

## Performance data indicate summation for pictorial depth-cues in slanted surfaces

T. S. MEESE\* and D. J. HOLMES

*Neurosciences Research Institute, School of Life and Health Sciences, Aston University, Birmingham B47ET, UK*

Received 24 July 2002; revised 12 January 2003; accepted 13 January 2003

**Abstract**—Over recent years much has been learned about the way in which depth cues are combined (e.g. Landy *et al.*, 1995). The majority of this work has used subjective measures, a rating scale or a point of subjective equality, to deduce the relative contributions of different cues to perception. We have adopted a very different approach by using two interval forced-choice (2IFC) performance measures and a signal processing framework. We performed summation experiments for depth cue increment thresholds between pairs of pictorial depth cues in displays depicting slanted planar surfaces made from arrays of circular ‘contrast’ elements. Summation was found to be ideal when size-gradient was paired with contrast-gradient for a wide range of depth-gradient magnitudes in the null stimulus. For a pairing of size-gradient and linear perspective, substantial summation ( $> 1.5$  dB) was found only when the null stimulus had intermediate depth gradients; when flat or steeply inclined surfaces were depicted, summation was diminished or abolished. Summation was also abolished when one of the target cues was (i) not a depth cue, or (ii) added in conflict. We conclude that vision has a depth mechanism for the constructive combination of pictorial depth cues and suggest two generic models of summation to describe the results. Using similar psychophysical methods, Bradshaw and Rogers (1996) revealed a mechanism for the depth cues of motion parallax and binocular disparity. Whether this is the same or a different mechanism from the one reported here awaits elaboration.

*Keywords:* Vision; texture; size; contrast; gradient; discrimination.

### INTRODUCTION

As any student of vision is aware, there are many routes to the perception of depth (e.g. Gibson, 1950; Rogers and Graham, 1979; Howard and Rogers, 1995; McKee and Smallman, 1998). For example, it is fairly straightforward to produce compelling classroom demonstrations using binocular disparity or motion parallax where flat stimuli appear three-dimensional and solid. In the case of binocular

---

\*To whom correspondence should be addressed. E-mail: [t.s.meese@aston.ac.uk](mailto:t.s.meese@aston.ac.uk)

disparity, the term stereopsis (literally, 'solid seeing') is used to describe this perception. However, with some care, pictorial depth cues (those used by artists to convey an impression of depth) can produce similarly solid perceptions, particularly if stimuli are viewed monocularly in order to remove competing binocular cues to flatness. This has been notably well illustrated by the artwork of Vasarely, of which a compelling example is reproduced in Goldstein (1989). Further examples include those in Todd and Oomes (2002) and Young *et al.* (1993).

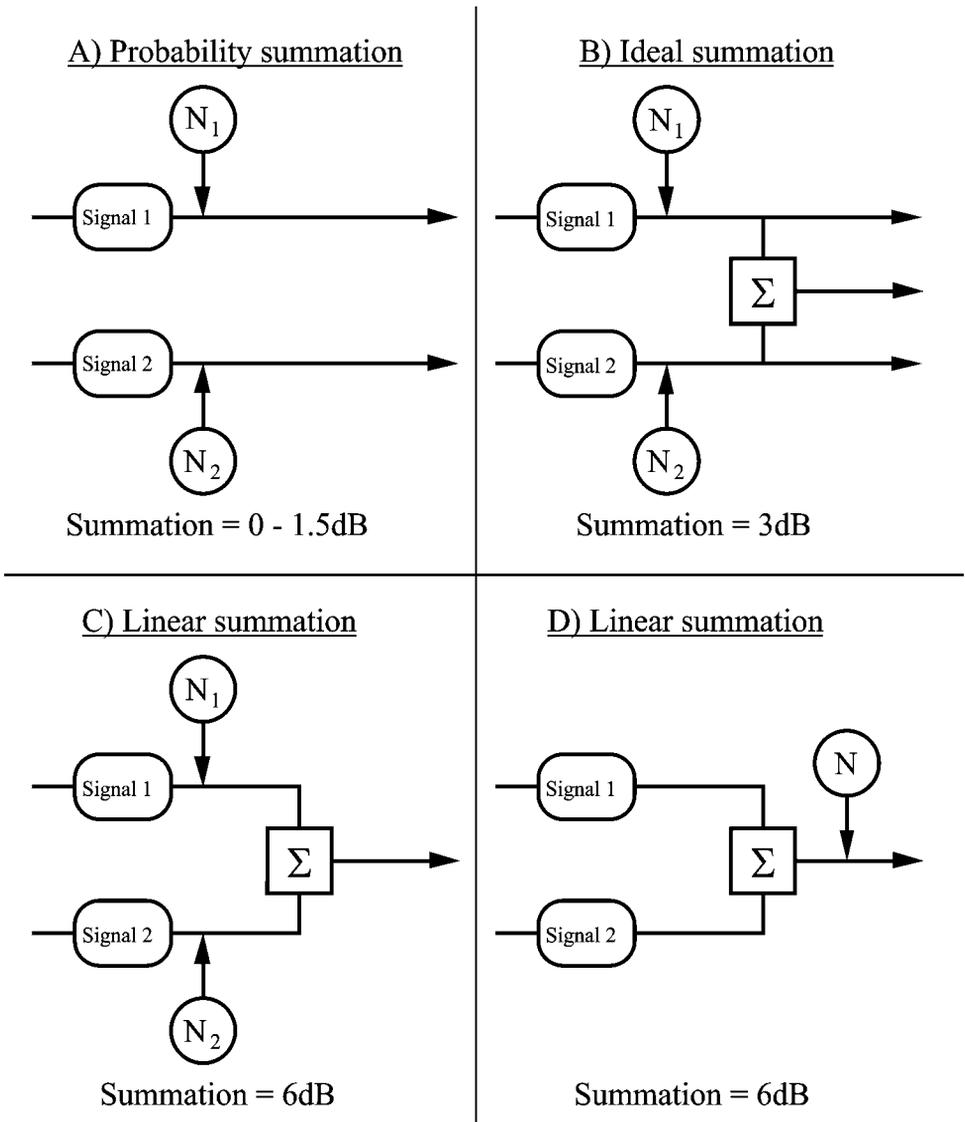
One widely studied class of pictorial depth cues is that of the texture-gradient. The importance of this variously classified collection of cues (e.g. size, foreshortening, density) in the perception of depth was emphasised in the seminal work by Gibson (1950) and has received considerable interest since, from both psychophysical and computational standpoints (Stevens, 1981; Epstein, 1981; Cutting and Millard, 1984; Todd and Akerstrom, 1987; Li and Zaidi, 2000; Todd and Oomes, 2002; Zaidi and Li, 2002). Other pictorial cues that contribute to perception of depth include linear perspective (Stevens and Brookes, 1988), height in plane (Bruno and Cutting, 1988), occlusion (Bruno and Cutting, 1988; Nakayama *et al.*, 1989), luminance contrast (Fry *et al.*, 1949; Troscianko *et al.*, 1991; O'Shea *et al.*, 1994; 1997), shading (Bülthoff and Mallot, 1988; Horn and Brookes, 1989), colour saturation (Troscianko *et al.*, 1991) and blur (Mather, 1996; O'Shea *et al.*, 1997).

Depth cues appear in such a variety of guises in the retinal image that it prompts the question of how they are combined to deliver a single unified percept of the world. In recent years this has been strongly emphasised in research on depth perception (Doshier *et al.*, 1986; Bruno and Cutting, 1988; Bülthoff and Mallot, 1988; Stevens and Brookes, 1988; Cutting *et al.*, 1992; Buckley and Frisby, 1993; Cumming *et al.*, 1993; Johnston *et al.*, 1993, 1994; Young *et al.*, 1993; Landy *et al.*, 1995). One consistent rule that has emerged from these studies is that cues are additive; each cue is assigned a weight (depending on prevailing circumstances) and the final computation of depth is the weighted sum of the cues (see Landy *et al.*, 1995 for further information and a review).

One feature of the cue combination studies above is that they have all used subjective perceptual judgements of the stimuli (typically using magnitude estimation or measuring a point of subjective equality<sup>1</sup>). This means that from the data alone it is not possible to determine whether summation was taking place in early sensory mechanisms (as often supposed), or whether higher level cognitive strategies were involved. A very different but complementary approach was taken by Bradshaw and Rogers (1996) who measured detection thresholds for stimuli in which depth was defined by motion parallax and/or binocular disparity. In an adaptation experiment, they found that adapting to one cue elevated threshold for the other, and in a subthreshold summation experiment, substantial summation was found between the two cues. The performance data from both of these experiments suggest a module for combining motion parallax and binocular disparity.

