Pergamon

PII: S0042-6989(96)00072-7

Vision Res., Vol. 36, No. 21, pp. 3457–3468, 1996

Printed in Great Britain
0042-6989/96 $15.00 + 0.00

The Interaction of Binocular Disparity and Motion Parallax in the Computation of Depth

MARK F. BRADSHAW,*† BRIAN J. ROGERS†

Received 18 July 1995; in revised form 1 February 1996

Depth from binocular disparity and motion parallax has traditionally been assumed to be the product of separate and independent processes. We report two experiments which used classical psychophysical paradigms to test this assumption. The first tested whether there was an elevation in the thresholds for detecting the 3D structure of corrugated surfaces defined by either binocular disparity or motion parallax following prolonged viewing (adaptation) of supra-threshold surfaces defined by either the same or a different cue (threshold elevation). The second experiment tested whether the depth detection thresholds for a compound stimulus, containing both binocular disparity and motion parallax, were lower than the thresholds determined for each of the components separately (sub-threshold summation). Experiment 1 showed a substantial amount of within- and between-cue threshold elevation and experiment 2 revealed the presence of sub-threshold summation. Together, these results support the view that the combination of binocular disparity and motion parallax information is not limited to a linear, weighted addition of their individual depth estimates but that the cues can interact non-linearly in the computation of depth. Copyright © 1996 Elsevier Science Ltd.

Binocular disparity Motion parallax Cue integration Threshold elevation Sub-threshold summation

INTRODUCTION

The depth perceived from binocular disparity or motion parallax cues has traditionally been considered to be the product of separate and independent processes. Indeed a common objective of many previous studies has been to demonstrate the effectiveness of each of the two cues when presented in isolation (Julesz, 1960; Rogers & Graham, 1979). Recently, however, considerable computational and psychophysical interest has centred on the question of whether disparity and parallax (and the many other sources of depth information) interact in the computation of depth when both are available to the visual system. Two general questions arise in this context: the first addresses whether the requisite mechanisms exist in the human visual system to support such interactions and the second addresses the computational advantages that accrue from the combination of the different cues (e.g., Richards, 1985; Waxman & Duncan, 1985). The first question is addressed in the present paper.

The mechanisms sensitive to disparity and parallax information are particularly good candidates for possible early (prior to depth computation) cue interactions. Not only can the nature of the information (spatial or spatio-temporal disparities) be related at a formal level but there are also similarities in the way we use the information and in the underlying mechanisms. Binocular disparity can be considered as the consequence of viewing the world from two spatially separated vantage points (the left and right eyes) at the same time, whereas motion parallax can be considered as the consequence of viewing the world from two spatially separated vantage points at different moments in time (Rogers & Graham, 1982). If we consider the case in which the observer's eye moves through the inter-ocular distance, and nothing moves in the world, then the problem of depth computation in the two cases is formally equivalent (e.g., Koenderink, 1986). Rogers & Graham (1982) determined the absolute sensitivity of the visual system for detecting the 3-D structure of sinusoidally corrugated surfaces which were specified by either motion parallax or binocular disparity. They found that the shape of the sensitivity functions were remarkably similar over a range of corrugation spatial frequencies (0.05–1.6 c/deg). The same authors also established that similar simultaneous and successive contrast effects could be created in both domains and in later work they used cross adaptation and depth biasing techniques to demonstrate interactions between domains (Graham & Rogers, 1982a,b, Rogers & Graham, 1984; see also Nawrot & Blake, 1991). These results all suggest that the information from both domains must come together at some stage in the visual system. Anstis &

---

*To whom all correspondence should be addressed at the Department of Psychology, University of Surrey, Guildford, Surrey GU2 5XH, U.K. [*Fax* 01483 32813; *Email* m.bradshaw@surrey.ac.uk].

†Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, U.K.

Harris (1974) using a contingent aftereffect paradigm provided further evidence in support of this suggestion. In their adaptation period, observers viewed a leftward moving pattern with a crossed disparity in alternation with a rightward moving pattern with an uncrossed disparity. This produced directional motion aftereffects that were contingent on disparity, and depth/disparity aftereffects that were contingent on the direction of motion. One interpretation of these results is that the aftereffects were the outcome of adaptation of neurons sensitive to both binocular disparity and the direction of motion. This is further supported by certain physiological findings. For example, cells have been found in the visual cortex that respond to both disparity and motion (Poggio & Talbot, 1981; Maunsell & van Essen, 1983). Of particular interest are cells in MT and MST which have been found to be sensitive to different components of the optic flow field such as rotations, dilations and deformations (Tanaka *et al.*, 1989; Orban *et al.*, 1992; Lagae *et al.*, 1994). Some of these cells, which may be involved in the computation of structure-from-motion, also respond selectively to disparity. However, such cells are evidently not merely 'double duty' as some modulate their response to motion when disparity is present. Roy *et al.* (1992) showed that the directional selectivity of certain cells in area MST was modified, depending on whether the moving stimulus was presented with crossed or uncrossed disparity. In area MT most cells are directionally selective and so will not respond if motions in opposite directions stimulate their receptive fields. However, Bradley *et al.* (1995) have recently reported that if the opposite directions of motion are separated in depth by disparity then the cells do respond. The properties of these cells are consistent with the possible function of detecting relative depth in the world from disparity and motion parallax information.

Cue conflict paradigm

The cue conflict paradigm has often been used to investigate how the human visual system processes the information obtained from different sources. In this paradigm, the cues of interest are presented to an observer in different degrees of conflict and the perceptual consequences monitored. Bülthoff & Mallot (1988) and Maloney & Landy (1989) have suggested that many of the results of cue combination experiments can be modelled by a weighted linear summation. A major problem with such models, however, is that the weight attributed to a particular cue is often crucially dependent on the specific stimulus parameters used and the degree of conflict between the cues in the stimuli (see Dosher *et al.*, 1986; Bradshaw *et al.*, 1991a). Clark & Yuille (1990) have suggested that these models can be distinguished according to whether they demonstrate weak or strong fusion (see also Landy *et al.*, 1995). Weak fusion is used to describe the situation where the cues are processed by separate mechanisms and then combined. Strong fusion is used to describe the possibility of an interaction between cues that occurs during, rather than after, the separate processes. Many examples of weak fusion are described in the literature (see Dosher *et al.*, 1986; Landy *et al.*, 1995). Evidence for strong fusion is less common. Rogers & Collett (1989) showed that when binocular disparity and motion parallax information specified depth profiles of slightly different peak-to-trough amplitudes, the two cues both influenced what was perceived, but in different ways. The magnitude of the perceived depth was determined primarily by the binocular disparity, whereas the motion parallax signal determined the perceived rotation of the stimulus surface (concave or convex) as it translated to-and-fro in the frontal plane. That is, perceived depth and perceived rotation were found to co-vary. More recently, Johnston *et al.* (1994) have demonstrated that the visual system can take advantage of the additional geometric information that is present when binocular disparity and motion information are presented together in the same stimulus (see also Richards, 1985). They found that shape judgements were accurate when both cues were available in a two-frame apparent motion sequence (they term this *promotion*), whereas perceived shape was subject to systematic distortions when either cue was presented in isolation. This suggests that the mechanisms sensitive to binocular disparity and relative motion may interact in the computation of depth (strong fusion).

