Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Vision Research 43 (2003) 1575–1587
Vision Research
www.elsevier.com/locate/visres

How many positions can we perceptually encode, one or many?

R.F. Hess [a,*], G. Barnes [b], S.O. Dumoulin [a], S.C. Dakin [c]

[a] McGill Vision Research, Department of Ophthalmology, McGill University, 687 Pine Avenue West (H4-14), Montreal, Que., Canada H3A 1A1
[b] Neurophysiology Unit, University of Aston, UK
[c] The Institute of Ophthalmology, London, UK

Received 2 April 2002; received in revised form 19 September 2002

Abstract

Here we show that our sensitivity for discriminating relative position across the visual field is limited. In experiment 1 we show that we are much worse at detecting a texture defined by the relative position of elements within an array than would be expected if we had access to multiple estimates of relative position across the visual field. In experiment 2 we show that human performance is impaired for positional judgments when there is uncertainty as to which of a number of possible elements is misaligned. This impairment is greater than one would expect from an ideal observer model and greater than that found for a comparable task involving orientation. It is consistent with positional thresholds being determined by only one estimate of relative position. In experiment 3 we estimate the number of suprathreshold positional signals that can be pooled at the same time across the visual field using a standard summation variance paradigm. The results suggest that the human visual system is limited to one estimate of position, but additional estimates can be built up serially over time; however, this process is slow and probably cognitive in nature. These experiments taken as a whole suggest that only one estimate of relative position (i.e. relative to a predefined reference) at a time is accessible at the perceptual level.
© 2003 Elsevier Science Ltd. All rights reserved.

Keywords: Positional accuracy; Positional coding; Relative position

1. Introduction

Vision is the most highly developed of the human senses, as much as 50% of the primate cortex is devoted to vision related tasks (Van Essen, Anderson, & Fellemen, 1992). Our visual sensitivity is impressive along a number of dimensions not least of which is positional accuracy. Positionally, we are accurate to a fraction of the size of an individual photoreceptor (i.e. less than 30″). There is evidence that the visual system extracts the centroid of the retinal light distribution to achieve such accuracy (Watt & Morgan, 1983) and it is assumed that this is done in parallel across the central visual field at an early stage of visual processing (Marr, 1982; Watt, 1988). Indeed, the idea that there is a local feature representation built up from the location of the edges of image components has formed the foundation on which some computational models of vision are based (Marr, 1982; Watt, 1988).

Human positional sensitivity has been measured using a number of different techniques. It is accepted that the relative position of abutting stimuli, such as vernier targets, is due in part to local contrast and orientation information rather than position per se (Carney & Klein, 1999). Targets that are well separated allow a better estimate of how accurately the visual system can discriminate relative position when the individual stimuli stimulate different neural populations (Toet & Koenderink, 1988). In this case, accuracy varies with the size of the individual elements whose positions are to be discriminated in a way that suggests that the visual system is computing something akin to the centroid of the retinal light distribution (Hess & Holliday, 1996; Watt & Morgan, 1983). The computation of relative position is assumed to occur in different parts of the field in parallel. However, it has never been clear how relative position is encoded within a neural population where the

* Corresponding author. Tel.: +1-514-843-1690; fax: +1-514-843-1691.
E-mail address: robert.hess@mcgill.ca (R.F. Hess).

0042-6989/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0042-6989(03)00122-6

two-dimensional spatial position of each neuron is unknown. Here we show, using three different experimental approaches, that at the level of conscious perception the human visual system's ability to discriminate position in different parts of the visual field at the same time, is severely limited. The results are consistent with only one estimate of relative position (i.e. relative to a known reference) being accessible at any one time at the perceptual level.

2. Experiment 1—Orientation discrimination of textures defined solely by relative position

2.1. Introduction

If the visual system can compute the relative position of image features in different parts of the central field at the same time and if these estimates are available to later stages of perception then it should be able to effortlessly detect simple textures that have been constructed purely from the relative positions of spatially distributed array elements. We created either a vertical or a horizontal 1-D Gaussian-profile texture bar, defined solely by a change of relative 2-D position of constituent bandpass array elements whose local contrast and orientation was randomized. Thus the 1-D Gaussian function controlled which elements were subjected to a 2-D positional displacement and the extent of this displacement. The 2-D displacement itself was Gaussian distributed with a mean equal to the original unjittered array spacing. Fig. 1 shows the perturbed grid positions (Fig. 1a) that define a vertically oriented texture bar. Fig. 1b shows an example of the vertically oriented texture bar composed of Gabor elements of random contrast and local orientation occupying these perturbed and unperturbed grid positions. We asked the question, "how sensitive is the human visual system for detecting such position-defined textures?" To answer this question we measured the