### Three models of summation

In the present study we used a summation paradigm to extend the work of Bradshaw and Rogers (1996) to investigate the combination of pairs of pictorial depth cues. We analysed our results in a signal processing framework for which three different possible outcomes are shown in Fig. 1. In Fig. 1a, two signals (different pictorial depth cues) are processed by independent modules, each with its own limiting source of (unit-variance, Gaussian) noise. As there are no connections between the



**Figure 1.** Summation models. (A) Independent signal detection and probability summation. (B) Ideal summation. (C and D) Formally equivalent versions of linear summation.

two signal lines, the presence of a second signal would not be expected to have an affect on the detectability of the first. In this case, either no summation is expected ( $0 \text{ dB}^2$ ) or, if observers are able to monitor all appropriate output lines effectively, a small amount of probabilistic summation might be expected (e.g. Graham, 1989). Calculations for this situation do not require a high threshold model as early work supposed (Robson and Graham, 1981) but are not trivial (Tyler and Chen, 2001) and are often approximated using Minkowski pooling with an exponent ( $k$ ) of four, sometimes referred to as fourth-root summation (Meese and Williams, 2000; Tyler and Chen, 2001). The expression is given in equation (1):

$$Sens_{\text{compnd}} = (\sum [Sens_i^k])^{1/k}, \quad (1)$$

where  $Sens_{\text{compnd}}$  is the sensitivity (reciprocal of increment threshold for the signal) to the compound stimulus and  $Sens_i$  is the sensitivity to the  $i$ th component. When the magnitudes of two components are adjusted so that they are equally detectable and component sensitivities are expressed in normalised units ( $Sens_1/Sens_i$ ), this predicts approximately 1.5 dB of summation (where summation =  $20 \cdot \log(Sens_{\text{compnd}}/Sens_1)$ ). In suprathreshold conditions, slightly higher values of  $k$  (i.e.  $> 4$ ) for combining multiple filter outputs were found to provide the best fits to contrast masking data (Watson and Solomon, 1997). This suggests slightly less summation than the fourth-root summation model.

Figure 1b is similar to Fig. 1a except that an additional output line is included which delivers the linear sum of the other two output lines. This model is often referred to as *ideal* because when two signals are present they are summed together, but when only one is present, one of the noise sources can be avoided by bypassing the summation device. To calculate the level of summation in this case we first make the simplifying assumption that observers select the most efficient output. (This requires complete knowledge of the stimulus, in the sense that the observer knows what the signal or signals are, and is consistent with our experimental design.) When both signals are present they are summed, but the variance of the two noise sources is also combined and so for normalised inputs the signal-to-noise ratio improves by 3 dB ( $\sqrt{2}$ ) over the single signal case (see Meese and Harris (2001) for further details). This is equivalent to Minkowski pooling with  $k = 2$ .

Figure 1c is similar to Fig. 1b except that the two independent output lines are not present. In this case, summation is compulsory and the two noise sources are combined regardless of whether one or both signals are present. Consequently, a second normalised signal improves the signal-to-noise ratio by a factor of two (6 dB) over the single signal case, and summation is said to be linear. This is equivalent to using a Minkowski exponent of one. A formally equivalent version of this model is shown in Fig. 1d, where the two noise sources are combined and this single source of limiting noise is placed after summation.

The models in Fig. 1b, c and d are sometimes described as physiological summation because of the putative neural substrate of the summation device. These

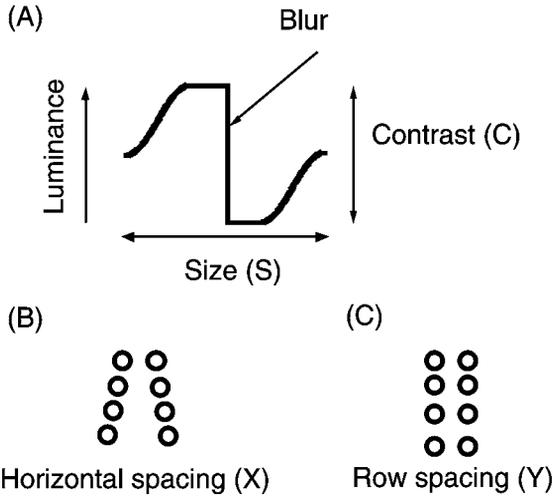
different levels of summation provide a useful gauge against which empirical estimates of psychophysical summation can be judged.

### *Comparison with weak fusion*

At first sight, the model in Fig. 1b bears some superficial similarity to the weak fusion model considered by Landy *et al.* (1995). In that model, the outputs of different depth cue modules are combined to produce a single estimate of depth. However, it is important to realise that the models in Fig. 1 will be used to address a very different data set from that considered by Landy *et al.* The data described by their models were reasonably assumed to reflect judgements about perception of depth (three-dimensional shape) and so the components in their weak fusion model were correctly depicted as multiple depth modules plus a resultant depth estimator at their combined output (or after ‘cue promotion’ in their modified weak fusion; see Landy *et al.*, 1995). Here, we started by making no such assumptions. We simply manipulated pairs of independent features of the stimulus and enquired whether observers were able to sum them using conventional discrimination performance methods of assessment. The stimulus features that we chose to manipulate were in fact depth-cues, but our task did not demand that observers use depth mechanisms; by being practised and receiving feedback, our observers were simply encouraged to perform the task as best they could, presumably using whatever mechanisms were most appropriate to optimise their performance. Feedback is an essential feature of experiments designed to investigate the limits of human ability, particularly when specific models are being tested (Sperling *et al.*, 1990). One consequence of all this is that the present experiment might have failed to reveal a depth cue combination mechanism simply because the task could have been performed without invoking that putative (noisy) mechanism. Indeed, as we discuss later in the paper, in the absence of data that address perceptual experiences (the way that things look, e.g. angle of surface slant), there are limitations to what performance data alone can tell us about the mechanisms involved in those experiences (the perception of depth being an example). On the other hand, if summation were found in the current experiments, this would imply the existence of summing circuitry for the depth cues in question. Such circuitry would quite reasonably be called a depth mechanism.

### *Pictorial depth cue stimuli*

The stimuli used in previous experiments investigating pictorial depth cues have fallen into several different classes. Depth judgements have been made between two or three flat planar surfaces (e.g. Bruno and Cutting, 1988; O’Shea *et al.*, 1994; Mather, 1996), on the shape of three-dimensional objects (e.g. Young *et al.*, 1993), on sinusoidal modulations of depth cues (Bradshaw and Rogers, 1996) and on the perceived angle of inclination of planar or curved surfaces (Cutting and Millard, 1984; Troscianko *et al.*, 1991). Before performing the present experiment we set two criteria for stimulus selection. First, we wanted to use stimuli that



**Figure 2.** Stimulus construction. (A) Luminance profile of stimulus element showing contrast and size parameters. Gradients of these parameters were applied to the two-dimensional array of stimulus elements. The element is shown with blur set to zero, though in the experiments a Gaussian blur with a standard deviation of one pixel was used. Gradients in blur were not used. (B) Column orientations (linear perspective) were controlled by a vertical spatial gradient in horizontal spacing between elements referred to as the  $x$ -gradient. (C) Proximity of rows was determined by the  $y$ -gradient.

could produce a compelling ('solid') impression of depth so as to enhance the likelihood that our experiment would tap a depth mechanism. Second, we wanted to be able to manipulate magnitudes of different depth cues straightforwardly and independently of each other. We were concerned that, without additional depth cues such as binocular disparity, pairs of planar surfaces depicted at different depth planes might not satisfy the first criterion. We had similar concerns regarding sinusoidal modulations. Of the remaining two alternatives, planar surfaces are the more straightforward to work with.

For these reasons we chose to use two-dimensional arrays of stimulus elements arranged to depict slanted planar surfaces. We have found that these can produce compelling impressions of depth, particularly when viewed monocularly, and that linear gradients of several pictorial depth cues can be manipulated independently (we consider this more fully later in the paper). Our basic stimulus is shown in Fig. 2a. The use of stimulus elements with no foreshortening (sometimes called compression) and vertical contours means that our main stimuli more closely resemble surfaces with vertically placed elements (such as blades of grass, bollards, trees and mountain goats) than a planar surface with a painted texture. In order to keep our stimuli and computations as simple as possible we chose to manipulate strictly linear gradients of depth cues (*cf.* Cutting and Millard, 1984). This was successful in producing stimuli that depicted slanted planar surfaces as illustrated by the selection of stimuli shown in Fig. 4. Finally, we might have decoupled

linear perspective cues from texture density cues by using randomly placed stimulus elements (see Cutting and Millard, 1984). But in the present experiments we chose not to do so for the sake of computational (experimental) speed. Hereafter then, we treat linear perspective and the vertical gradient of the horizontal packing density of stimulus elements as a single cue.

## METHODS

### *Equipment and stimuli*

Stimuli were displayed on a Sony Trinitron 20se II monitor under the control of a PC computer and using the framestore of a Cambridge Research Systems VSG2/4 stimulus generator operating in pseudo 15-bit mode.

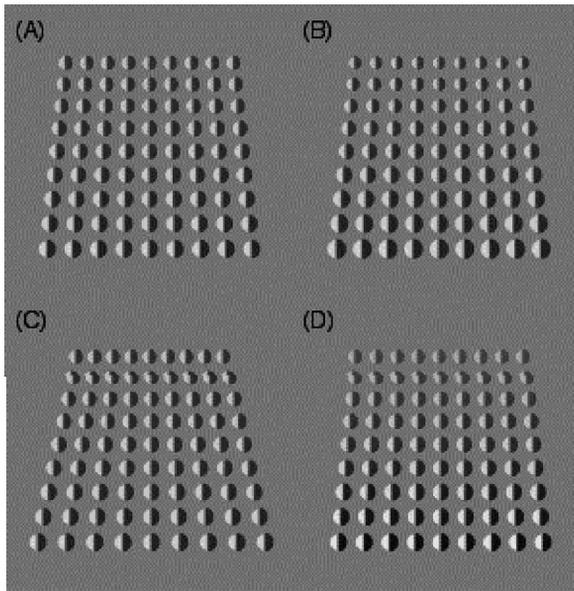
Stimuli had a mean luminance of  $60 \text{ cd/m}^2$  and were made from two-dimensional arrays of 81 stimulus elements, the basic square arrangement of which (Fig. 4a) was transformed in several different ways (see below). The luminance profile of the basic stimulus element is shown in Fig. 2a. It consists of a central luminance step-edge with a contrast of 50% ( $100 \cdot [L_{\max} - L_{\min}] / [L_{\max} + L_{\min}]$ , where  $L_{\min}$  and  $L_{\max}$  are minimum and maximum luminance values respectively), and a width of fourteen pixels plus two half cycles of a raised cosine function each two pixels wide. Thus, the middle element subtended a visual angle of 0.88 deg. In the basic square arrangement of elements (Fig. 4a), the centre-to-centre spacing of elements in both  $x$ - and  $y$ -dimensions was 24 pixels and the total stimulus width subtended a visual angle of approximately 10.2 deg.

