern analysers*. Oxford, U.K.: Oxford University Press.

Julesz, B. & Miller, J. E. (1975). Independent spatial-frequency-tuned channels in binocular fusion and rivalry. *Perception, 4*, 125–143.

Koenderink, J. J. & van Doorn, A.J. (1974). Detection of two-dimensional band limited noise. *Vision Research, 14*, 515–518.

Legge, G. E. (1978). Sustained and transient mechanisms in human vision: temporal and spatial properties. *Vision Research, 18*, 69–81.

Legge, G. E. & Gu, Y. (1989). Stereopsis and contrast. *Vision Research, 29*, 989–1004.

Mitchell, O. R. (1976). Effect of spatial frequency on the visibility of unstructured patterns. *Journal of the Optical Society of America, 66*, 327–332.

Shiori, S., Hatori, T., Yaguchi, H. & Kubo, S. (1994). Spatial frequency channels for stereoscopic depth perception. *Optical Review, 1*, 311–313.

Smallman, H. S. & MacLeod, D. I. A. (1994). Size disparity correlation in stereopsis at contrast threshold. *Journal of the Optical Society of America A, 11*, 2169–2183.

Stromeyer, C. F.III, Julesz, B. (1972). Spatial-frequency masking in vision: critical bands and spread of masking. *Journal of the Optical Society of America, 62*, 1221–1232.

Tyler, C. W., Barghout, L. & Kontsevich, L. L. (1994). Computational reconstruction of the mechanisms of human stereopsis. *Proceedings of SPIE conference 6–9 July, Asilomar, California, 2054*, 52–68.

Wilson, H. R., McFarlane, D. K. & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research, 23*, 873–882.

Yang, Y. & Blake, R. (1991). Spatial frequency tuning of human stereopsis. *Vision Research, 31*, 1177–1189.

Acknowledgements—Supported by a MRC Career Development Award. Part of this work was presented at a meeting of the Physiological Society, Keele, U.K., 1995 and the European Conference for Visual Perception, Tübingen, Germany, 1995. We are very grateful to Yuede Yang and Randolph Blake for supplying us with a copy of their data and to Christopher Tyler, Lenny Kontsevich and Harvey Smallman for helpful comments.