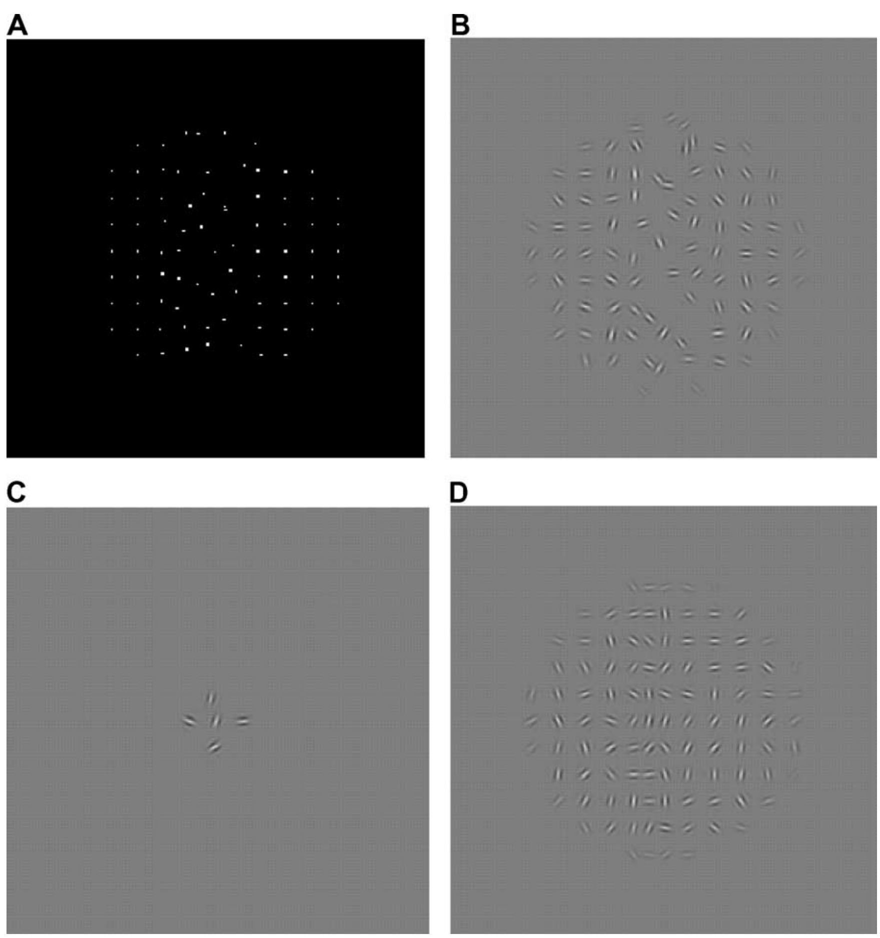

Fig. 1. Illustration of the positionally defined texture used in experiment 1. The grid positions (A) of an array of Gabor element (whose contrast and local orientation are random) are perturbed according to a 1-D Gaussian function to produce either a vertical or horizontal texture bar (a vertically oriented texture bar is illustrated in B). The same texture bar has been masked down to only its central five elements (C). In D, a similar texture bar is defined by a pure density change of regularly spaced elements.

minimum positional disturbance necessary to do the horizontal/vertical discrimination of a positional-texture defined bar.

2.2. Methods

2.2.1. Apparatus

An Apple Macintosh computer controlled stimulus presentation and recorded subjects' responses. Programs for running the experiment were written in the Matlab programming environment (Mathworks Ltd.) using Psychtoolbox code (Brainard, 1997; Pelli, 1997). Stimuli were displayed on a 21" Nanao FlexScan monochrome monitor, with a frame refresh rate of 75 Hz. Pseudo 12-bit contrast accuracy was achieved by electronically combining the RGB outputs from the computer using a video attenuator (Pelli & Zhang, 1991).