In order to prevent observers using positional cues, the position of the entire set of elements was jittered by up to  $\pm 0.5$  of the width of the central element (i.e. up to  $\pm 9$  pixels) both vertically and horizontally on each stimulus presentation. To achieve subpixel accuracy, the step-edge of all elements was blurred by a Gaussian function with a standard deviation of one pixel.

To reduce competing cues to flatness, the stimuli were viewed monocularly (using an eye-patch over the non-preferred eye) through a tubular viewing tunnel, lined with matte black card. This was constructed to give a viewing distance of 57 cm through a circular viewing aperture of 17.5 deg. Observers placed their chin on a chin-rest and fixated a small fixation point in the centre of the display which was visible throughout the experiments.

### *Linear gradients of depth-cues*

Linear stimulus gradients ( $g$ ) are expressed in arbitrary percentage units and given by  $100(B_0 - B_{96})/B_0$ , where  $B_0$  was a parameter of the basic stimulus element in the centre of the array (5th row, 5th column), and  $B_{96}$  was the value of the same parameter for a (virtual) element whose centre was placed 96 pixels above that of the central element. (In the untransformed array in Fig. 4a, this was the distance



**Figure 3.** Examples of stimuli used in Experiment 1. (A) The null stimulus (with a null-gradient of 10%). (B, C and D) The same as (A) but with the addition of an easily detectable increment in size-gradient,  $x$ -gradient and contrast-gradient respectively. Note that the elements in the middle row were identical for all stimuli throughout the experiments.

between the element centres on the top and middle rows.) This defines a linear spatial gradient from the bottom to the top of the display monitor and from which depth-cue parameter values for all of the stimulus elements could be determined. Gradients in the spacing between elements in the  $x$ - and  $y$ -dimensions simulated perspective and texture density, and are referred to as  $x$ - and  $y$ -density gradients, or simply  $x$ - and  $y$ -gradients. The magnitude of the  $y$ -gradient controlled the vertical spacing between elements. In this case,  $B$  was the distance between the centre of one row and the centre of the row below it. The magnitude of the  $x$ -gradient controlled the horizontal spacing between elements and equivalently, the lengths of the element rows. For example, an  $x$ -gradient of 10% meant that the (virtual) rows 96 pixels above and below the middle row were 10% shorter and longer than the middle row respectively. Note, however, that the vertical transformation of row positions caused by non-zero values of the  $y$ -gradient meant that the actual top and bottom rows in the displays were typically not 96 pixels from the centre, meaning that their actual lengths were not given straightforwardly by the magnitude of  $g$  (see Fig. 3).

Note that all four depth-cue gradients ( $x$ ,  $y$ , size and contrast) could be manipulated independently of each other; changing any one gradient had no impact on the gradient of any other parameter (see Fig. 3). However, because the  $y$ -gradient controlled the relative positions of element rows it did affect other stimulus statistics (e.g. the size of the smallest and largest elements in the display). Because of

this we did not attempt to explore summation of the  $y$ -gradient with other depth cues. Nevertheless, this cue was present in most stimuli in an attempt to enhance the perception of depth (Landy *et al.*, 1995).

In all experiments, the  $x$ -,  $y$ -, size- and contrast-gradients of the null stimuli were equated and referred to as the null-gradient. In Experiments 1 and 2, the null-gradient was 10% (Fig. 3a) and in Experiment 3 it was 0%, 10%, 20% and 30% (Fig. 4).

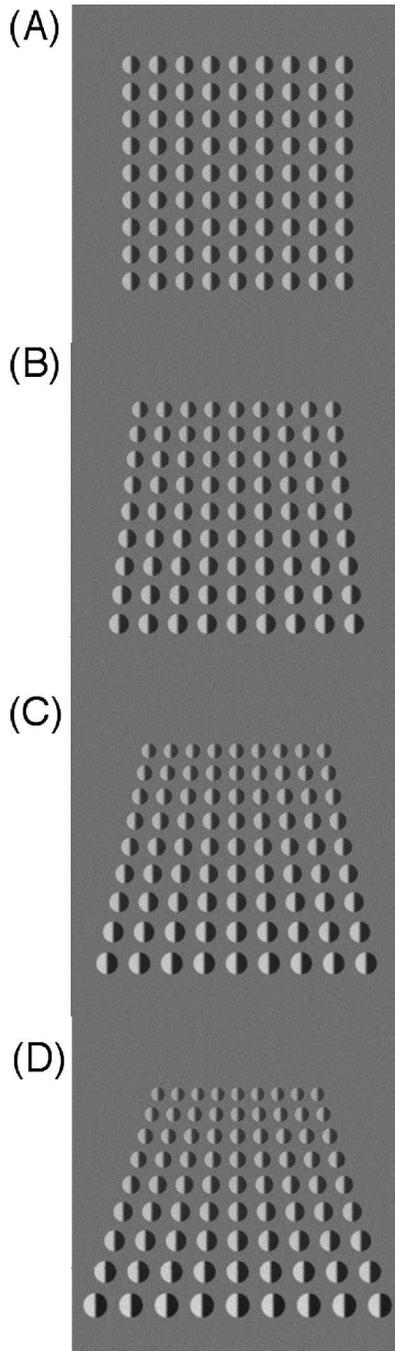
In a series of control conditions, summation was measured between depth cues and an orientation-gradient. In these conditions, an orientation gradient was included in the null stimulus and is shown in Fig. 5. In this case,  $B_0$  was vertical and  $B_{96}$  had an orientation of 10 deg. Increments in orientation gradient were applied to the test stimuli in the same way as for the depth cues and expressed as the difference between  $B_{96}$  in the null and test intervals.

### Procedure

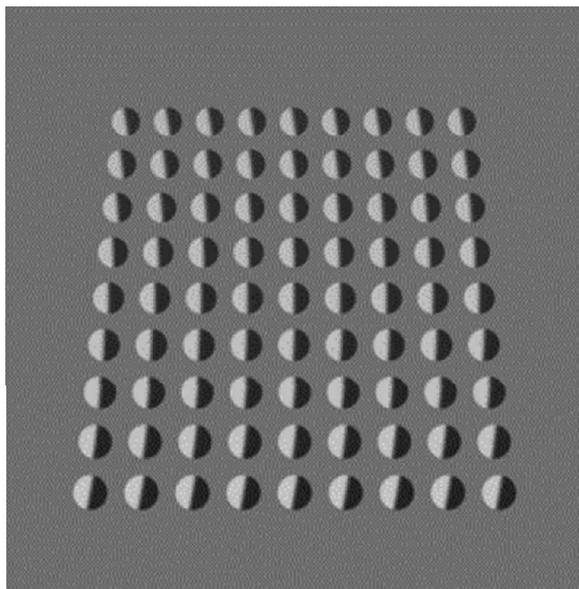
Thresholds (75% correct) were measured using pairs of interleaved staircases and probit analysis in a two-interval forced-choice (2IFC) paradigm. One of the stimuli was always a null stimulus (e.g. Fig. 4) and the other was the same, but with an increment in the gradient ( $\Delta g$ ) of one or two depth cues. Increments were selected by a '3-down, 1-up' staircase procedure (Wetherill and Levitt, 1965) from those placed evenly along a logarithmic scale. The initial 'step-size' was 0.45 log units, and this was reduced to 0.15 log units after the first 'reversal'. Each staircase terminated after 10 subsequent reversals and probit analysis was performed on the data from the (approximately 80–100) trials collapsed across the two staircases after the second reversal for each staircase. The observer selected the interval thought to contain  $\Delta g$  by pressing one of two mouse buttons and was given auditory feedback regarding the correctness of response. Each experimental session measured summation for a single pair of depth cues and consisted of three sequential stages: (1) threshold estimation for one of the cues ( $t_1$ ), selected at random, (2) threshold estimation of the other cue ( $t_2$ ) and (3) threshold estimation of a compound stimulus ( $T_1, T_2$ ) containing normalised increments of the two cues. Specifically, in the third stage, the level of the second cue was ( $t_2/t_1$ ) times that of the first cue in order that the two cues were equally detectable. This whole procedure was repeated at least five times for each pair of depth cues. As a matter of convenience, summation is expressed in dB, and for each session is given by  $20 \cdot \log(t_1/T_1)$ . Results are plotted as the means and standard errors of these measures.

### Observers

The two authors (TSM and DJH) served as observers and both wore their normal spectacle correction. Prior to formal data collection, the two observers had many



**Figure 4.** Null stimuli used in Experiment 3. From top to bottom the gradients are 0%, 10%, 20% and 30%.



**Figure 5.** Null stimulus for the control conditions in Experiment 1. This stimulus is identical to that in Fig. 3a, but with the addition of an orientation gradient.