In summary, the mechanisms involved in the computation of depth from binocular disparity and motion parallax cues share many empirical similarities. These similarities may be a consequence of the formal relationship between the respective sources of information, which raises the possibility that there may be mechanisms which are sensitive to both types of information. Psychophysical and physiological findings both support this possibility. The present paper reports two experiments which investigate whether mechanisms to support such an interaction exist in the human visual system. To do this we have adapted two classical psychophysical techniques from spatial contrast vision: (i) threshold elevation following adaptation; and (ii) sub-threshold summation.

GENERAL METHODS

Stimuli

The stimuli were 50% density random dot patterns visible within a circular aperture of 25 deg diameter. Dot separation was 6.25 arc min. Horizontally oriented corrugations with a sinusoidal profile in depth (i.e., modulated as a function of vertical position) were defined by either (i) binocular disparity; (ii) motion parallax; or (iii) both cues together, depending on the experimental condition. The spatial frequency and peak-to-trough depth of the modulations were variable and are given in the appropriate sections below. The horizontal centre line of the corrugated surface was marked by two horizontal white lines, 1 deg long, superimposed towards the left and right edges of the pattern. The pattern of random dots was changed on every trial.

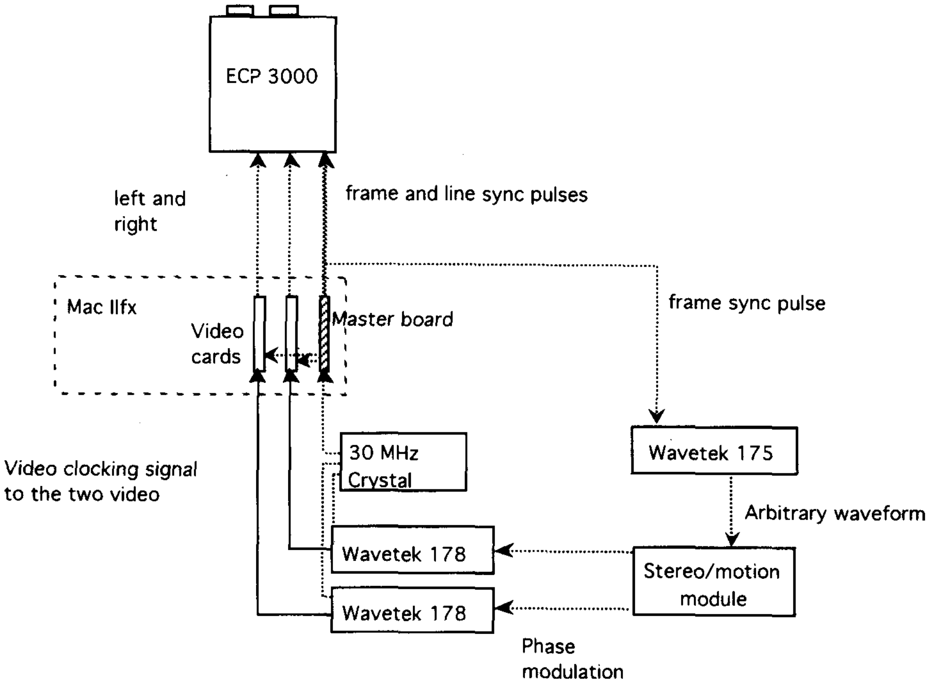

FIGURE 1. A schematic representation of the equipment used to generate and display the experimental stimuli (see text for details).

Binocular disparity and motion parallax signals

Three separate video cards, resident in an Apple Macintosh IIfx computer were required to create the experimental stimuli. Two of the video cards provided the random dot patterns seen by the two eyes and the third was used as a reference in the phase-shifting techniques described below (the method of dichoptic separation is described in the following section). Binocular disparities between the left and right eye's images were created by phase-shifting the video signal in an equal and opposite direction on the two cards. The phase-shifting technique was chosen as it permitted sub-pixel displacements (disparities) to be created so that the depth profiles of the corrugated surfaces appeared to be smooth. Identical patterns of continuous relative motion were created by phase-shifting the video signal in the same direction on the two cards.

A block diagram of the equipment is shown in Fig. 1. The video clocking signals for the three video cards were derived from a single external 30 MHz crystal so that the three cards were always synchronised.

The first video card, which was connected directly to the crystal, was used solely to supply the line and frame sync pulses for the projection T.V.. The other two cards provided the video signals for display to the left and right eyes. The clocking signals for these two cards were provided by two Wavetek 178 programmable waveform synthesisers (each running at 30 MHz) and phase locked with the crystal. The output waveform of the 178s could be phase-modulated by an externally applied voltage. This meant that the video clocking signals of the other two video cards (each connected to a separate Wavetek 178) could be advanced or retarded separately by precise, sub-pixel, amounts relative to the line and frame sync pulses created by the first card. A phase shift of the video signal relative to the line sync pulses will create a spatial displacement of all dots on a given raster line of the display and hence a binocular disparity between corresponding dots in the two eye's views. The magnitude of the phase shift applied to each line was determined by a Wavetek 175 arbitrary waveform generator which was synchronised to the frame rate of the master card (card 1). By varying the frequency and/or the amplitude of the waveform provided by the Wavetek 175, corrugations with different spatial frequencies and peak-to-trough depth profiles could be created.

Binocular disparity was created by phase-shifting each line of the video signal by an equal and opposite amount on the two cards. This was accomplished by a purpose-built module (stereo/motion module in Fig. 1) which created equal and opposite time varying signals from the output of the Wavetek 175. To produce motion parallax, each line of the video signal was advanced or retarded by applying identical phase shifts to both cards, the amplitude and direction being determined by the horizontal position of the observer's head (see below). The modulation in depth, when specified by either binocular disparity or motion parallax, appeared smooth and continuous.

Apparatus

Dichoptic presentation was achieved using crossed pairs of polaroid filters. The left and right eye's images were superimposed on a non-depolarising screen by an Electrohome Projection Television (ECP 3000). The projection T.V. was fitted with two green guns, each of which was driven by a separate video card to provide the (different) images for the left and right eyes. Each gun

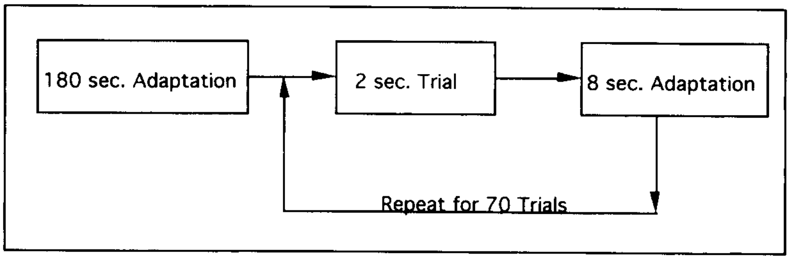

FIGURE 2. The adaptation test cycle of experiment 1.

had crossed polarising filters placed in front of its lens. The observer was seated with his or her head supported by a chin rest and viewed the stimuli through polaroid spectacles which ensured that the left and right eyes received only their respective images (one from each video card). Cross-talk from the opposite eye's image was negligible (greater than 1.25 log units down).

The chin rest was constrained to move horizontally and parallel to the screen with end-stops placed 13 cm apart. During the experiments, observers were required to move their head (in the chin rest) to-and-fro at a rate of 1 Hz paced by a metronome. A potentiometer, attached to the chin rest, monitored the position of the head and the voltage was used to modulate the horizontal position of dots along a raster line to mimic observer-produced parallax in natural viewing, as described previously by Rogers & Graham (1979).

