Masking efficiency as a function of stimulus onset asynchrony for spatial-frequency detection and identification

ANDREI GOREA*
Laboratoire de Psychologie Expérimentale (Université René Descartes et EPHE 3e section), associé au CNRS, 28 rue Serpente, 75006 Paris, France

Abstract—Detection and identification thresholds for grating targets were measured in the presence of a compound mask grating as a function of the stimulus onset asynchrony (SOA). The detection and identification SOA functions are both reversed U-shaped but they are not parallel. The detection-to-identification ratio is itself a reversed U-shaped function of SOA, even for stimuli two octaves apart, with a peak between + 20 and + 60 ms SOA (backward masking). It is argued that these results support the hypothesis according to which detection and identification are serial processes.

INTRODUCTION

Given the existence of low-level 'feature' detectors (Barlow, 1972; Graham, 1985) the problem arises as to the relationship between detecting and identifying perceptual primitives. This problem has been recently approached in relation to specific achromatic visual 'primitives' such as orientation (Thomas and Gille, 1979), temporal and spatial frequency (Olzak and Thomas, 1981; Watson and Robson, 1981; Thomas *et al.* 1982; Gorea, 1984; Olzak, 1985), direction of motion (Gorea, 1985), etc. One possibly exhaustive classification of the relationships between detection and identification processes can be represented as a two-by-two entry diagram. Detection and identification neural sites may coincide or may not. This is a 'hard'-type distinction. While confounded, detection and identification performances may depend on the same or on different processing rules. This is a 'soft'-type distinction. A commonly accepted detection rule is based on the high threshold assumption which posits that a detection event occurs any time that *at least* one feature detector exceeds its own threshold—an OR rule (Sachs *et al.*, 1971). To the extent that one detection event may entail an identification event the detecting unit is said to be 'labelled' (Watson and Robson, 1981). A correct identification might nevertheless depend on the weighted mean activation of many detectors—an AND rule. Initially introduced to account for hue perception (Hurvich and Jameson, 1955), this approach has been applied in spatial frequency, size and orientation discrimination studies (Gelb and Wilson, 1983; Regan and Beverly, 1983; Wilson and Gelb, 1984; Regan, 1985; etc.). Note that this approach requires the existence of labelled underlying detecting units.

---

*Present address: AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974, USA

If detection and identification occur at different sites, then they may be serial or parallel. OR and AND rules may hold in both cases. While the parallel wiring hypothesis allows for the counterintuitive possibility that identification may be better than detection, it remains anatomically and physiologically implausible. It implies the existence of two mutually exclusive populations of feature detectors one of which detects without identifying and the other which identifies without detecting! Note that this argument is unrelated to the possibility of a parallel processing of different visual characteristics such as 'positive' vs. 'negative' contrast (see Jung, 1972; for a more recent review see Legge and Kerstein, 1983), transient vs. sustained stimulations (see Lennie, 1980), or of different texton-stimuli (Julesz, 1980, 1981; Bergen and Julesz, 1983; Treisman and Souther, 1985). Also note that, under particular experimental conditions, identification performance may indeed be better than detection performances (Olzak and Thomas, 1981; Thomas et al., 1982). This is currently accounted for in terms of the decision processes involved in the two tasks although methodological artefacts might also account for such results (Klein, 1985).

Soft- and hard-type distinctions may lead to different predictions concerning the spatial and temporal processing characteristics in detection and identification tasks. The OR–AND distinction at a common processing level will predict larger spatial integration constants for the identification task (the AND rule implies larger spatial pooling), but possibly identical temporal integration constants. The serial vs. parallel distinction does not lead to clear-cut predictions of this type. In both cases the spatial and/or temporal integration characteristics of the two processes may or may not coincide. It has been recently shown that spatial-frequency detection and identification thresholds are parallel functions of stimulus duration (Gorea, 1986a). One interpretation of this result is that the two processes have identical time constants which can be taken as evidence of a common processing for the two tasks. (In fact, the common neural substratum hypothesis is widely shared in the literature—Thomas, 1985a.) Nonetheless, a serial processing of detection and identification might produce similar results if the time constant of the identification stage is substantially shorter than the time constant of the detection stage. This would be so because the contribution of the identification stage to the overall time constant of the system (as measured for the identification task) will be negligible (see Gorea and Tyler, 1983a; Tyler and Gorea, 1986).