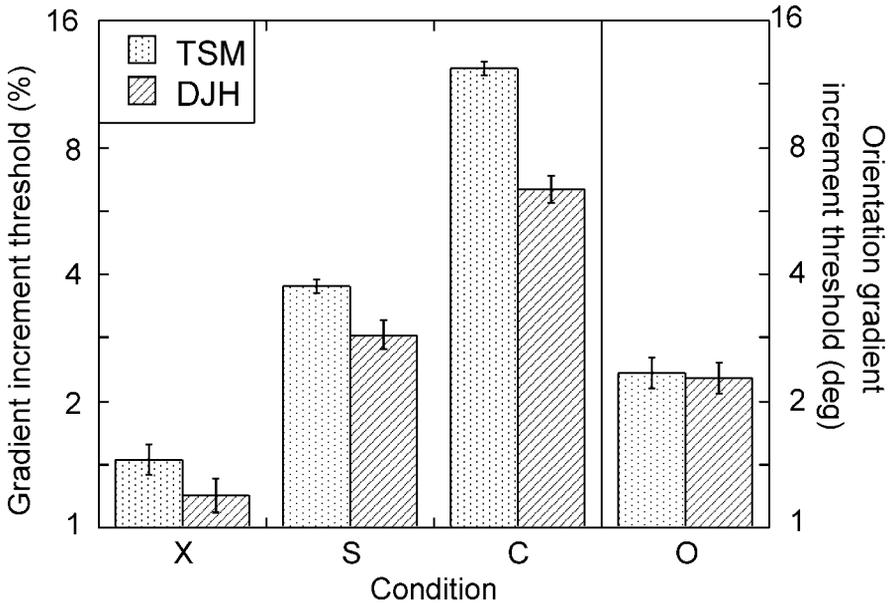
hours of practice (spanning several months) with stimuli that were either identical or similar to those used in the experiments.

## RESULTS AND DISCUSSION

### *Experiment 1: Summation for depth-cue gradients*

Sensitivities to the individual components are shown for both observers (different shading) in Fig. 6. The pattern of results is very similar for the two observers, though DJH was consistently more sensitive than TSM.

The summation results are plotted in Fig. 7 (see figure caption for statistical analysis). Summation (in dB) is shown for the three pairings of depth cue gradients (test conditions) and the three pairings of orientation-gradient with the depth-cue gradients (control conditions). Predictions are shown for no summation (0 dB), fourth-root summation (1.5 dB), ideal summation (3 dB) and linear summation (6 dB) (see Introduction). For both observers, summation was close to zero or fourth-root summation in the control conditions suggesting an absence of physiological summation for these gradient pairs. In the test conditions there were some differences between the two observers. TSM showed substantial summation (slightly greater than ideal) for all three pairings of depth cues, whereas DJH showed substantial summation (close to linear) for only the size- and contrast-gradient pairing. Note that the pattern of sensitivities to the individual components (Fig. 6) offers no clue to the origin of these differences. We shall return to these individual

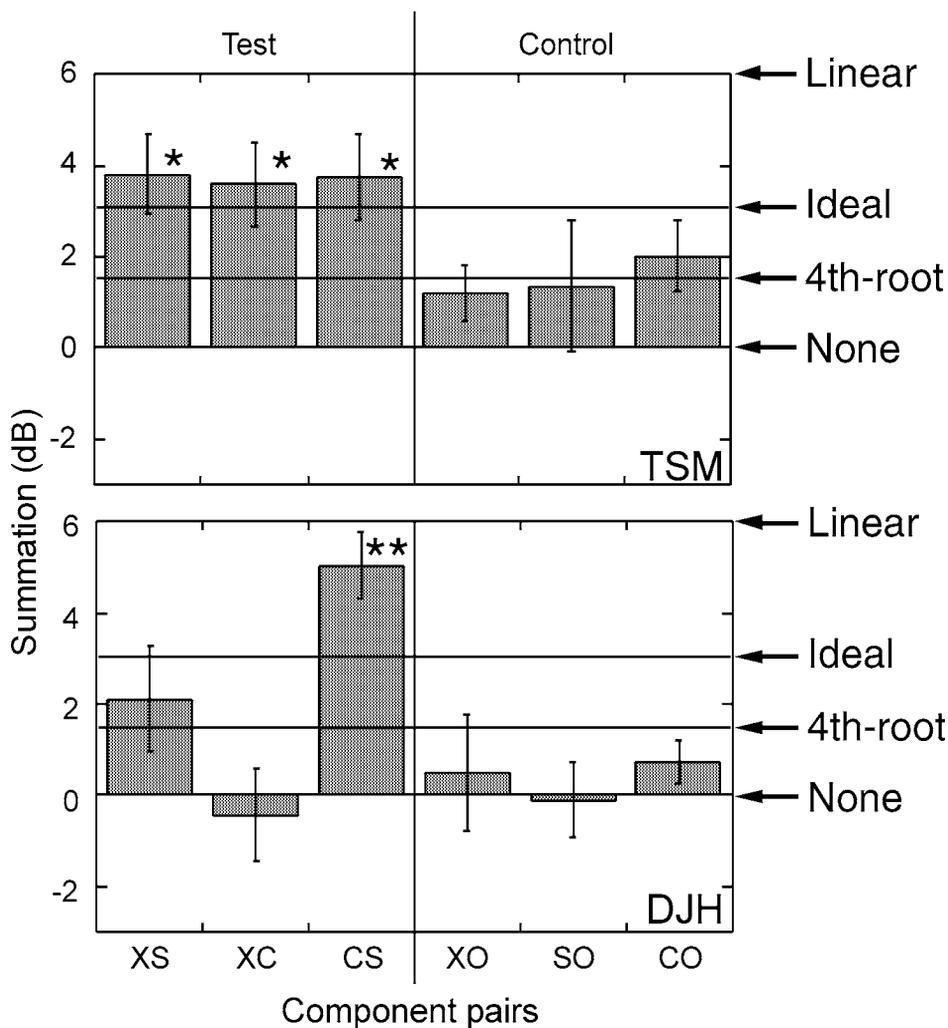


**Figure 6.** Gradient increment thresholds for Experiment 1 (error bars show  $\pm 1$  SE). Results are for the four different stimulus components averaged across each of the components with which it was paired. Note the different units used for orientation-gradient (right ordinate) and for  $x$ -, size- and contrast-gradients (left ordinate). Different shading is for different observers.

differences in Experiment 3, but first we consider and reject two simple explanations of our results that are not concerned with depth cues.

#### *Row-length discrimination does not explain summation*

For TSM there was substantial summation between size- and  $x$ -gradients. In principle, this could have taken place within a mechanism that simply performed row-length discrimination. For example, consider how the following transformations affect the top row of stimulus elements. An increment in  $x$ -gradient reduces its length (compare Fig. 3a and c). An increment in size-gradient does not change the centre positions of the stimulus elements, but does reduce the length of the row when measured to the outside edges of the elements because the elements are smaller (compare Fig. 3a and b). In other words, increments in  $x$ - and size-gradients physically sum to change the length of the rows, but is this physical summation of sufficient magnitude to account for the psychophysical summation? At discrimination thresholds for  $x$ -gradient and size-gradient ( $\Delta g = 1.5\%$  and  $3.7\%$ , respectively; see Fig. 6) the length of the top row decreases by  $1.63\%$  and  $0.36\%$ , respectively. This would give  $1.7$  dB of summation for a mechanism that measures row length and is considerably less than the  $3.7$  dB of summation seen in the data. Furthermore, the implicit Weber fraction for row-length discrimination of  $1.63\%$  is less than that usually found for angular (size) judgment tasks which are typically between  $2\%$  and  $4\%$  (Ono, 1967; Wilson and Gelb, 1984; Burbeck, 1987; Norman *et al.*, 1996;



**Figure 7.** Summation results for Experiment 1 (error bars show  $\pm 1$  SE). Compound stimuli contained component pairs of  $x$ -gradient and size-gradient (XS);  $x$ -gradient and contrast-gradient (XC) and contrast-gradient and size-gradient (CS) (left columns) and orientation-gradient paired with  $x$ -gradient (OX), size-gradient (OS) and contrast-gradient (OC) (right columns). The model predictions are described in the text. Different panels are for different observers. A single asterisk (\*) indicates that the experimental summation was significantly greater than 4th-root summation and a double asterisk (\*\*) indicates that it was significantly greater than ideal summation (one tailed  $t$ -test;  $p < 0.05$ ;  $3 < df < 7$ ). Two-tailed  $t$ -tests confirmed that there was no significant difference between the results from any of the control conditions and fourth-root summation.