Psychophysical techniques

The "method of constant stimuli" was used to determine thresholds for detecting the depth corrugations. The observer's task was to report whether the horizontal corrugation lying across the centre of the dot pattern, and marked by the two white lines, was concave (a trough) or convex (a peak). On each trial, the amplitude and phase of the depth signal (disparity and/or motion parallax) was randomly chosen from seven possible values corresponding to -3, -2, -1, 0, 1, 2 or 3 times the smallest step size—where negative amplitudes indicate the sinusoidal modulation was in the opposite phase. An experimental session consisted of 280 trials (in four blocks), corresponding to 40 trials of each of the seven stimulus levels. Frequency of seeing plots were generated from each data set and the best-fitting cumulative gaussian curve was determined using the probit technique (Finney, 1971). The 75% correct point on the psychometric function was taken as the threshold value.

EXPERIMENT 1: THRESHOLD ELEVATION

Introduction

The purpose of the first experiment was to examine the independence of mechanisms tuned to binocular disparity and motion parallax using a threshold elevation paradigm (e.g., Pantle & Sekuler, 1968; Schumer & Ganz, 1979). This procedure was used to determine whether thresholds for detecting sinusoidal corrugations defined by binocular disparity or motion parallax increased after prolonged viewing of similar corrugations defined by either the same cue (within-cue adaptation) or the other cue (between-cue adaptation). Following the rationale of Pantle & Sekuler (1968), we assume that the degree to which an adapting pattern can affect the detectability of a subsequent pattern reflects the extent to which both are processed by the same mechanism. The existence of between-cue adaptation, therefore, will be taken as evidence against the hypothesis of cue-independent mechanisms.

Methods

The "method of constant stimuli" described above was integrated into an *adaptation test* cycle (see Fig. 2). Observers adapted to supra-threshold corrugations (defined by either binocular disparity or motion parallax) which were phase-reversed every 2 sec. This phase reversal, together with the to-and-fro head movements and the replacement of the random dot pattern every 2 sec eliminated the possibility that any local disparity, motion or luminance negative aftereffects might develop. No such aftereffects were reported by any observer.

The peak-to-trough amplitude of the adapting surface for motion parallax and disparity-defined corrugations was 4.5 arc min disparity or equivalent disparity (approximately 20–30 times threshold). The initial adapting period was 3 min followed by a test trial lasting 2 sec and then a further 8-sec top-up adaptation period (two presentations of each of the phases). The duration of each block of 70 test trials was ~15 min.

Observers moved their head to-and-fro continuously throughout all adaptation conditions and always viewed the stimuli binocularly. Thresholds were determined for the detection of corrugations defined by (i) binocular disparity; and (ii) motion parallax in the following conditions:

1. The baseline conditions (no adaptation).
2. Following adaptation to disparity defined corrugations.
3. Following adaptation to parallax defined corrugations.

As a control to ensure that the length of the adaptation period *per se* did not cause threshold elevation, thresholds were also determined in a *flat adaptation* (i.e., no depth modulation) control condition. The same sequence as depicted in Fig. 2 was followed but the adapting

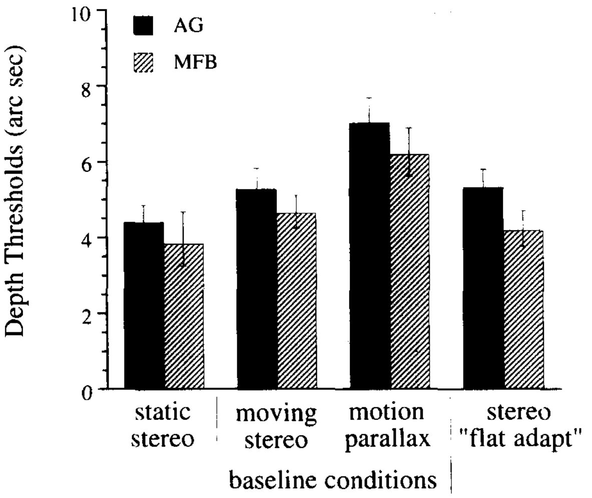

FIGURE 3. The depth discrimination thresholds found in (i) head static stereo; (ii) head moving binocular baselines; and (iii) the flat adapt conditions for two observers. The head static conditions are shown for reference only. Errors bars depict the SEs of the standard deviations of the cumulative gaussians, fitted to the data by the Probit technique.

TABLE 1. Threshold values determined for both observers following adaptation in the within- and between-cue conditions

	Stereo thresholds (arc sec)		*Motion thresholds* (arc sec equivalent disp.)	
	AG	MFB	AG	MFB
Within-cue Adaptation	13.5 (1.4)	8.3 (0.72)	12.9 (1.39)	9.9 (0.97)
Between-cue adaptation	8.0 (0.86)	7.49 (0.71)	11.8 (1.14)	6.9 (0.69)

SEs of the threshold values are given in parentheses. The baseline values shown in Fig. 3 were used with these values to compute the threshold elevations represented in Fig. 4.

surface contained no depth modulation. Hence there was a fourth condition:

4. Following adaptation to a flat surface.

The order of presentation of these four conditions was randomly interleaved across subjects. Pilot experiments were used to establish a suitable step size for the method of constant stimuli for each observer and in each condition. The pilots also provided observers with practice in making the appropriate evenly paced head movements. Two experienced psychophysical observers took part in the experiment (each with 6/5 visual acuity). The spatial frequency of the depth corrugations was 0.2 c/deg.

Results

In order to directly compare the motion parallax thresholds with the stereo thresholds the former were converted into "equivalent disparities". Equivalent disparity is the maximum amount of relative displacement created between a peak and trough in the stimulus as the head moves through the inter-ocular distance—6.5 cm (see Rogers & Graham, 1982). In other words, a surface with a given peak-to-trough depth will create the same *equivalent* disparity for a binocular observer as it would motion parallax for a moving observer. This equivalent disparity ordinate is labelled "depth threshold".

Thresholds for the baseline (1) and control conditions (4) are plotted in Fig. 3. The results of a "static disparity" (i.e., no head movements) condition, which was included in the experiment for reference purposes, are also shown (left bars). Thresholds for the static disparity condition were 3.9 and 4.4 arc sec peak-to-trough disparity for the two observers.

In the "head movement disparity" condition, thresholds were slightly higher (4.6 and 5.2 arc sec peak-to-trough disparity, respectively). This may be attributable to the slight cue conflict between the binocular disparity and the absence of appropriate motion parallax to accompany the head movements, but it is more likely to be due to the additional difficulty of making judgements while maintaining paced head movements. There was no appreciable difference in performance between the "head movement disparity" condition and the "flat adaptation" (binocular disparity) condition. This suggests that there was no artefact as a result of the long adaptation period (e.g., fatigue due to prolonged head movements) causing thresholds to rise. Therefore, thresholds for the baseline conditions were determined without the interposition of the adaptation periods which would have been rather arduous for observers. Thresholds for detecting the structure of corrugations defined by motion parallax were higher than those for binocular disparity (5.7 and 7.0 arc sec peak-to-trough equivalent disparity).

The values of the thresholds determined following within- and between-cue adaptation are shown in Table 1.