One possibility of assessing the existence of such a serial wiring would be to selectively 'disrupt' processing at the identification stage. Traditionally, selective 'disruption' may be achieved in masking experiments where the temporal interval between the mask and the target stimulus is systematically varied (Breitmeyer, 1984). While masking efficiency as a function of stimulus onset asynchrony may be interpreted in various ways (see Discussion), masking experiments remain an interesting experimental approach to the understanding of the detection–identification relationships (e.g., Sagi and Julesz, 1985).

METHODS AND PROCEDURE

Stimuli
These were vertical sinusoidal gratings generated by a Picasso CRT Image Generator under computer control (M/OS-80 Mostek microsystem) and displayed on a Tektronix 608 monitor (P4 white phosphor). The inspection field was limited by a circular

aperture 5.4 deg in diameter (at 100 cm from the observer) and surrounded by a large (100 × 80 cm) white surface. The mean luminance of both the inspection and the surrounding field was set at 88 cd m^{-2}. Fixation was facilitated by means of four tiny black dots 1 cm apart. Two spatial-frequency target pairs were used in the medium–low (*a*, 1 and 1.5 cycles/deg; *b*, 1 and 4 cycles/deg) and in the medium–high (*a*, 5 and 7.5 cycles/deg; *b*, 2.2 and 5 cycles/deg) spatial-frequency range. The target pairs were chosen such as to activate, presumably, overlapping and nonoverlapping detection channels. The mask stimulus was a compound grating obtained by the electronic superposition of the two target frequencies to be detected and discriminated. In order to prevent the reciprocal cancellation of the target and of one of the mask components when 180 deg out of phase, the two gratings composing the mask stimulus were actually set at spatial frequencies 5% less than the target frequencies. The available apparatus did not permit the generation of a broad band noise mask which would have been more efficient in preventing such cancellation effects. However, pilot experiments have shown that, with sufficient training, variations within a range of −5% to +5% (including 0%) of the spatial frequency of the mask components (relative to the spatial frequency of the target) do not induce significant differences in the measured masking effects.

The spatial phases of the two mask- and of the target-stimuli were independently randomized from trial to trial. This prevented the observers from basing their identification judgements on the *systematic* beat effects in the three-component stimulus (Thomas, 1985b). The relative phases of the two masking gratings were also randomized across the two temporal intervals composing one experimental trial (see below). The contrast of each component of the mask stimulus was set at 20%. Both target and mask stimuli were flashed for 20 ms. Their temporal separation (stimulus onset asynchrony—SOA) could vary from −225 ms (forward masking) to +225 ms (backward masking). All stimuli were viewed binocularly by the author and by a second, well-trained observer. They both had normal vision.

The limitations of the spatial-frequency range used in this study were determined by two constraints. (1) Because the contrast of the two masking components could not be adjusted independently, they had to be selected within a spatial-frequency range where sensitivity (for briefly flashed stimuli) remains approximately constant (e.g., Nachmias, 1967). This ensured an approximately constant masking efficiency across spatial frequency and determined the upper bound of the frequency range. (2) Pilot experiments showed that, under masking conditions, discrimination contrast thresholds for targets within the low-frequency range (i.e. smaller than 1 cycle/deg) frequently exceeded the linear range of the screen. The lowest spatial frequency was consequently set at 1 cycle/deg.

Procedure

Detection and identification thresholds were measured by means of a 2 × 2 alternative forced choice (2 × 2 AFC) staircase procedure. The target stimulus appeared in one of two temporal intervals, the beginning and end of which were marked by auditory tones. It could be one of the two members of a spatial-frequency pair. In the masking conditions, the mask stimulus was presented in both temporal intervals. The observer had to decide which of the two intervals contained the target and to identify it as 1 or 2. Auditory feedback was provided for incorrect detection and/or identification responses. Four independent staircases were used concurrently. Two of them were

detection dependent and the other two were identification dependent. Each trial was randomly selected to belong to one of them. Three consecutive correct detection (or identification) responses resulted in a 2 dB contrast decrease. One incorrect response entailed an identical increase. This makes the staircase converge to 79.6% correct responses.