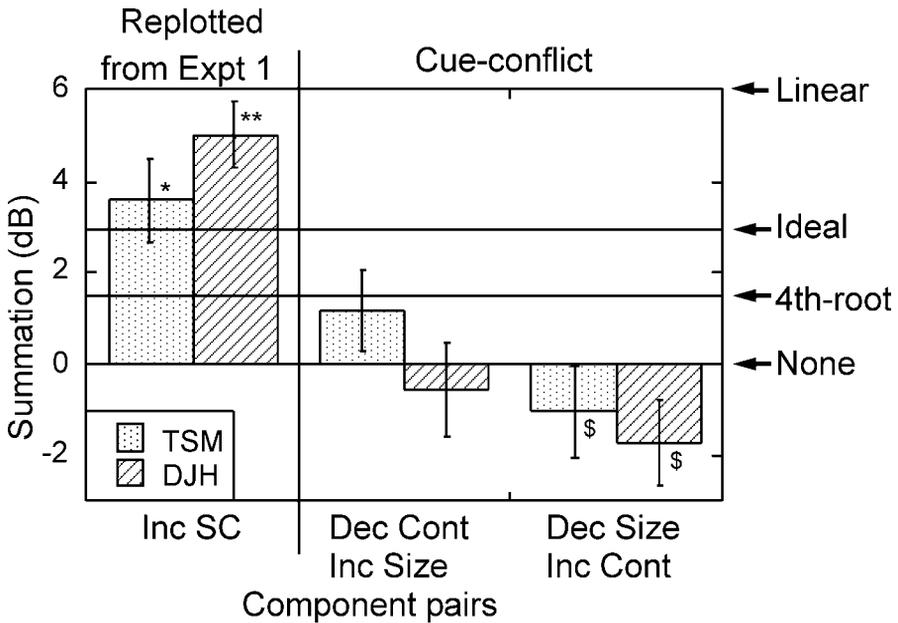
McKee and Smallman, 1998). Note that these figures are for stimuli measured in the high acuity region of the central fovea, whereas the row-length discrimination figure for the present study is derived from a stimulus placed about 5 deg away from the fovea where spatial discriminations are typically less fine (Wilson *et al.*,

1990; Greenlee, 1992). Thus, a putative row-length mechanism would need to be more sensitive than conventional estimates allow and, in any case, physical summation within such a mechanism is insufficient to account for the psychophysical data. Another possibility is that performance is governed by a mechanism sensitive to column orientation. An  $x$ -gradient increment threshold of 1% (slightly less than the data, see Fig. 6), equates to an orientation change of  $0.57^\circ$  which is within the region that might be expected for orientation discrimination (e.g. Regan and Beverley, 1985). However, with regard to summation, the consequences of the geometry are much the same as for the putative row-length mechanism making this an equally improbable account. A more likely explanation is that the observer was using a mechanism that integrated spatial information over much of the display. That this mechanism was selective for spatial cues involved in depth perception (i.e. unselective for the orientation-gradient cue) suggests that this mechanism might be a (pictorial) depth cue mechanism.

### *Experiment 2: No summation between depth cue increments and decrements*

Both observers showed substantial summation for increments in size- and contrast-gradients. When a size-gradient increment was applied to the stimulus, the bottom row of elements increased in size, and so one possible explanation of the data is that observers were more sensitive to contrast increments when the stimuli were large. Certainly, this is true at contrast detection threshold (Robson and Graham, 1981), though the situation is less clear for contrast discrimination (Legge and Foley, 1981; Bonnef and Sagi, 1999; Meese *et al.*, 2001). To test this possibility we performed the experiment again but replaced the contrast-gradient increments with decrements (in both stages of the experiment). If the above explanation were correct, this should not affect the results because if contrast discrimination is influenced by element-size it should be equally so for increments and decrements. However, if the results were due to physiological summation within a module for processing depth cues, then psychophysical summation should be diminished because the two cues change inconsistently with changes in depth information. It might even result in negative summation (i.e.  $< 0$  dB) because if performance in Experiment 1 were at its most efficient and driven exclusively by a depth mechanism (e.g. the linear summator of Fig. 1b), then the conflict would require that a less efficient mechanism should mediate performance for the new compound stimulus. The stimulus arrangement in this general situation is sometimes referred to as cue-conflict. As a further test of its predictions for depth cue combination we also performed the experiment with the decrements applied to size-gradients and the increments applied to contrast-gradients. In all cases, the null stimulus was as before (Fig. 3a).

Results for the cue-conflict conditions are shown in the right part of Fig. 8 (see figure caption for statistical analysis). There are two main observations. First, summation was always less for the decrement conditions than for the corresponding increment condition (left columns, replotted from Fig. 7). Second, summation was always less than the fourth-root summation prediction, and in three cases was

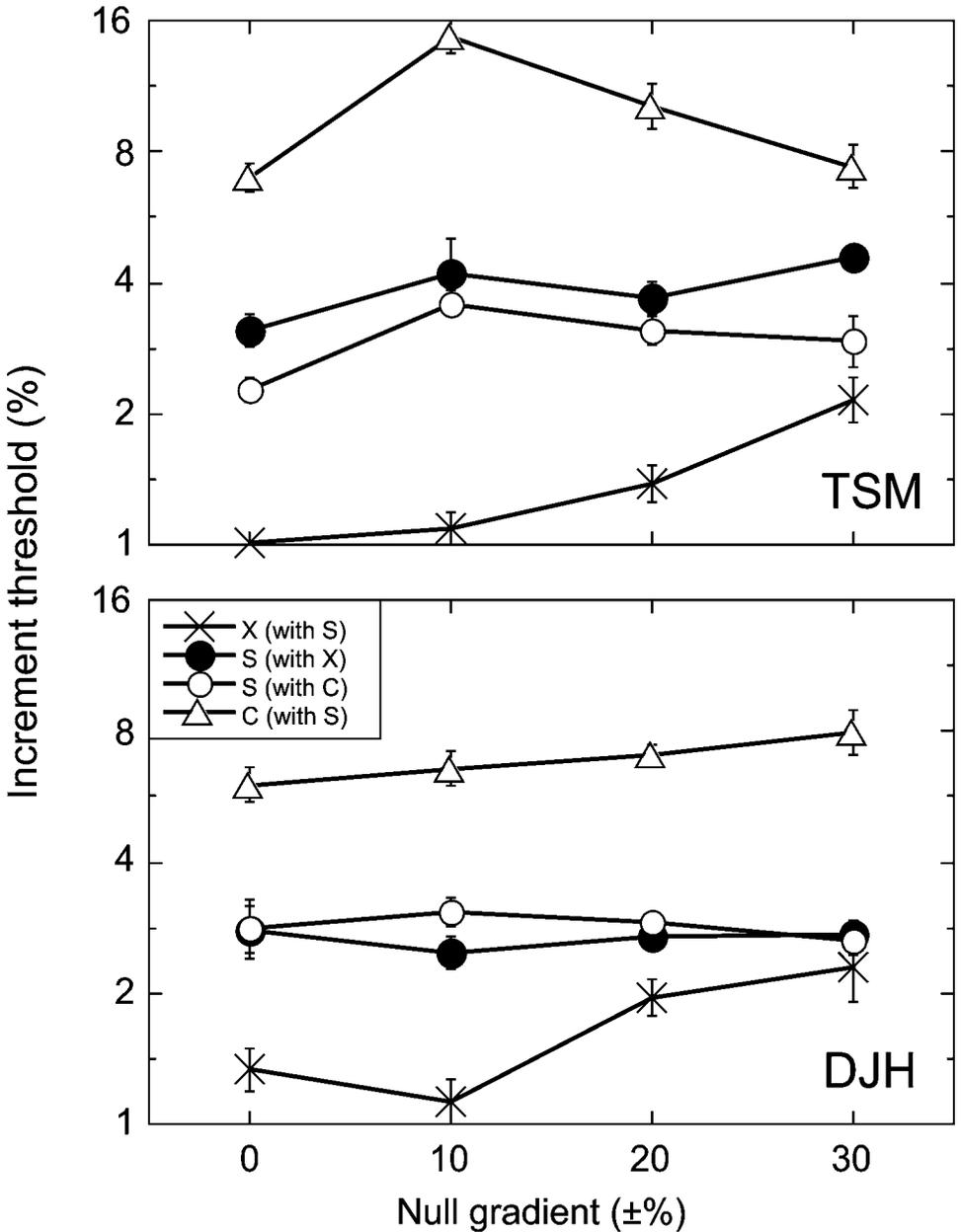


**Figure 8.** Summation results for Experiment 2 (error bars show  $\pm 1$  SE). The data in the left part of the figure are for increments in size- and contrast-gradients and are replotted from Fig. 7 (see Fig. 7 caption for meaning of \* and \*\*). The remaining data are for stimuli in which decrements were applied to one of the depth-cues and increments to another (cue-conflict). For the middle columns, the decrements were applied to the contrast-gradient and for the right columns, the decrements were applied to the size-gradient. Different shading is for different observers. The dollar symbols (\$) indicate that the experimental summation was significantly less than fourth-root summation (one tailed  $t$ -test;  $p < 0.05$ ;  $df = 4$ ). In no case was it significantly less than zero. For both observers, the results from both of the cue-conflict conditions were significantly less than those from the other condition (one tailed  $t$ -test; TSM: Dec Cont,  $p = 0.04$ ,  $df = 8$ ; Dec Size,  $p = 0.004$ ,  $df = 9$ ; DJH: Dec Cont,  $p = 0.002$ ,  $df = 8$ ; Dec Size,  $p = 0.0002$ ,  $df = 8$ ).

less than zero (though not significantly so) indicating that performance for the compound stimulus tended to be worse than for either of the components. The first observation indicates that the summation from Experiment 1 (left columns) was not due to within-element effects. The second observation suggests that when the two cues were added with opposite sign this interfered with performance. As explained above, this was to be expected if observers were using perception of depth (e.g. surface slant) to drive performance.