The elevation of depth thresholds following within-cue adaptation was large and is plotted, both as a threshold elevation ratio and as a percentage change, in Fig. 4(a). Averaged over the two observers, disparity thresholds more than doubled (112%) after adaptation to depth corrugations defined by binocular disparity; and parallax thresholds rose by 76% after adaptation to depth corrugations defined by motion parallax.

Figure 4(b) shows that there was also an appreciable amount of between-cue threshold elevation. Averaged over the two observers, disparity thresholds rose by 50% after adaptation to corrugations defined by motion parallax; and parallax thresholds rose by 45% after adaptation to corrugations defined by binocular disparity. Threshold elevation in the between-cue conditions was not as large as that found in the within-cue conditions. The implications of this finding are discussed below. Figure 5 replots the amount of between-cue threshold elevation of each observer, normalised by the individual adaptability to each cue (as indicated by the amount of within-cue threshold elevation).

When expressed in this manner and averaged across observers, between-cue parallax thresholds rose to 62% of their within-cue values and between-cue disparity thresholds rose by 47% of their within-cue values.

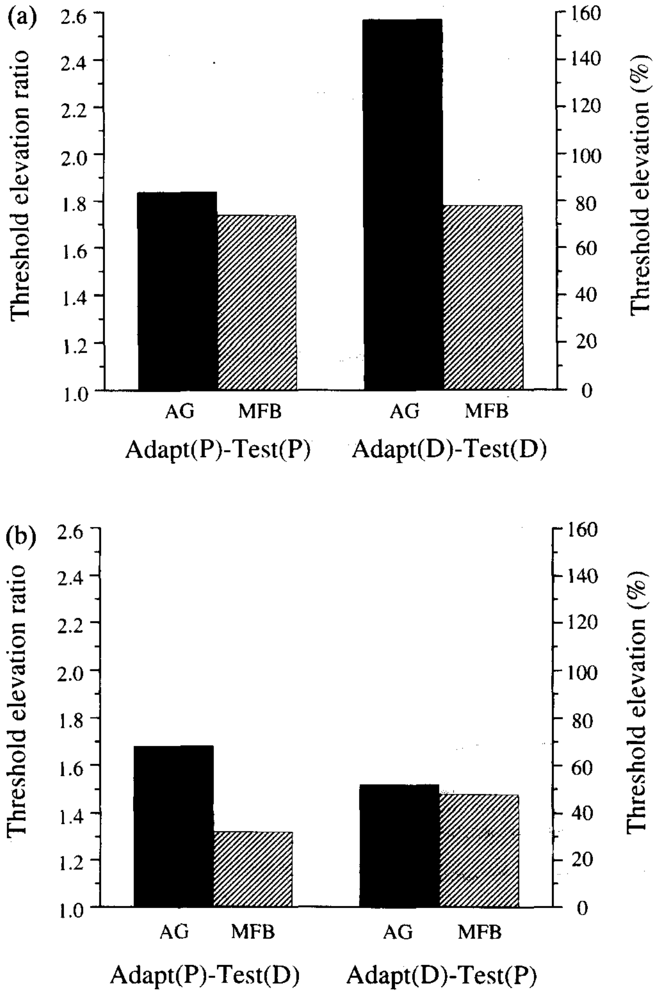

FIGURE 4. (a) The within-cue threshold elevation plotted separately for two observers. (b) The between-cue threshold elevation plotted separately for the same two observers. The ordinate plots threshold elevation as a ratio and the alternative *y* axis plots the same value as a percentage increase. "*D*" indicates disparity conditions and "*P*" indicates motion parallax conditions.

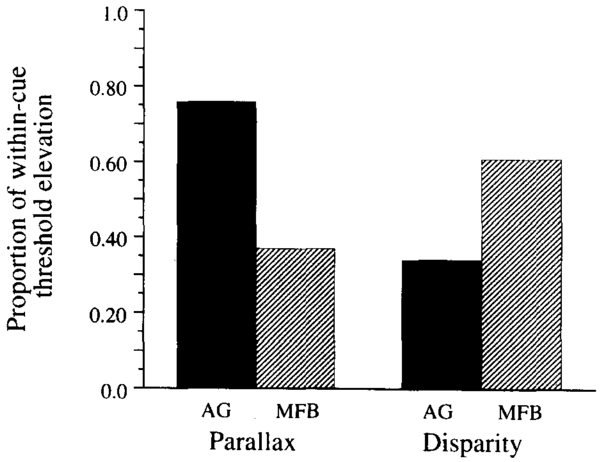

FIGURE 5. The amount of between-cue threshold elevation of each observer as a proportion of their individual adaptability to each cue, as indicated by the amount of within-cue threshold elevation, is replotted separately for two observers. The left-hand bars indicate the between-cue adaptability of the parallax system and the right-hand bars indicate the between-cue adaptability of the disparity system.

Discussion

Significant between-cue threshold elevation strongly suggests that mechanisms tuned to binocular disparity and motion parallax are not completely independent but must be linked at some stage in the computation of depth. The amount of within-cue adaptation was large and its magnitude is comparable to previous results in the depth domain. For example, Schumer & Ganz (1979) found threshold elevations in the disparity domain of ~75% (cf. Fig. 4, 1979). Graham & Rogers (1982b) assessed the amount of within- or between-cue signal that was necessary in order to cancel, or null, a large negative depth aftereffect produced by the prolonged viewing of a surface modulated in depth and defined by either disparity or motion parallax. The strength of the aftereffect was measured by determining the amplitude of the same, or different, cue required to cancel the impression of depth when a flat surface was viewed (i.e., a depth modulation, 180 deg out of phase with the adapting surface, was superimposed on the test surface and the observer adjusted its amplitude until the test surface appeared flat). They found that 75% of the within-cue motion parallax signal was required to cancel the depth aftereffect produced by disparity and 34% of the within-cue disparity signal was required to cancel the depth aftereffect produced by adaptation to parallax corrugations. These values are similar in magnitude to the between-cue threshold elevation found in the present experiment (62% and 47%, respectively).

The fact that the within- and between-cue threshold elevation is different in magnitude suggests that the human visual system contains two separate pools of neurons sensitive to either binocular disparity or motion parallax, together with a third pool of neurons sensitive to both cues. In this scheme, within-cue threshold elevation would result from the joint adaptation of cells tuned to a single cue (binocular disparity *or* motion parallax) and cells tuned to both cues (binocular disparity *and* motion parallax) whereas between-cue threshold elevation results only from the adaptation of cells tuned to both cues. If it is assumed that the outputs of each class of cell are integrated at some stage, and the size of an aftereffect is proportional to the number of cells adapted, then this model can account qualitatively and quantitatively for the within- and between-cue threshold elevation (see also Moulden, 1980).

Alternatively, between-cue adaptation could reflect the operation of a more central "depth" mechanism, which receives separate inputs from the motion parallax and binocular disparity mechanisms. If this central mechanism were adaptable then it could account for between-cue threshold elevation. However, to account quantitatively for the difference in the magnitude of within- and between-cue effects, this hypothesis requires additional assumptions. The simplest of these would be to assume that performance in the detection task could be affected by factors at different levels of the system. That is, thresholds may increase due to the adaptation of separate mechanisms sensitive to disparity and parallax, or they may increase due to the adaptation of subsequent mechanisms sensitive to depth (with inputs from disparity and parallax). These mechanisms would jointly or separately affect post-adaptation thresholds in the within- or between-cue adaptation/test cycles, respectively. Post-adaptation thresholds in the within-cue conditions would by influenced by the adaptation of both mechanisms, whereas in the between-cue condition they would only be influenced by adaptation of the common depth mechanism. In support of this idea Bradshaw *et al.* (1995) reported that thresholds for discriminating structure in depth corrugations defined by binocular disparity did not depend on the stimulus disparity alone. Rather, at close viewing distances, depth (i.e., scaled disparity with respect to viewing distance) determined performance. Therefore, different factors can determine threshold performance depending on the prevailing circumstances. However, whether separable thresholds exist in the motion domain (for shearing motion or depth, for example) is questionable (see Bradshaw *et al.*, 1991b).