One experimental session consisted of at least 240 trials equally distributed among the two stimuli to be discriminated and the two response-dependent rules. Masked and unmasked thresholds were collected in separate sessions. In the masking conditions, SOAs for each stimulus pair were randomly chosen from session to session. Once a complete SOA function for a given stimulus pair was obtained, another stimulus pair was randomly chosen. More than half of the experimental conditions were repeated twice.

RESULTS

Datum points in the first two figures were obtained in the following way. The masked detection and identification thresholds for each stimulus in a pair and for each SOA
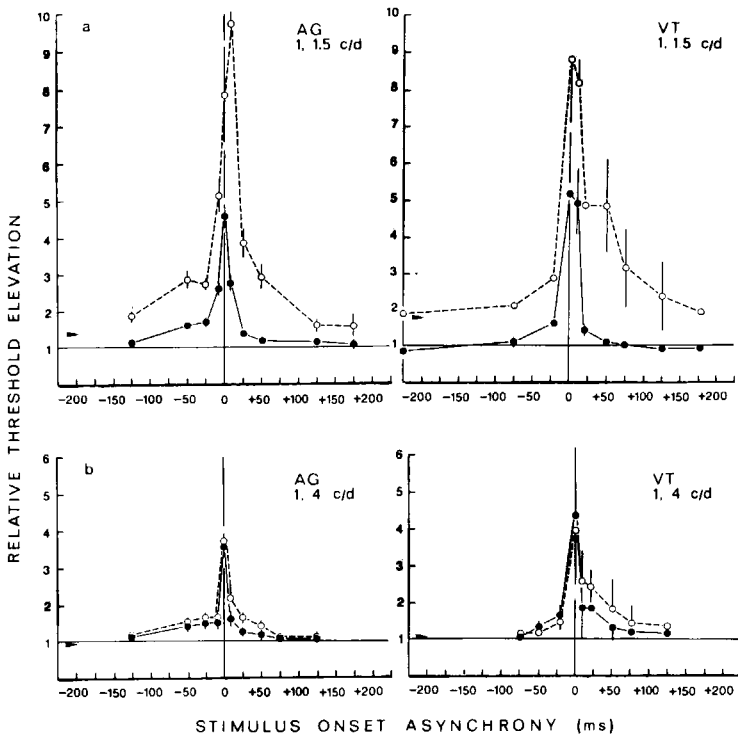

Figure 1. Relative detection (filled symbols) and identification (open symbols) threshold increments as a function of the target-mask stimulus onset asynchrony (SOA). Results for medium–low spatial-frequency stimuli (a) 0.58 and (b) 2 octaves apart. Negative and positive SOAs refer to forward and backward masking, respectively. Target and mask were displayed for 20 ms. The contrast of the two components of the mask was set at 20%. The horizontal lines (ratio of 1) and arrows represent the normalized detection and identification thresholds, respectively, in the absence of the mask stimulus. Results for the two observers are displayed in separate panels. Vertical bars show + 1 and/or − 1 standard errors.
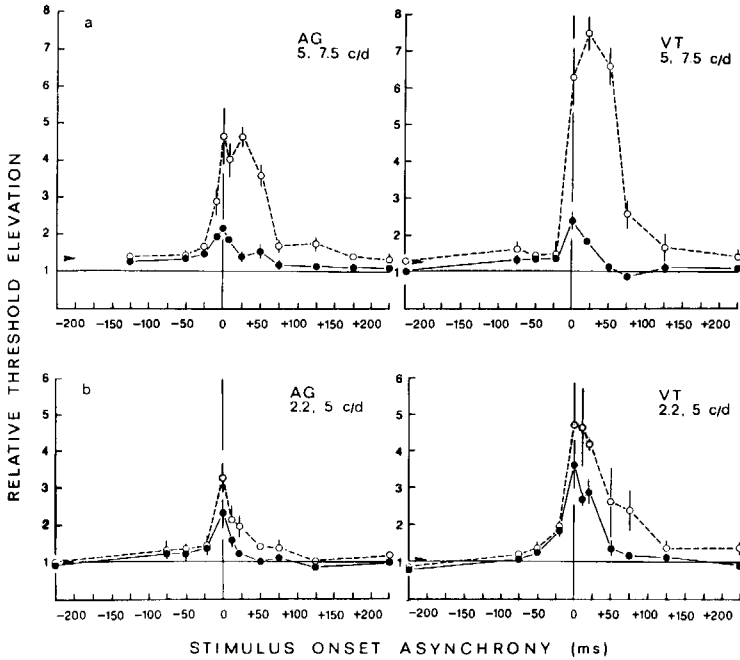
Figure 2. As in Fig. 1 but for the medium–high spatial-frequency pairs. Stimuli (a) 0.58 and (b) 1.18 octaves apart.