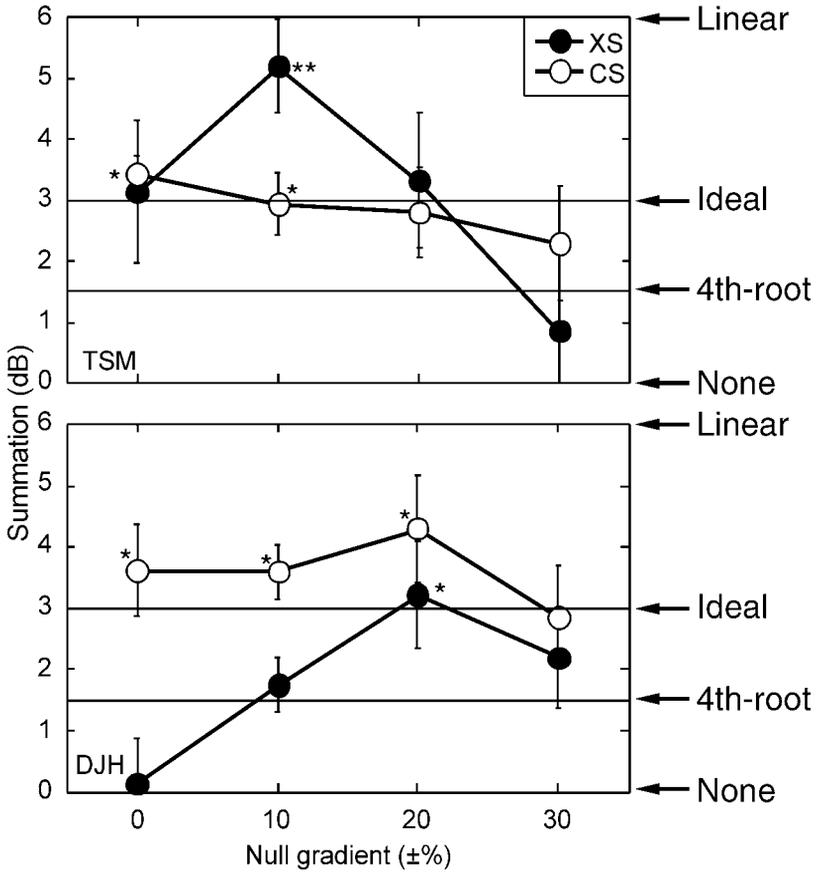
### *Experiment 3: Effects of gradient magnitude in the null stimulus*

In Experiment 3 we extended the work of Experiment 1 by manipulating the magnitude of the depth-gradients in the null interval. We chose to do this for pairings of (i) size- and contrast-gradients, because both observers showed substantial summation in this condition (see Fig. 7) and (ii)  $x$ - and contrast-



**Figure 9.** Gradient increment thresholds for Experiment 3 (error bars show  $\pm 1$  SE). Different symbols are for different conditions. Note, size-gradient (circles) is shown twice because it was paired with both  $x$ -gradient (filled circles) and contrast-gradient (open circles). Different panels are for different observers.

gradients, because DJH showed only a hint of summation in this condition and we wanted to know whether it would be possible to reveal more.



**Figure 10.** Summation results for Experiment 3 (error bars show  $\pm 1$  SE). Stimulus pairs were  $x$ - and size-gradient (filled circles) and contrast- and size-gradient (open circles). The null-gradient of 10% repeats some of the conditions of Experiment 1. Different panels are for different observers. A single asterisk (\*) indicates that experimental summation was significantly greater than 4th-root summation and a double asterisk (\*\*) indicates that it was significantly greater than ideal summation (one tailed  $t$ -test;  $p < 0.05$ ;  $3 < df < 10$ ).

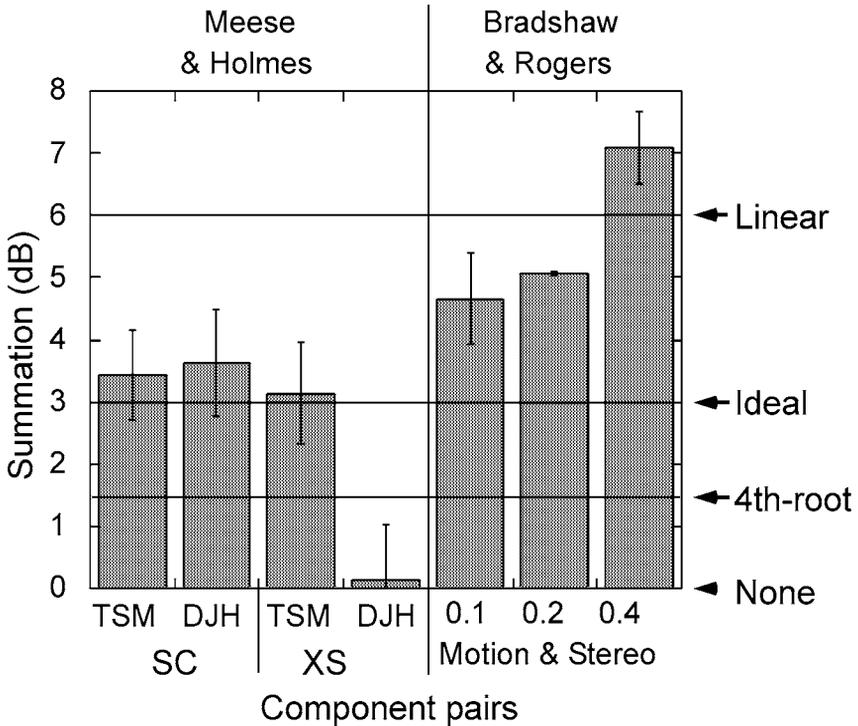
Gradient increment thresholds are shown in Fig. 9. The pattern of results is similar for the two observers (different panels), though curiously, TSM was consistently more sensitive to the size-gradient in sessions where thresholds were also measured for contrast-gradient (open circles) than in sessions where thresholds were also measured for the  $x$ -gradient (filled circles). As the stimuli were identical in these two cases, higher level processes such as attention might have been involved for this observer, though why or how remains unclear. Certainly, the subjective reports from this observer did not shed any light on the situation. We do note, however, that the effect is small and inconsistent across observers; where the effect is present for DJH it is in the opposite direction to that for TSM.

Summation results for this experiment are shown in Fig. 10. Although there are clear individual differences (different panels), there are several features of the data

that are consistent across the two observers. First, summation for contrast- and size-gradients (open symbols) is fairly insensitive to the gradient in the null stimulus and is around 3 dB (ideal). Second, summation for  $x$ - and size-gradients is greatest for intermediate null gradients declining, in some cases substantially, at higher and lower values. Note, however, that as in Experiment 1, the component detection data offer no obvious clue towards the origins of the variation in the summation data (Fig. 9).

### Comparison with Bradshaw and Rogers

Bradshaw and Rogers (1996) performed a summation experiment analogous to ours, but used random dot stimuli containing sinusoidal modulations of motion parallax and binocular disparity. We have replotted their data in Fig. 11 for their three different spatial frequencies of modulation (0.1, 0.2 and 0.4 c/deg). As their null stimulus always had zero depth of modulation (i.e. no depth cues), the most appropriate comparison with our own data is the condition in which the null stimulus had a gradient of 0%. These data are also replotted in Fig. 11. Overall, there was more summation for Bradshaw and Rogers stimuli than there was for ours, though at



**Figure 11.** Comparison of results from the present study and those of Bradshaw and Rogers (1996) (error bars show  $\pm 1$  SE). In both cases, the null stimulus contained no depth cues. The data for TSM and DJH are replotted from the rightmost symbols in Fig. 10. Those for Bradshaw and Rogers are for conditions in which the spatial frequency of their sinusoidal modulations were 0.1, 0.2 and 0.4 c/deg.

the lower spatial frequencies (0.1 and 0.2 c/deg) the differences are not large. In both studies, conditions were found in which summation was between 3 dB and 6 dB. We note, however, that Bradshaw and Rogers did not use feedback in their task. This can lead to observers underperforming (Sperling *et al.*, 1990) though whether this is responsible for the differences in summation between the two studies is not clear.

## GENERAL DISCUSSION

The work in this paper is presented in the spirit of a first attempt at using established empirical and analytic methodologies to provide fresh insight into the problem of depth cue combination. Our approach has rarely been used before in depth perception and because its assumptions and analytic framework are different from more direct but subjective methods (e.g. Landy *et al.*, 1995), it offers a complementary perspective. In this preliminary work we have demonstrated that the approach can be effective when using stimuli that depict a (slanted) planar surface and contain only pictorial depth cues.

### *Summary of results*

We found substantial summation (= 3 dB and statistically  $> 1.5$  dB) for several pairings of the depth-cues (perspective ( $x$ -gradient), size-gradient, and contrast-gradient). Summation was abolished or diminished: (i) when one of the cues was not a depth cue (Experiment 1), (ii) when cues were added in conflict (Experiment 2) and (iii) for extreme values of the null gradient for a cue pair of  $x$ -gradient and size-gradient (Experiment 3). We ruled out two explanations of our results in terms of physical summation and mechanisms that act on individual columns, rows or elements and suggest that our results imply a mechanism that is able to perform summation between the independent depth cues. Generally, less summation was found for our pictorial cues than was found by Bradshaw and Rogers (1996) for motion parallax and binocular disparity. Whether the mechanism inferred from our own results is the same or different from that revealed by Bradshaw and Rogers awaits further investigation.