The importance of the present experiment is that it clearly demonstrates that the mechanisms which support the processing of depth from binocular disparity and motion parallax information are not completely independent. However, it does not distinguish between the possibilities of late, linear interactions and early, non-linear co-operative interactions between the two cues. The second experiment addresses this issue more directly.

EXPERIMENT 2: SUB-THRESHOLD SUMMATION

Introduction

In a sub-threshold summation paradigm, thresholds for detecting a compound stimulus, in this case a surface specified by both binocular disparity and motion parallax cues, are compared to thresholds for detecting simple stimuli specified by each cue separately (Graham & Nachmias, 1971; Graham, 1989). If the mechanisms responding to the compound are not completely independent, thresholds may be lower for the compound stimulus than for either of its components. Independence implies that detection decisions are made separately for each cue without the influence of other mechanisms. Adaptation is not used in this technique.

In our second experiment we determined the detection thresholds for surfaces defined by (i) binocular disparity; (ii) motion parallax; and (iii) both cues together. Several outcomes of the experiment can be envisaged. If we assume that the disparity and parallax mechanisms are independent, then thresholds for a compound stimulus containing both disparity and parallax may either be reliably detected at the point where the most sensitive component reaches its own individual threshold (a first-past-the-post rule) or they may decrease owing to the effects of probability summation (see Graham, 1989). Alternatively, if the disparity and parallax mechanisms are not independent then thresholds for the compound stimulus may decrease, owing to the interaction between the mechanisms. The combination of the cues may be linear or non-linear.

Method

In the main experiment, separate thresholds were determined for surfaces defined by (i) binocular disparity; (ii) motion parallax; and (iii) both cues ("compound stimulus"). In each of these conditions the observer made side-to-side head movements and viewed the stimulus binocularly, as described for the first experiment. The three conditions were randomly interleaved. Three spatial frequencies of depth modulation were used: 0.1, 0.2 and 0.4 c/deg. Three experienced observers took part in the experiment, two with 6/5 and one with 6/6 visual acuity.

Prior to the main experiment, a pilot study was carried out to establish the thresholds for detecting binocular disparity and motion parallax defined surfaces for each of the three observers at the three different spatial frequencies. The purpose of this pilot study was to establish each observer's relative sensitivity to the cues so that an appropriate compound stimulus could be constructed for presentation in the main experiment. The aim was to create a compound stimulus for each observer in each condition which contained disparity and parallax in a proportion that reflected the observer's relative sensitivity to the two cues in that condition. If relative sensitivity is expressed as a ratio (k) of parallax/disparity thresholds, this means that a particular observer requires k times more parallax than disparity (in units of equivalent disparity) in order to detect the 3-D structure of the corrugation. The compound stimulus for this observer, in this condition, would, therefore, comprise k times more motion parallax than binocular disparity in order that both cues would reach threshold at approximately the same point as the amplitude of the compound is increased.

Results

In order to compare the thresholds from binocular disparity and motion parallax directly, the results from the latter conditions were again converted into units of equivalent disparity, as described above.

The relative sensitivity of each subject, at each spatial frequency was determined in the pilot experiment. These results, expressed as parallax/disparity ratios, are presented in Table 2. Ratios were similarly determined from the thresholds obtained in the main experiment and these are also presented in Table 2. Since these ratios provided the basis for the construction of the compound stimulus containing a particular proportion of binocular disparity and motion parallax, it is important to establish that relative sensitivity of each observer did not change systematically between the pilot and main experiments. The sensitivity ratios, collapsed over spatial frequency, did not differ significantly ($t = 1.546$; df = 2; $P > 0.05$). The possible effect of the small fluctuations on the thresholds for the compound stimulus is discussed below.

Figure 6 plots the thresholds determined in each

TABLE 2. The sensitivity ratios of parallax/disparity for each of the three observers in the three spatial frequency conditions.

	(a)			(b)		
	0.1	0.2	0.4	0.1	0.2	0.4
MFB	1.2	2.5	2.25	1.04	2.38	2.3
AG	2.16	2.43	3.29	1.87	2.52	2.5
BDB	1.62	2.59	2.71	1.80	2.59	2.4

(a) plots the ratios determined before the main experiment and used to construct the compound stimuli; and (b) plots the ratios determined during the experiment.

condition of the main experiment: binocular disparity, motion parallax and the compound stimulus. The compound stimulus is plotted relative to the magnitude of its most sensitive component—binocular disparity—but this choice is arbitrary.

To take the example of one observer (MFB) at one spatial frequency (0.2 cpd): the threshold for detecting 3-D structure from motion parallax was 7.6 arc sec; from binocular disparity it was 3.2 arc sec; and when both cues were present the threshold was 1.72 arc sec. If the two mechanisms involved in processing disparity and parallax were independent and thresholds were determined by a first-past-the-post rule then the threshold for the combined stimulus should be equivalent to its most sensitive component (in this case 3.2 arc sec). To put the results into perspective, the threshold for the compound stimulus corresponds to a depth difference between the peaks and troughs of the corrugations of less than 1/20 mm, at a viewing distance of 57 cm.

The psychophysical functions for the three stimulus conditions for the example described above are plotted in Fig. 7, together with best fitting cumulative gaussians from probit analysis. It can be seen that the compound stimulus produces a steeper slope than the binocular disparity condition, whereas the 50% point (the bias of the psychometric function) is not significantly affected.

Figure 8 plots the reduction in thresholds (averaged over observers) as ratios of disparity/(disparity + parallax). If the mechanisms are completely independent, the predicted ratio should be 1 because the thresholds for the most sensitive component and the compound stimuli containing that component should be the same (first-past-the-post rule). Figure 8 shows that the ratios are considerably higher than 1. The chi-square test was used to establish whether the reduction in thresholds was significantly greater than chance. The chi square was of the form:

$$\chi^2 = \frac{(obs - pred)^2}{\sigma^2} \qquad (1)$$

where *obs* was the obtained thresholds, *pred* was based on the model of independence (i.e., 1, or the threshold determined by a component presented alone) and σ^2 is an estimate of the precision of the thresholds determined by probit analysis. For each subject in each spatial frequency condition the reduction in thresholds was found to be highly significant ($P < 0.01$). The group mean ratios of disparity/(disparity + parallax), depicted in Fig. 8, were also significantly greater than 1 at each spatial frequency ($\chi^2 = 22.4$, $P < 0.005$; $\chi^2 > 1000$, $P < 0.005$; $\chi^2 > 1000$, $P < 0.005$; df = 2, from 0.1 cpd, respectively).