were expressed as a fraction of the respective unmasked *detection* threshold. Computed detection and identification ratios were then geometrically averaged across the two stimuli in the pair and across repetitions.

Figure 1 shows relative threshold increments as a function of SOA for detection (filled symbols) and identification (open symbols) of stimuli within the medium–low frequency range. Figure 2 displays similar data obtained with stimuli within the medium–high frequency range. Results for the two observers are shown in separate panels. The normalized threshold increments should be compared to the respective reference ratios as obtained without the mask stimulus. For the detection task the normalized reference ratio is 1 by definition (horizontal line in each panel), while for the identification task it is the average detection-to-identification (D/I) ratio (without mask) for each stimulus-pair (arrows).

Inspection of Figs 1 and 2 leads to the following observations:

1. While the overall shape of the detection and identification functions are rather similar for the two observers, observer VT shows more variability than observer AG, particularly for the identification performances under backward masking conditions. While these larger standard errors are partly related to the higher identification thresholds measured under these conditions, they probably reflect as well a higher level of difficulty of the experimental task combined with the lower training level of this observer.

2. The detection performances are rather typical of SOA masking functions obtained with medium to high spatial frequency content stimuli (see Breitmeyer, 1984). Previous studies demonstrated a double peak SOA masking function obtained with

low spatial-frequency targets (Green, 1981; Rogowitz, 1983) but the precise shape of the SOA masking function depends on a range of stimulus characteristics such as the duration of the target and of the mask, the contrast of the mask, the performance level set by the experimental procedure, etc. (Breitmeyer, 1984).

3. The peak of the identification masking function appears to be both spatial frequency and spatial-frequency-difference dependent; it is displaced toward positive SOAs (backward masking) when spatial frequency increases and when the difference in the spatial frequency of the stimuli to be discriminated decreases. Given the technical limitations described in the Procedure section, the spatial-frequency effect could not be tested more extensively.

4. At least for the small spatial-frequency differences (Figs 1a and 2a) the detection and identification performances appear to follow non-parallel functions of SOA. This effect is illustrated in Fig. 3 which displays the D/I sensitivity ratios as a function of SOA for the four experimental conditions and for the two observers (AG: circles; VT: squares).

In Fig. 3, the masked D/I ratios were normalized with respect to the D/I ratios obtained in the unmasked conditions for each observer and geometrically averaged across the two stimuli in a pair and across repetitions. An analysis of variance performed with the normalized D/I ratios and restricted to the SOA values where both observers were tested in all four experimental conditions (i.e. $-25, 0, 25, 50$ and 125 ms) shows a strong main effect of SOA ($F_{4,4} = 95.9$, $P < 0.0005$) and of spatial-frequency difference ($F_{3,3} = 35.8$, $P < 0.01$). Moreover, the comparison between the SOA effect for small (stimulus pairs 1, 1.5 and 5, 7.5 cycles/deg) and large (stimulus pairs 1, 4 and 2.2, 5 cycles/deg) spatial-frequency differences also yields a highly significant effect
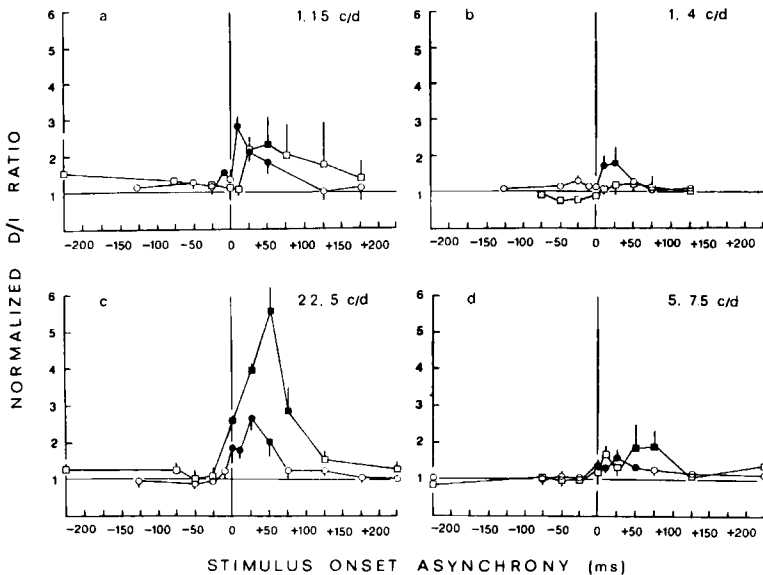