### *Operation of the summation device*

Although the experiments in this paper imply the presence of a summation device such as those shown in Fig. 1, performance data are unable to reveal its purpose. Nevertheless, as the components being summed are pictorial depth cues it is reasonable to describe this mechanism as a (pictorial) depth cue mechanism. As the magnitude of pictorial depth cue gradients relate to the magnitude of surface slant, it is natural to consider the possibility that the summation device is involved in delivering an estimate of slant, or at least some related measure of depth. Certainly, this is the essence of the type of computation delivered by Landy *et al.*'s (1995)

(modified) weak fusion model. If this were so, then at least for some conditions, psychophysical performance in the present experiments might have been driven by the magnitude of slant delivered by the depth mechanism. However, there is at least one other possibility. We have noticed that manipulating the depth cues in our planar stimuli not only changes perception of slant magnitude, but can also affect the extent to which a surface is seen as (i) a compelling representation of a solid three-dimensional surface ('solid mode'), (ii) a two-dimensional representation of a three-dimensional surface ('pictorial mode') or (iii) a two-dimensional representation of a flat two-dimensional surface ('flat mode')<sup>3</sup>. For example, if one of the cues becomes 'too large', instead of increasing the perceived angle of slant, perception degenerates to a flat representation of a set of circular elements, presumably because the magnitude of one cue is so inconsistent with those of the other cues. At present it is not clear whether these subjective classifications represent distinct modes of operation or form part of a continuum, but either way they could contribute to the decision variable. Furthermore, evidence from elsewhere suggests a potential association between the mode of perception and the mode of summation. When stimuli are seen as three dimensional surfaces in depth ('solid mode') it can be difficult to access the two-dimensional component features of the surface (Burbeck, 1987; He and Nakayama, 1994; Davis and Driver, 1998; also see: McKee and Welch, 1992; Bennett and Cortese, 1996; and see Mack, 1978 for some counter examples) in which case vision might operate in 'solid mode' and summation would be linear (Fig. 1c). An intermediate situation arises when vision is in 'pictorial mode'. In this case, observers have 'dual perception' (Mamassian and Landy, 1998) and can make judgements about both the distal and proximal stimuli, though cross-talk means that the judgements are not veridical (Sedgwick and Nicholls, 1993). The ideal summator of Fig. 1b bears some similarity to this situation. When vision is in 'flat mode', no depth is seen and observers perform according to models that operate on basic retinal descriptions (Wilson and Gelb, 1984; Burbeck, 1987). In this case, little or no summation might be expected.

The conjecture that observers switch between the different models in Fig. 1 is captured by the generic summation model in Fig. 12a which permits summation between zero dB and 6 dB (as found in Experiments 1 and 3). An alternative model in which the gating is determined by changes in the variance of noise on the output lines is shown in Fig. 12b (see figure caption for details). This version of the model also illustrates a situation alluded to in the Introduction. If the variance of  $N_3$  were relatively high it is possible that the summation device would have been entirely hidden from the psychophysical probe used in the present experiments. Clearly, this was not generally the case.

In view of the above, we suggest that caution should be exercised on matters regarding the specific computations (and phenomenology) associated with the performance measures reported in this paper. This picture will only become clearer when experiments are performed in which both performance (discrimination thresholds) and phenomenology (perception of surfaces) are measured in association.

### Nonlinearities

In Fig. 1 we outlined three different models of summation. In general, however, our data did not fall neatly onto these predictions. For example, none of our data showed quite as much summation as the linear model (6 dB), though in several cases it fell between this prediction and that of the ideal model (3 dB). In the previous subsection we presented two generic summation models that could accommodate this. Here we point out that the inclusion of invertible nonlinearities can also achieve intermediate summation in some circumstances.

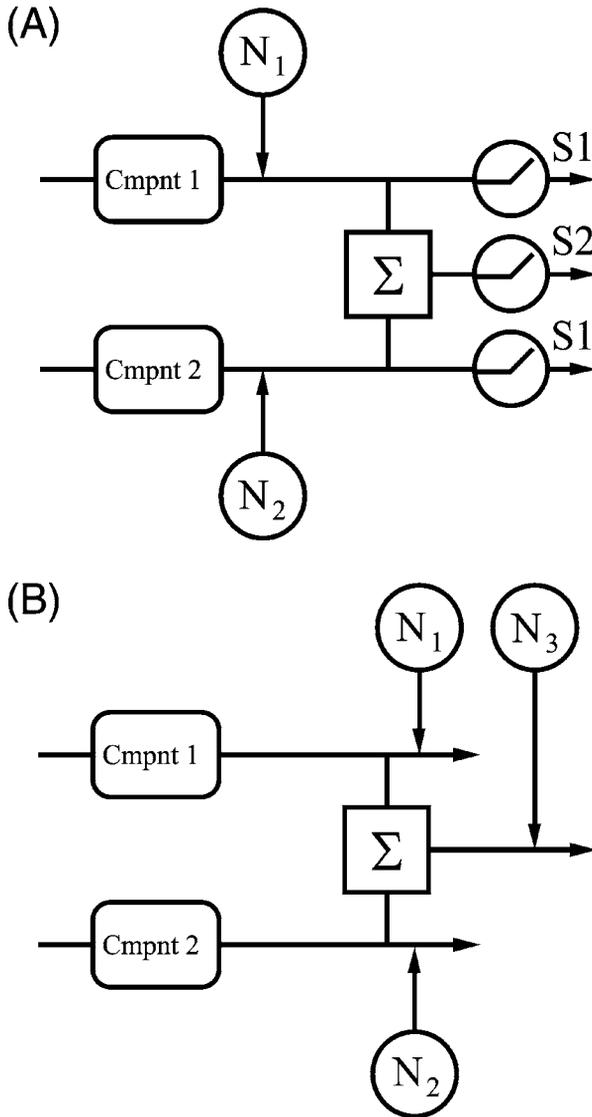
First consider the case where the null gradient is 0% and so the output of the putative depth mechanism is also zero. In the linear model (Fig. 1c;  $k = 1$  in equation 1), if the responses to the two signals pass through an expansive nonlinearity prior to summation (the  $Sens_i$  terms in equation (1) are first raised to a power greater than 1) then this would reduce the predicted level of summation below 6 dB. Alternatively, in the ideal model (Fig. 1b;  $k = 2$  in equation (1)), if the responses to the two signals pass through a compressive nonlinearity (the  $Sens_i$  terms in equation (1) are first raised to a power less than 1) then this would raise the predicted level of summation above 3 dB. In the first case, the effect of the nonlinearities is reduced as the magnitude of the null gradient increases. In the second case, the magnitude of summation reduces as the null gradient increases.

At first sight some aspects of these suggestions seem at odds with the prevailing view outlined in the Introduction, that summation of depth cues is linear. However, some care is needed here. That view is born out of experiments that have typically involved judgements of depth, and it is these measures that combine linearly. Our experiments have not addressed the relationship between the various stimulus gradients and their perception but they could be nonlinear (e.g. Fry *et al.*, 1949; Cutting and Millard, 1984; Freeman *et al.*, 1995). In that case, these nonlinearities could be inserted at various places in the physiological summation models in Fig. 1 so as to modify summation predictions for discrimination but maintain linear summation of depth cues once reexpressed in terms of depth estimates.

With all this uncertainty, it is clearly beyond the scope of the present paper to speculate on exactly why the summation data in Figs 7, 8 and 10 have the particular form that they do or at what processing stages lie the differences between the two observers. For now, we merely treat these data as an indication that summation is measurable between different pictorial depth cues within a putative depth mechanism. To address the questions above, a more complete model of performance measures for depth cue-combination is required. Experiments that describe the relation between stimulus cues and perceived depth would be particularly useful in characterising and constraining such a model.

### CONCLUSIONS

We performed summation experiments at increment threshold for pairs of pictorial depth cues. We found ideal summation when the signal pair was contrast-gradient



**Figure 12.** Generic summation models. (A) Constant variance noise and switched output lines. When only S1 is closed, the model is the independent model in Fig. 1a (mode 1), when both S1 and S2 are closed, the model is the ideal model in Fig. 1b (mode 2) and when only S2 is closed the model is the linear model in Fig. 1c (mode 3). Intermediate levels of summation are achieved by varying the frequency with which the switches are opened and closed. (B) Changeable variance noise and constant output lines. This is a modified version of the model in Fig. 1b in which a third source of noise is added after the summation device and the locations of the first two sources of noise are moved beyond the branches to the summation device. In this model, summation can be raised from 0 dB to 6 dB by either: (i) keeping the variance of  $N_1$  and  $N_2$  fixed, and progressively decreasing the variance of  $N_3$ , or (ii) keeping the variance of  $N_3$  fixed and progressively increasing the variance of  $N_1$  and  $N_2$ .

and size-gradient for a wide range of null gradients (0% to 30%). This suggests a depth cue summation model of the form shown in Fig. 1b for these two cues. When the contrast-gradient was replaced with a linear perspective cue (here called *x*-gradient), the level of summation depended upon the magnitude of the gradient in the null-interval and upon observer. This result requires a more flexible model than any of those shown in Fig. 1. We propose the two models in Fig. 12 as viable alternatives, though recognise that nonlinearities might also play an important role. At this stage it is not clear what controls the fluctuations in the summation data, but whether the stimuli are seen as compelling solid representations of surfaces might be one important factor.

In summary, we have demonstrated that summation of sensitivities can be found between pictorial depth cues and suggest that this might provide a suitable avenue for further investigating the nature of image representation at and beyond early cortical processes.

### Acknowledgements

This work was supported by a pilot-study grant awarded by the Leverhulme Trust. We would like to thank Mark Bradshaw for providing us with the data replotted in the right part of Fig. 9.