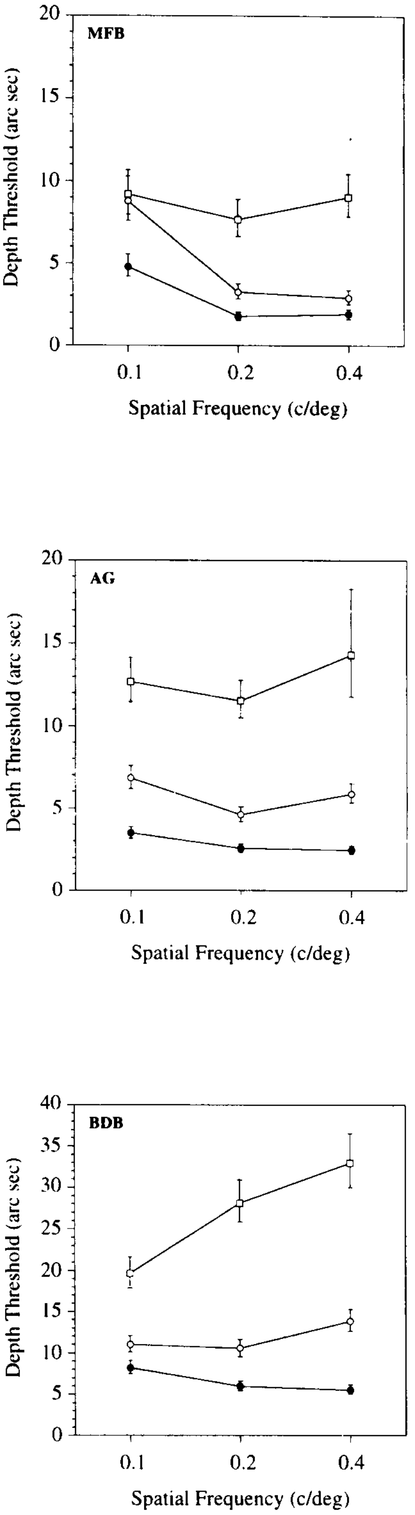

FIGURE 6. The thresholds for each subject plotted for each condition: parallax alone (open squares); disparity alone (open circles); and disparity plus parallax (solid circles).

Moreover, the magnitude of sub-threshold summation found in the present experiment averaged over subjects

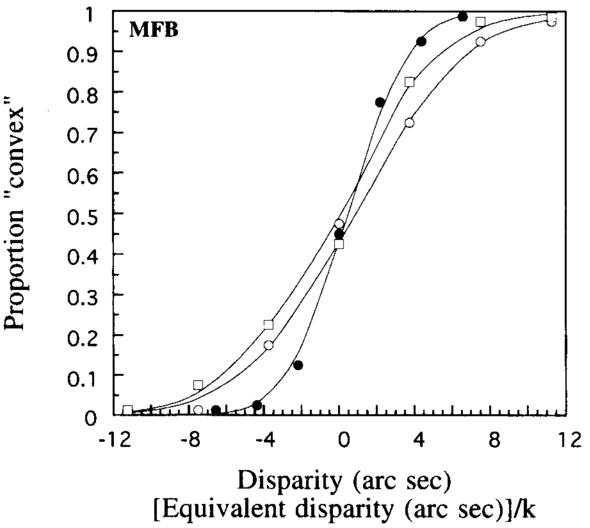

FIGURE 7. The psychophysical functions for depth discriminations based on disparity alone (open circles); motion parallax alone (open squares); and the compound stimulus: disparity and parallax (solid circles). The abscissa represents disparity or equivalent disparity (parallax) expressed in arc sec. The motion parallax data are normalised by the ratio of disparity and parallax sensitivity (k).

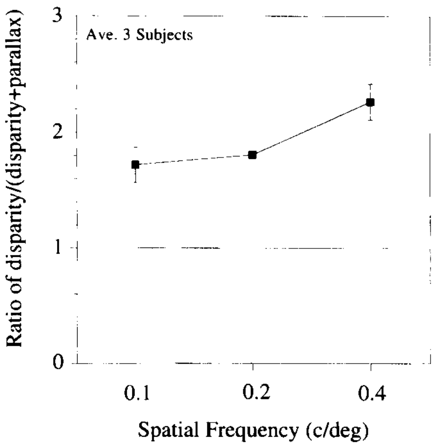

FIGURE 8. The ratio of disparity/(disparity + parallax) is plotted so that the magnitude of sub-threshold summation can be readily assessed. If the mechanisms were completely independent (i.e., thresholds determined by a first-past-the-post rule) this ratio should equal 1. The results indicate the mean and SE of three observers plotted for the three spatial frequencies.

and spatial frequency was 1.92 (48% reduction in thresholds) which is significantly greater than that predicted on the basis of linear summation, 1.41 ($\chi^2 = 15.51$; df = 2; $P < 0.05$).

Possible effect of probability summation

Probability summation refers to the fact that when more than one source of information is available for a detection judgement performance may improve, simply because that on any trial there are effectively two chances to detect the stimulus. That is, if one component of the compound stimulus is not detected there is still the chance that the other component will be. This improvement can occur even when the mechanisms are completely independent (see Graham, 1989). Therefore, the issue must be addressed here.

Probability summation, however, cannot account for the marked decrease in thresholds (48%) found in the present experiment. Its effect depends on the nature of the mechanisms, the probability distribution of responses and the decision rules of the system.

A simple model of probability summation based on the hypothesis of complete cue independence that we set out to test shows that it cannot account for our empirical results (see also Treisman, 1996). In the compound stimulus the relative amount of disparity (D) and motion parallax (M) was normalised so that the probability of a particular response was equivalent for both cues (i.e., both cues should reach threshold at the same time). Let a compound stimulus with value $D_i + M_i$ be presented such that D_i alone gives the probability of seeing a convex (cx) stimulus of P and M_i alone gives the same probability, P. That is:

$$P(cx_i) = p \text{ and } P(cx_i) = p. \quad (2)$$

The hypothesis of complete cue independence leads us to the assumption that a decision is made on each dimension separately and the results are combined in such a way that (i) when both dimensions give the same response, that response is chosen; and (ii) when opposite responses are given (convex from one and concave from the other) then, because we find no bias to choose one or the other, it is a matter of chance which is chosen. From this we get:

$$P(cx_i + M_i) = P(cx_i)P(cx_i) + 0.5[P(cx_i)(1 - P(cx_i))]$$

$$+0.5[P(cx_i)(1 - P(cx_i))] \quad (3)$$

$$= P^2 + 0.5[P(1 - P)] + 0.5[P(1 - P)]] \quad (4)$$

$$= P^2 + P - P^2 \quad (5)$$

$$= P. \quad (6)$$

Therefore, given the assumptions based on cue independence, there is no benefit from probability summation in the present design. This argument is extended by Treisman (1996) who develops more complex models of probability summation but shows that the form of the psychometric functions based on probability combination models is qualitatively different from the empirical functions established here (see Fig. 7). This concurs with the fact that no model of probability combination could account for an improvement in excess of full linear summation between the cues. The magnitude of sub-threshold interaction found in the present experiment was 1.92 which is larger than what would be expected on the basis of linear summation (1.41). Probability summation should never increase performance above linear summation.

Discussion

The results from the second experiment suggest strongly that we can reject the hypothesis of cue-independent mechanisms. Rather, they suggest that there are interactions between the mechanisms sensitive to binocular disparity and motion parallax prior to the decision about depth being made. In turn this implies that a common mechanism, sensitive to both disparity and parallax, may exist. Thresholds to detect the 3-D structure in the surface when both cues were available to the visual system were reduced on average by 48% (a ratio of 1.92).