Figure 3. Relative detection-to-identification (D/I) sensitivity ratios as a function of SOA for the four experimental conditions. Datum points were normalized with respect to the D/I ratio obtained without mask for the two observers (AG: circles; VT: squares) and for each experimental condition. Vertical bars show $+1$ or -1 standard errors. Filled symbols refer to D/I ratios significantly higher (at a 0.05-level or more) than the reference, unmasked D/I ratio. See text for more details.

($F_{1,1} = 1.9 \times 10^6$, $P \ll 0.0001$). A test for multiple comparisons with a control (Dunnett, 1955 *passim*; Winer, 1970, pp. 89–92) was also performed to find out the D/I ratios which were significantly higher (one-tailed test) than the reference (unmasked) D/I ratio for each observer and for the four experimental conditions. Critical values at a 0.05 level or more were obtained from the Dunnett tables with (k, n) degrees of freedom (where k stands for the number of means to be compared including the control and n specifies the degrees of freedom for the MS error). The closed symbols in Fig. 3 represent D/I ratios significantly higher than the reference (unmasked) ratio. With one exception (1, 4 cycles/deg stimulus pair, Obs. VT–Fig. 3(b)), all the D/I SOA-functions display at least one, but typically two or more, D/I datum points significantly higher than the reference. They are all obtained for positive SOAs (backward masking) ranging from 0 to $+ 75$ ms with a peak increment somewhere between $+ 20$ and $+ 60$ ms (median SOA of about $+ 35$ ms).

It can thus be concluded that, under masking conditions: (a) the D/I ratio is SOA dependent; (b) the D/I variation with SOA depends on the spatial-frequency difference between the stimuli to be discriminated; and (c) in seven out of eight cases the D/I ratio is a reversed U-shaped function of SOA displaying a maximum for positive SOAs somewhere between 20 and 60 ms. The fact that in most of the studied cases the mask stimulus impairs identification to a greater extent than it impairs detection is evidence that the two tasks are, at least partly, processed at different neural sites.

DISCUSSION

Gorea (1986a) demonstrated that the detection and the identification of spatial frequency have similar temporal integration characteristics. The present study showed that the D/I sensitivity ratio varies with the delay between the test and the mask stimuli and hence provides evidence that the two processes do not take place at the same neural site. Given that parallel processing for these specific tasks is unlikely (see Graham, 1985; Thomas, 1985a), it is convenient to conclude that the two processes are serial, with detection taking place at the first processing stage. If the system is linear, the apparent identity of the measured temporal integration constants of the two processes may then reflect an identification time constant much shorter than that of the detection process (see also Gorea and Tyler, 1983a; Tyler and Gorea, 1986).

The interpretation of the detection and the identification SOA functions is necessarily dependent on the model one adopts to account for SOA masking functions in general. The literature provides a large variety of such models (see Breitmeyer, 1984, ch. 5) and it is not the aim of the present study to discriminate among them. Instead, I discuss below one possible qualitative (and intuitive) explanation. It is based on the idea that detection and identification are serial processes.

It is reasonable to posit that the information (i.e. the internal response) integrated at the detection level is transferred to the identification level only after its current, integrated value has attained some critical (threshold) point. The neurophysiological meaning of this 'gating' process is that the energy integrated at a given processing level is transferred to the next level after its transformation into spikes. This is a highly nonlinear operation.