2.2.2. Multi-element disarray modulation and task

Gabor elements (comprising a 1-D sinusoid multiplied by a 2-D Gaussian envelope) having a peak spatial frequency of 5 c/deg and an envelope sigma size of 0.06° were generated from a 256×256 pixel array (the grid subtended $4.2° \times 4.2°$, it has 16×16 array elements and an inter-element separation of 16 min). A 1-D Gaussian-profile controlled the magnitude of the 2-D spacing of the Gabors within the array, its sigma was set to 0.53° and its peak position, with respect to the center of the array, was jittered from trial to trial (Fig. 1a shows how the grid positions were modulated by this 1-D Gaussian function). The Gaussian function whose magnitude controlled how the positions of the array elements deviated from regularity was itself either horizontal or vertical in orientation and its peak height was the experimental variable (see Fig. 1b in which a vertically oriented positional-texture bar is illustrated). The whole stimulus was contained in a circular window with a raised cosine profile and presented for a duration of 500 ms. The contrast and the orientation of each Gabor were randomized across the array for each presentation (uniformly distributed between 20–40% for contrast and 0°–180° for orientation) so that local orientation or contrast could not be used to help discriminate the orientation of the 2-D positional disturbance.

Thresholds were derived by fitting a Weibull function (Weibull, 1951) to frequency of seeing data obtained from a 2 AFC task using the method of constant stimuli. Subjects were asked, "is the orientation of the bar produced by the element disarray, horizontal or vertical?" Three threshold estimates were averaged; each of these was obtained from individual runs of 20 trials per 11 stimulus levels. In one experiment, the density of the elements constituting the bar that was itself defined by element disarray was varied to directly compensate for any density cue that occurred secondary to the disarray (see below).

2.2.3. Masked multi-element stimulus and task

The multi-element stimulus was masked down so that only the central five elements were visible (see Fig. 1c). The 2-D positional disturbance was exactly the same as that described above (Fig. 1b). The subjects' task was the same as that described above, namely to decide whether the positional disturbance was oriented horizontally or vertically. Thresholds were derived in the same way to that described above.

2.2.4. Multi-element density modulation and task

Same as that described above for the multi-element disarray stimulus and task except that now the 1-D Gaussian function controlled the density (i.e. the inter-element distance) and not the irregularity of the element array (Fig. 1d shows a vertical texture bar defined solely by density). Subjects were asked, "is the orientation of the bar produced by element density, horizontal or vertical?"

2.3. Results and discussion

The results displayed in Fig. 2 show performance for discriminating the orientation of the positional disarray. Subjects exhibited thresholds of around 6′ (unfilled circles). This corresponds to the peak 2-D disarray for the elements corresponding to the middle of the Gaussian texture-defined bar.

The first issue is whether performance in this case is based on disarray or perceived density. Previous work has implicated the latter in tasks where irregularity is introduced in element spacing (Allik & Tuulmets, 1991). To address this issue we first measured the threshold for a change in density using a comparable stimulus arrangement and task to those already described. Subjects now had to discriminate between a horizontal or vertical Gaussian-profile texture bar defined by the density of the elements (e.g., Fig. 1d). Thresholds for RFH were around 0.2 (the spacing of the elements corresponding to the peak of the 1-D Gaussian disturbance was reduced by 20% compared with those outside the Gaussian disturbance) whereas for SOD thresholds were around 0.15. We reasoned that if the threshold disarray that we had previously measured was due to a change in perceived density, then by compensating for any density change we would expect to see an elevation in thresholds for the disarray stimulus. No such elevation was observed (crosses in Fig. 2), suggesting that the task is not done on the basis of perceived density.

To ascertain the extent to which the original multiple positional estimates were aiding performance we compared the results to the same stimulus when only the