If the two depth signals are combined in an additive fashion then an improvement of $\sqrt{2}$ (1.41) would be predicted, owing to the increase in the signal to noise ratio in the summed signal (see Campbell & Green, 1965). The improvement found in the present experiment was 1.92, which was found to be significantly greater than 1.41. This suggests that the combination of information from binocular disparity and motion parallax does not take place in a simple linear fashion. This was also borne out by the shape of the psychometric functions from the compound stimuli, which were qualitatively different from those based on probability combination (cf. Treisman, 1996). That is, the two signals are combined in a non-linear way which leads to facilitation when both cues are present.

A potential experimental artefact must be considered in this context. The process of cue normalisation in the compound stimulus, if not completely effective [i.e., the probability P in equation (2) was not the same in both cases] may have contributed to the observed decrease in thresholds for the compound stimuli. Table 2 presents the ratios of disparity–motion sensitivity on which the composition of the compound stimulus for each subject was based. These ratios were determined both before and during the main experiment. In all but one case these ratios changed slightly. Five of the ratios became smaller and three slightly larger. Recall that these fluctuations were not statistically significant. Nevertheless, this means that the relative proportions of disparity and parallax in the compound stimulus may not have reflected precisely the relative sensitivity of each observer to parallax and disparity during the main part of the experiment. Thresholds determined for the compound stimulus, therefore, which were defined in terms of its disparity component, may have been subject to a slight estimation error. However, it is possible to estimate the maximum net influence of this potential effect from the sensitivity ratios. It may have decreased the estimated thresholds for the compound stimuli by up to 3%. This is small relative to the size of the main effect (48%) and, therefore, cannot account for the observed decrease in thresholds, nor for the fact that the magnitude of the reductions exceeded that predicted by a simple linear summation model ($\sqrt{2}$). Therefore, we conclude that the slight fluctuations in the relative sensitivity to binocular disparity and motion parallax exhibited in some conditions cannot account for the reduction in thresholds.

GENERAL DISCUSSION

The two experiments reported in the present paper investigated whether the mechanisms which support the computation of depth from binocular disparity and motion parallax are independent. Taken together, the results suggest that the hypothesis of complete cue independence should be rejected. These results have important implications for models of cue combination and add further impetus to the investigation of the possible computational advantages which such mechanisms provide.

The major result of experiment 1 was that prolonged viewing of a stimulus defined by motion parallax can affect the detectability of a subsequently presented stimulus defined by binocular disparity (and vice versa). This is contrary to the view that the recovery of depth on the basis of motion parallax or binocular disparity is the product of independent processes. The difference in magnitude between the within- and between-cue adaptation conditions suggests that there may be a mechanism in the human visual system—a pool of neurons—which is sensitive to both binocular disparity and motion parallax, in addition to mechanisms tuned to each individual cue. The adaptation of this mechanism, which would also be stimulated by either binocular disparity or motion parallax alone, could account for the between-cue threshold elevation. It is difficult, however, on the basis of this experiment alone, to distinguish between this explanation and alternatives which could also account for the within- and between-cue effects. For example, the site of the aftereffect might be more central and result from the adaptation of mechanisms tuned to local depth variations and excited by disparity, parallax and/or other cues (Graham & Rogers, 1982a). However, it is the most parsimonious explanation when evidence from our second experiment is taken into account.

The second experiment established that there is substantial sub-threshold summation between motion parallax and binocular disparity prior to the computation of depth. Moreover, the interaction between the cues was found to be non-linear. This is consistent with the concept of strong fusion. The finding supports previous results which have also suggested that mechanisms may exist in the visual system, which support non-linear interactions between disparity and motion. In an ingenious adaptation experiment, Anstis & Duncan (1983) created separate monocular and binocular motion aftereffects. Their adaptation cycle consisted of three phases in alternation. In the first phase, clockwise motion was presented to the left eye, in the second phase clockwise motion was presented to the right eye, and in the third phase counterclockwise motion was presented to both eyes at the same time. (Note that during the adaptation period, equal and opposite motions were presented to each eye and so no aftereffects would be expected.) Anstis & Duncan (1983), however, found both strong monocular and binocular aftereffects. To account for these results they suggested that the visual system must possess three channels tuned to motion, two monocular and one binocular. Moreover, to account for the binocular aftereffect, which was evident despite the fact that each eye was exposed to equal and opposite motion during the adaptation phases of the experiment, they suggested that the response to binocular input must be non-linear. The monocular aftereffects were accounted for by making an additional assumption that the binocular channel inhibits the monocular channels during binocular stimulation. The characteristics of Anstis and Duncan's binocular motion channel are rather similar to those required to account for the results in the present experiments. Non-linear interactions have also featured in physiological findings (Pettigrew *et al.*, 1968; Bishop *et al.*, 1971; Cynader & Regan, 1978; Poggio & Talbot, 1981).

Taken together, the results from the two experiments reported in the present paper support the view that there is a mechanism which is sensitive to both binocular disparity and motion parallax information in the human visual system. This is contrary to the hypothesis of cue-independent mechanisms.

The composition of the mechanisms which link disparity and motion is presently the subject of further investigation in our laboratory. Three possible mechanisms, based on the manner in which the disparity and motion information is linked together, can be distinguished. The mechanisms may respond to the disparity of particular binocular motions, the movement of a feature with a particular disparity, or to either. That is, they may result from either two logical *AND* mechanisms and one logical *OR* mechanism. The first AND mechanism would result from binocular receptive fields which only respond if they are both stimulated by a pre-selected motion. Such a unit would respond to the disparity of motion-defined contours and would have similar properties to the cells found by Bradley *et al.* (1995) reviewed above. The second AND mechanism would result from the selective movement of a particular disparity derived from a surface feature. This type of mechanism could detect changes of disparity in space or time and its existence is supported by the fact that a MAE can be achieved from a disparity defined motion (Papert, 1964; Patterson *et al.*, 1992; Cumming, 1994). Finally, the OR mechanism would respond to either monocular motion or static binocular disparity (but may respond more vigorously when both cues are present). A potential benefit of this type of mechanism would be to increase the visual system's robustness to noise and so lower thresholds for depth detection. The exact nature of the mechanism involved in the perception of depth from relative disparity or relative motions, as established in the present paper, remains to be established. These models constitute three hypothetical mechanisms that may be involved in the encoding of disparity–parallax defined features.

An issue that arises in this respect is whether the putative disparity–parallax mechanisms are selective for the spatial frequency of depth modulation. Thresholds for detecting depth corrugations defined by disparity or motion show a marked dependency on the spatial frequency of depth modulation (Tyler, 1974; Rogers & Graham, 1982; Bradshaw & Rogers, 1993; Cobo-Lewis & Yeh, 1994). These sensitivity functions may reflect the envelope of separate narrowly tuned disparity or motion parallax channels or they could reflect a single broad-band channel in both domains. In the disparity domain there is evidence to suggest that the disparity sensitivity function is underpinned by several channels which overlap in spatial frequency sensitivity (Tyler, 1975; Schumer & Ganz, 1979). It would also be of interest to establish whether the between-cue threshold elevation and sub-threshold summation, found in the present experiments, generalises to the case in which the corrugations defined by disparity and parallax differed in their frequency of depth modulation.

In summary, the results of the present experiments suggest that the mechanisms which support the computation of depth from binocular disparity and motion parallax are not independent. These findings should add further impetus to research which addresses the computational advantages that such mechanisms provide.