To the extent that the internal responses to the mask and test extend over time, there will be a given range of positive and negative SOAs within which the two responses will overlap. Within this SOA range the detection of the target will be impaired because the

signal to noise ratio, i.e. the target + mask to mask ratio, at a given moment in time will be decreased. Because of the serial nature of the system, any detection impairment obtained in forward and simultaneous masking will necessarily entail an identification impairment. The D/I ratio will then remain roughly constant. This is true *independently* of the time required by the gating operation at the detection stage.

For backward masking the situation is different. Linear integration at the detection level is typically completed within 30 to 40 ms, but it continues as a decelerated function of duration beyond this limit (Gorea and Tyler, 1983b, 1986). Therefore, if the mask is delayed with respect to the target by more than a critical duration, the integration of the former will begin after the integration of the latter has been already completed and detection will then remain unaffected. However, if the integration of the mask occurs while the information integrated at the detection level is transferred to the identification level, the signal to noise ratio at this second stage will be decreased and the identification performances will be impaired. According to this interpretation, the maxima of the D/I functions displayed in Fig. 3 reflect, at least partly, the time course of the transfer operation which includes the time required for the detection stage to reach its own threshold.

The 'computational' approach described above can be reformulated within a system-analysis framework. Nonetheless, the assumption of a gating operation as well as of other types of nonlinearity (including noise characteristics) at the detection stage (see Gorea 1986b; Gorea and Tyler, 1986) will prevent the inference of the impulse response of the identification stage through division of the Fourier transform of its internal response (directly related to the shape of the duration vs. contrast threshold functions — see Gorea and Tyler, 1986) by the Fourier transform of the internal response of the detection stage. This means that the similarity between the detection and identification duration vs. contrast threshold functions demonstrated by Gorea (1986a) does not necessarily imply a much shorter impulse response for the identification stage as it would have been the case if the system was assumed to be linear. Given these considerations and because the overall shape of the SOA masking function is strongly related to the shape of the impulse response (Breitmeyer, 1984; Gorea, 1986b), it may be conjectured that a sufficient condition for obtaining the non-monotonic D/I functions as measured in this study would be that the impulse response of the identification stage presents a more rapid ascending part then the impulse response of the detection stage. Simulations run in my laboratory show indeed that the slope of the descending part (backward masking) of an SOA masking function is inversely correlated with the slope of the ascending part of the impulse response used to generate it.

In conclusion, the serial processing of the detection and identification tasks appears as the most straightforward explanation of the differences between the detection and the identification SOA masking functions. It is less clear how these differences could be accounted for on the basis of a 'soft'-type distinction between the two processes. While this point remains to be elucidated, the seriality hypothesis, already put forth by Sagi and Julesz (1985), remains the most plausible interpretation of the present results.

REFERENCES

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1, 371–394.

Bergen, J. R. and Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature* **303**, 696–698.

Breitmeyer, B. G. (1984) *Visual Masking: An Integrative Approach.* Oxford University Press, New York.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Am. Stat. Assoc.* **50**, 1096–1121.

Gelb, D. J. and Wilson, H. R. (1983). Shifts in perceived size due to masking. *Vision Res.* **23**, 589–597.

Gorea, A. (1984). Spatial and temporal characteristics for two types of detection/identification tasks. *J. Opt. Soc. Am.* **A1**, 1290A.

Gorea, A. (1985). Spatial integration characteristics in motion detection and direction identification. *Spatial Vision* **1**, 85–102.

Gorea, A. (1986a). Temporal integration characteristics in spatial frequency identification. *Vision Res.* **26**, 511–515.

Gorea, A. (1986b). Contrast perception over time: A multiple purpose model. *Invest. Ophthal. Vis. Sci. Suppl.*, A343.

Gorea, A. and Tyler C. W. (1983). On whether phase and contrast are processed by different mechanisms. *Invest. Ophthal. Vis. Sci. Suppl.*, A35.

Gorea, A. and Tyler, C. W. (1986). New look at Bloch's Law for contrast. *J. Opt. Soc. Am.* **A3**, 52–61.