### NOTES

1. Doshier *et al.* (1986) made perceptual judgements about the direction of rotation in the Kinetic Depth Effect and used this to estimate the contribution of stereo and luminance contrast cues. See Landy *et al.* (1995) for some discussion of this work.

2. See Methods for calculations in dB.

3. The ‘solid’ and ‘flat’ modes identified here are similar to the ‘proximal’ and ‘distal’ modes discussed by Mack (1978). The ‘pictorial’ mode identified here is similar to ‘pictorial’ perception discussed by Hagen (1978).

### REFERENCES

- Bennett, P. J. and Cortese, F. (1996). Masking of spatial frequency in visual memory depends on distal, not retinal, frequency, *Vision Research* **36**, 233–238.
- Bonneh, Y. and Sagi, D. (1999). Contrast integration across space, *Vision Research* **39**, 2597–2602.
- Bradshaw, M. F. and Rogers, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth, *Vision Research* **36**, 3457–3468.
- Bruno, N. and Cutting, J. E. (1988). Minimodularity and the Perception of Layout, *J. Exp. Psychol.: General* **117**, 161–170.
- Buckley, D. and Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges, *Vision Research* **33**, 919–933.
- Bülthoff, H. H. and Mallot, H. A. (1988). Integration of depth modules: stereo and shading, *J. Opt. Soc. Amer.* **5**, 1749–1758.

- Burbeck, C. A. (1987). Locus of spatial frequency discrimination, *J. Opt. Soc. Amer. A* **4**, 1807–1813.
- Cumming, B. G., Johnston, E. B. and Parker, A. J. (1993). Effects of different cues on curved surfaces viewed stereoscopically, *Vision Research* **33**, 872–838.
- Cutting, J. E. and Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces, *J. Exp. Psychol.: General* **113**, 198–216.
- Cutting, J. E., Bruno, N., Brady, N. P. and Moore, C. (1992). Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth, *J. Exp. Psychol.: General* **121**, 364–381.
- Davis, G. and Driver, J. (1998). Kanizsa subjective figures can act as occluding surfaces at parallel stages of visual search, *J. Exp. Psychol.: Human Perception and Performance* **24**, 169–184.
- Dosher, B. A., Sperling, G. and Wurtz, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure, *Vision Research* **26**, 973–990.
- Epstein, W. (1981). The relationship between texture gradient and perceived slant-in-depth: direct or mediated? *Perception* **10**, 695–702.
- Freeman, T. C. A., Harris, M. G. and Meese, T. S. (1995). On the relationship between deformation and perceived surface slant, *Vision Research* **36**, 317–322.
- Fry, G. A., Bridgman, C. S. and Ellerbrock, V. J. (1949). The effects of atmospheric scattering on binocular depth perception, *Amer. J. Optometry* **26**, 9–15.
- Gibson, J. J. (1950). *The Perception of the Visual World*. Houghton Mifflin: Boston, USA.
- Goldstein (1989). *Sensation and Perception*, 3rd edn. Brooks/Cole, California, USA.
- Graham, N. (1989). *Visual Pattern Analyzers*. Oxford University Press, Oxford, UK.
- Greenlee, M. W. (1992). Spatial frequency discrimination of band-limited periodic targets — Effects of stimulus contrast, bandwidth and retinal eccentricity, *Vision Research* **32**, 275–283.
- Hagen, M. A. (1978). An outline of an investigation into the special character of pictures, in: *Modes of Perceiving and Processing Information*, Pick, H. I. and Saltzman, E. (Eds), pp. 23–38. LEA, Hillsdale, New Jersey.
- He, Z. J. and Nakayama, K. (1994). Perceiving textures: Beyond filtering, *Vision Research* **34**, 151–162.
- Horn, B. K. P. and Brookes, M. J. (Eds) (1989). *Shape From Shading*. MIT Press, Cambridge MA.
- Howard I. P. and Rogers, B. J. (1995). *Binocular Vision and Stereopsis*. Oxford University Press, Oxford.
- Johnston, E. B., Cumming, B. G. and Parker, A. J. (1993). Integration of depth modules: Stereopsis and texture, *Vision Research* **33**, 813–826.
- Johnston, E. B., Cumming, B. G. and Landy, M. S. (1994). Integration of stereopsis and motion shape cues, *Vision Research* **34**, 2259–2275.
- Landy, M. S., Maloney, L. T., Johnston, E. B. and Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion, *Vision Research* **35**, 389–412.
- Legge, G. E. and Foley, J. M. (1981). Contrast masking in human vision, *J. Opt. Soc. Amer.* **70**, 1458–1471.
- Li, A. and Zaidi, Q. (2000). Perception of three-dimensional shape from texture is based on patterns of oriented energy, *Vision Research* **40**, 217–242.
- McKee, S. P. and Smallman, H. S. (1998). Size and speed constancy, in: *Perceptual Constancy: Why Things Look as They Do*. Walsh, V. and Kulikowski, J. (Eds), pp. 373–408. Cambridge University Press, Cambridge, UK.
- McKee, S. P. and Welch, L. (1992). The precision of size constancy, *Vision Research*, **32**, 1447–1460.
- Mack, A. (1978). Three modes of visual perception, in: *Modes of Perceiving and Processing Information*, Pick, H. I. and Saltzman, E. (Eds), pp. 171–186. LEA, Hillsdale, New Jersey.
- Mamassian, P. and Landy, M. (1998). Observer biases in the 3D interpretation of line drawings, *Vision Research* **38**, 2817–2832.
- Mather, G. (1996). Image blur as a pictorial depth cue, *Proc. Roy. Soc. London B.* **263**, 169–172.

- Meese, T. S. and Williams, C. B. (2000). Probability summation for multiple patches of luminance modulation, *Vision Research* **40**, 2101–2113.
- Meese, T. S. and Harris, M. G. H. (2001). Independent detectors for expansion and rotation, and for orthogonal components of deformation, *Perception* **30**, 1189–1202.
- Meese, T. S., Hess, R. F. and Williams, C. B. (2001). Spatial coherence does not affect contrast discrimination for multiple Gabor stimuli, *Perception* **30**, 1411–1422.
- Nakayama, K., Shimojo, S. and Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping and the recognition of occluded objects, *Perception* **18**, 55–68.
- Norman, J. F., Todd, J. T. Perotti, V. J. and Tittle, J. S. (1996). The visual perception of three-dimensional length, *J. Exp. Psychol.: Human Perception and Performance* **22**, 173–186.
- O’Shea, R. P., Blackburn, S. G. and Ono, H. (1994). Contrast as a depth cue, *Vision Research* **34**, 1595–1604.
- O’Shea, R. P., Govan, D. G. and Sekuler, R. (1997). Blur and contrast as pictorial depth cues, *Perception* **26**, 599–612.
- Ono, H. (1967). Difference thresholds for stimulus length under simultaneous and nonsimultaneous viewing conditions, *Perception and Psychophysics* **2**, 213–316.
- Regan, D. and Beverley, K. I. (1985). Postadaptation orientation discrimination, *J. Opt. Soc. Amer. A* **2**, 147–155.
- Robson, J. G. and Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field, *Vision Research* **21**, 409–418.
- Rogers, B. J. and Graham, M. E. (1979). Motion parallax as an independent cue for depth perception, *Perception* **8**, 125–134.
- Sedgwick, H. A. and Nicholls, A. L. (1993). Cross-talk between the picture surface and the pictured scene: Effects on perceived shape, *Perception* **22** (suppl.), 109.
- Sperling, G., Doshier, B. A. and Landy, M. S. (1990). How to study the kinetic depth effect experimentally, *J. Exp. Psychol.: Human Perception and Performance* **16**, 445–450.
- Stevens, K. A. (1981). The information content of texture gradients, *Biological Cybernetics* **42**, 95–105.
- Stevens, K. A. and Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces, *Vision Research* **28**, 371–386.
- Todd, J. T. and Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture, *J. Exp. Psychol.: Human Perception and Performance* **13**, 242–255.
- Todd, J. T. and Oomes, A. H. J. (2002). Generic and non-generic conditions for the perception of surface shape from texture, *Vision Research* **42**, 837–850.
- Troscianko, T., Montagnon, R., Le Clerc, J., Malbert, E. and Chanteau, P.-L. (1991). The role of colour as a monocular depth cue, *Vision Research* **31**, 1923–1930.
- Tyler, C.W. and Chen, C. C. (2001). Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation, *Vision Research* **40**, 3121–3144.
- Watson, A. B. and Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking, *J. Opt. Soc. Amer. A* **14**, 2379–2391.
- Wetherill, G. R. and Levitt, H. (1965). Sequential estimation of points on a psychometric function, *Brit. J. Math. Stat. Psychol.* **18**, 1–10.
- Wilson, H. R. and Gelb, D. J. (1984). Modified line-element theory for spatial-frequency and width discrimination, *J. Opt. Soc. Amer. A* **1**, 124–131.
- Wilson, H. R., Levi, D., Maffei, L., Rovamo, J. and DeValois, R. (1990). The Perception of form: Retina to striate cortex, in: *Visual Perception: The Neurophysiological Foundations*, Spillman, L. and Werner, J. S. (Eds), pp. 231–272. Academic Press, London.
- Young, M. J., Landy, M. S. and Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues, *Vision Research* **33**, 2685–2696.
- Zaidi, Q. and Li, A. (2002). Limitations on shape information provided by texture cues, *Vision Research*, **42**, 815–835.