REFERENCES

Anstis, S. M. & Duncan, K. (1983). Separate motion aftereffects from each eye and from both eyes. *Vision Research, 23*, 161–169.

Anstis, S. M. & Harris, J. P. (1974). Motion aftereffects contingent on binocular disparity. *Perception, 3*, 153–168.

Bishop, P. O., Henry, G. H. & Smith, C. J. (1971). Binocular interaction fields of single units in the cat striate cortex. *Journal of Physiology, 216*, 39–68.

Bradley, D. C., Qian, N. & Anderson, R. A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature, 373*, 609–611.

Bradshaw, M. F. & Rogers, B. J. (1993). Sensitivity to horizontally and vertically oriented stereoscopic corrugations as function of corrugation frequency. *Perception, 22*, 117.

Bradshaw, M. F., Rogers, B. J. & Frisby, J. P. (1991a). The combination of disparity and structure-from-motion information in human vision. *Perception, 20*, 90.

Bradshaw, M. F., Rogers, B. J. & Glennerster, A. (1995). Are binocular disparity thresholds affected by viewing distance? *Investigative Ophthalmology and Visual Science, 36*, 230.

Bradshaw, M. F., Rogers, B. J., Ono, H. & Ohmi, M. (1991b). Thresholds for perceiving (i) 3-D structure from motion parallax and (ii) 2-D relative motion as a function of dot lifetime. *Investigative Ophthalmology and Visual Science, 32, Supplement*, 831.

Bülthoff, H. H. & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America A, 5*, 1749–1758.

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

Clark, J. J. & Yuille, A. L. (1990). *Data fusion for sensory information processing systems.* Boston: Kluwer.

Cobo-Lewis, A. B. & Yeh, Yei-Yu (1994). Selectivity of cyclopean masking for the spatial frequency of binocular disparity modulation. *Vision Research, 34*, 607–620.

Cumming, B. (1994) Motion-in-depth. In Smith, A. & Snowden, R. (Eds), *The visual perception of motion* (pp. 334–366). London: Academic Press.

Cynader, M. & Regan, D. (1978). Neurons in cat prestriate cortex tuned to the direction of motion in three-dimensional space. *Journal of Physiology (London), 274*, 549–569.

Dosher, B. A., Sperling, G. & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research, 26*, 973–990.

Finney, D. J. (1971). *Probit analysis*, 3rd edn. Cambridge, U.K.: Cambridge University Press.

Graham, M. E. & Rogers, B. J. (1982a). Simultaneous and successive contrast effects in the perception of depth from motion-parallax and stereoscopic information. *Perception, 11*, 247–262.

Graham, M. E. & Rogers, B. J. (1982b). Interactions between monocular and binocular depth aftereffects. *Investigative Ophthalmology and Visual Science, 22, Supplement*, 272.

Graham, N. V. S. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Graham, N. V. S. & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channel models. *Vision Research, 11*, 251–259.

Johnston, E. B., Cumming, B. G. & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research, 34*, 2259–2275.

Julesz, B. (1960). Binocular depth perception of computer generated patterns. *Bell Systems Technical Journal, 39*, 1125–1162.

Koenderink, J. J. (1986). Optic flow. *Vision Research, 26*, 161–180.

Lagae, L., Maes, H., Raiguel, S., Xiao, D. & Orban, G. A. (1994). Responses of macaque STS neurons to optic flow components: A comparison of areas MT and MST. *Journal of Neurophysiology, 71*, 1597–1626.

Landy, M. S., Maloney, L. T., Johnston, E. B. & Young, M. (1995). Measurement and modelling of depth cue combination: In defense of weak fusion. *Vision Research, 35*, 389–412.

Maloney, L. T. & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. In Pearlman, W. A. (Ed.), *Visual communications and image processing IV, Proceedings of the SPIE*, 1199, pp. 1154–1163.

Maunsell, J. H. R. & van Essen, D. C. (1983). Functional properties of neurons in the middle temporal visual area of the macaque monkey, II: Binocular interactions and sensitivity to binocular disparity. *Journal of Neuroscience, 49*, 1148–1167.

Moulden, B. (1980). After effects and the integration of patterns of neural activity within a channel. *Philosophical Transactions of the Royal Society of London B, 290*, 39–55.

Nawrot, M. & Blake, R. (1991). The interplay between stereopsis and structure from motion. *Perception and Psychophysics, 49*, 230–244.

Orban, G. A., Lagae, L., Verri, A., Raiguel, S., Xiao, D., Maes, H. & Torre, V. (1992). First-order analysis of optical flow in monkey brain. *Proceedings of the National Academy of Science USA, 89*, 2595–2599.

Pantle, A. & Sekuler, R. (1968). Size detecting mechanisms in human vision. *Science, 162*, 1146–1148.

Papert, S. (1964). Stereoscopic synthesis as a technique for localizing visual mechanisms. *M.I.T. Quarterly Progress Report No., 73*, 239–244.

Patterson, R., Ricker, C., McGary, J. & Rose, D. (1992). Properties of cyclopean motion perception. *Vision Research, 32*, 149–156.

Pettigrew, J. D., Nikara, T. & Bishop, P. O. (1968). Binocular interaction on single units in cat striate cortex: Simultaneous stimulation by single moving slit with receptive fields in correspondence. *Experimental Brain Research, 6*, 391–416.

Poggio, G. F. & Talbot, W. H. (1981). Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *Journal of Physiology, 315*, 469–492.

Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America A, 2*, 343–349.

Rogers, B. J. & Collett, T. S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *Quarterly Journal of Experimental Psychology, 41A*, 697–717.

Rogers, B. J. & Graham, M. E. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8*, 125–134.

Rogers, B. J. & Graham, M. E. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research, 22*, 261–270.

Rogers, B. J. & Graham, M. E. (1984). Aftereffects from motion parallax and stereoscopic depth: Similarities and interactions. In Spillman, L. & Wooten, B. R. (Eds), *Sensory experience, adaptation and Perception: Festschrift Ivo Kohler* (pp. 603–619). Hillsdale NJ: Lawrence Erlbaum.

Roy, J. P., Komatsu, H. & Wurtz, R. H. (1992). Disparity sensitivity of neurons in monkey extrastriate area MST. *Journal of Neuroscience, 12*, 2478–2492.

Schumer, R. & Ganz, L. (1979). Independent stereoscopic channels for different extents of spatial pooling. *Vision Research, 19*, 1303–1314.

Tanaka, K., Fukada, Y. & Saito, H. (1989). Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology, 62*, 642–656.

Treisman, M. (1996). Combining sensory information: When do we get probability summation and when do we get probability averaging? Submitted to *Vision Research*.

Tyler, C. W. (1974). Depth perception in disparity gratings. *Nature (London), 251*, 140–142.

Tyler, C. W. (1975). Stereoscopic tilt and size aftereffects. *Perception, 4*, 187–192.

Waxman, A. M. & Duncan, J. H. (1985). Binocular image flows: Steps toward stereo-motion fusion. (Report CAR-TR-74) College Park, University of Maryland, Center for Automation Research.

Acknowledgements—The present work was supported by an Esprit Basic Research Grant 6019 and SERC (U.K.). The results described in this paper were presented at ARVO 1992 and 1993. We wish to thank Richard Eagle, Michel Treisman, Bart DeBruyn and Andrew Glennerster for their help in the preparation of this paper.