Graham, N. (1985). Detection and identification of near-threshold visual patterns. *J. Opt. Soc. Am.* **A2**, 1468–1482.

Green, M. A. (1981). Spatial frequency effects in masking by light. *Vision Res.* **21**, 971–984.

Hurvich, L. M. and Jameson, D. (1955). Some quantitative aspects of an opponent-colors theory. II. Brightness, saturation and hue in normal and dichromatic vision. *J. Opt. Soc. Am.* **45**, 602–616.

Julesz, B. (1980). Spatial nonlinearities in the instantaneous perception of textures with identical power spectra. *Phil. Trans. R. Soc. Lond.* **290**, 83–94.

Julesz, B. (1981). Textons, the elements of texture perception, and their interaction. *Nature* **290**, 91–97.

Jung, R. (1972). Visual perception and neurophysiology. In: *Handbook of Sensory Physiology. VII/3. Central Visual Information*, R. Jung (Ed). Springer-Verlag, Berlin.

Klein, S. A. (1985). Double-judgment psychophysics: problems and solutions *J. Opt. Soc. Am.* **A2**, 1560–1585.

Legge, G. E. and Kerstein, D. (1983). Light and dark bars: contrast discrimination. *Vision Res.* **23**, 473–483.

Lennie, P. (1980). Parallel visual pathways. *Vision Res.* **20**, 561–594.

Nachmias, J. (1967). Effect of exposure duration on visual contrast sensitivity with square-wave gratings, *J. Opt. Soc. Am.* **57**, 421–427.

Olzak, L. A. (1985). Interactions between spatially tuned mechanisms: converging evidence. *J. Opt. Soc. Am.* **A2**, 1551–1559.

Olzak, L. A. and Thomas, P. (1981). Gratings: why frequency discrimination is sometimes better than detection. *J. Opt. Soc. Am.* **71**, 64–70.

Regan, D. (1985). Masking of spatial-frequency discrimination. *J. Opt. Soc. Am.* **A2**, 1153–1159.

Regan, D. and Beverley, K. I. (1983). Spatial-frequency discrimination and detection: comparison of postadaptation thresholds. *J. Opt. Soc. Am.* **73**, 1684–1690.

Rogowitz, B. E. (1983). Spatial/temporal interactions: backward and forward metacontrast masking with sine-wave gratings. *Vision Res.* **23**, 1057–1073.

Sachs, M. B., Nachmias, J. and Robson, J. G. (1971). Spatial frequency channels in human vision. *J. Opt. Soc. Am.* **61**, 1176–1186.

Sagi, D. and Julesz, B. (1985). 'Where' and 'What' in vision. *Nature* **228**, 1217–1219.

Thomas, J. P. (1985a). Detection and identification: how are they related? *J. Opt. Soc. Am.* **A2**, 1457–1467.

Thomas, J. P. (1985b). Effect of static-noise and grating masks on detection and identification of grating targets. *J. Opt. Soc. Am.* **A2**, 1586–1592.

Thomas, J. P. and Gille, J. (1979). Bandwidth and orientation channels in human vision. *J. Opt. Soc. Am.* **69**, 652–660.

Thomas, J. P., Gille, J. and Barker, R. A. (1982). Simultaneous visual detection and identification. *J. Opt. Soc. Am.* **72**, 1642–1651.

Treisman, A. and Souther, J. (1985). Search asymmetry: a diagnostic for preattentive processing of separable features. *J. Exp. Psychol. General* **114**, 282–310.

Tyler, C. W. and Gorea, A. (1986). Different encoding mechanisms for phase and contrast. *Vision Res.* **26**, 1073–1082.

Watson, A. B. and Robson, J. G. (1981). Discrimination at threshold: labelled detectors in human vision. *Vision Res.* **21,** 1115–1122.

Wilson, H. R. and Gelb, D. J. (1984). A modified line-element theory for spatial frequency and width discrimination. *J. Opt. Soc. Am.* **A1,** 124–131.

Wilson, H. R. and Regan, D. (1984). Spatial-frequency adaptation and grating discrimination: predictions of a line-element model. *J. Opt. Soc. Am.* **A1,** 1091–1096.

Winer, B. J. (1970). *Statistical Principles in Experimental Design.* McGraw-Hill, London.