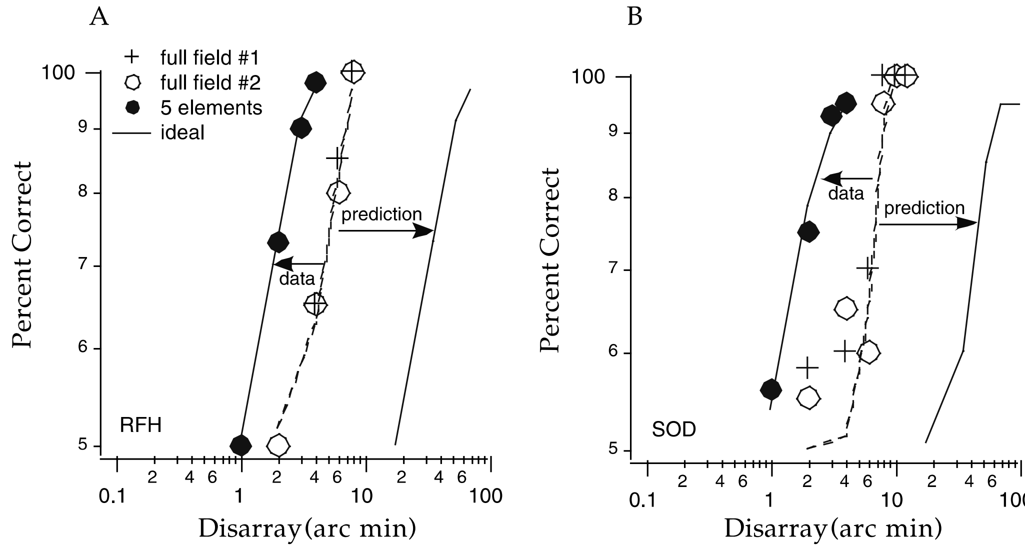

Fig. 2. Psychometric data for two subjects comparing performance on detecting the orientation of the positionally defined texture bar using all the elements in a multi-element display (see Fig. 1B for illustration; open circles in A & B) and just the central five elements (see Fig. 1C for illustration; filled symbols in A & B). The crosses represent data for the multi-element display in which any perceived density has been compensated for. The solid curve represents the predicted performance of the ideal observer for the five-element display given the multi-element thresholds. Sensitivity is predicted by the ideal observer to be worse (factor of 9) on the five-element display compared to the multi-element display however the experimental results show the opposite (factor of 2–3 better).

central five elements were present (e.g., Fig. 1c). The results for two subjects are shown by filled circles in Fig. 2 where they are compared with the previously discussed results obtained using the multiple element array condition (unfilled circles). Surprisingly, performance in the masked condition (filled symbols—Fig. 2) that only contains a fraction of the positional information available in the original stimulus (unfilled circles—Fig. 2) is as much as a factor of 2–3 *better*. To obtain predictions based on the informational difference between these two tasks we developed an ideal observer prediction (see Appendix A for more details). The ideal observer encodes all element positions with equal accuracy and uses a relative position metric. In the case of the multiple element stimulus, since there are many more positional samples, the ideal observer predicts better performance (nine times) for the multiple element case compared with the five-element case. In Fig. 2 (solid curve) we display the ideal observer's predictions for the five-element case based on knowing the threshold for the multiple element case.

Experimentally we find that the opposite is true, performance in the five-element case is about a factor of 5 *better* than that found for the multiple element display. Not only are these extra positional samples in the multiple element case of no help, they actually reduce performance. The inescapable conclusion is that the visual system, unlike the ideal observer, does not compute the position of multiple elements with equal accuracy. It seems that the visual system in unable to utilize any more than a few positional estimates at any one time. But how many is a few?

3. Experiment 2—Positional accuracy with stimulus uncertainty

3.1. Introduction

To test how limited the visual system's capability is for judgments of relative position, we measured positional accuracy for a stimulus comprising three well separated elements (Fig. 3a) where subjects were uncertain as to which one of the three elements was misaligned. In one condition subjects knew which was the signal element (element certainty condition) whereas in the other condition subjects were uncertain which one of the three elements was misaligned (element uncertainty condition). We reasoned that in the element certainty task, only one relative position needs to be encoded (i.e. relative to the known reference). In the element uncertainty task involving *n* elements, for sensitivity to be maintained, *n* relative positional estimates are required. Performance should be much worse in the stimulus uncertainty case if the visual system is limited to only one estimate of relative position but unchanged if the visual system can derive at least two estimates of relative position at the same time.

3.2. Methods

3.2.1. Three/four-element stimuli and alignment task

Three (Fig. 3) and in a later experiment, four (Fig. 4) Gabor elements (spatial frequency 2 c/deg; sigma 0.13°; separation 3.3°; contrast 80%) were arranged in a vertical line (see Fig. 3a). The absolute position of the elements on the screen was randomized from trial to trial (±0.83°)

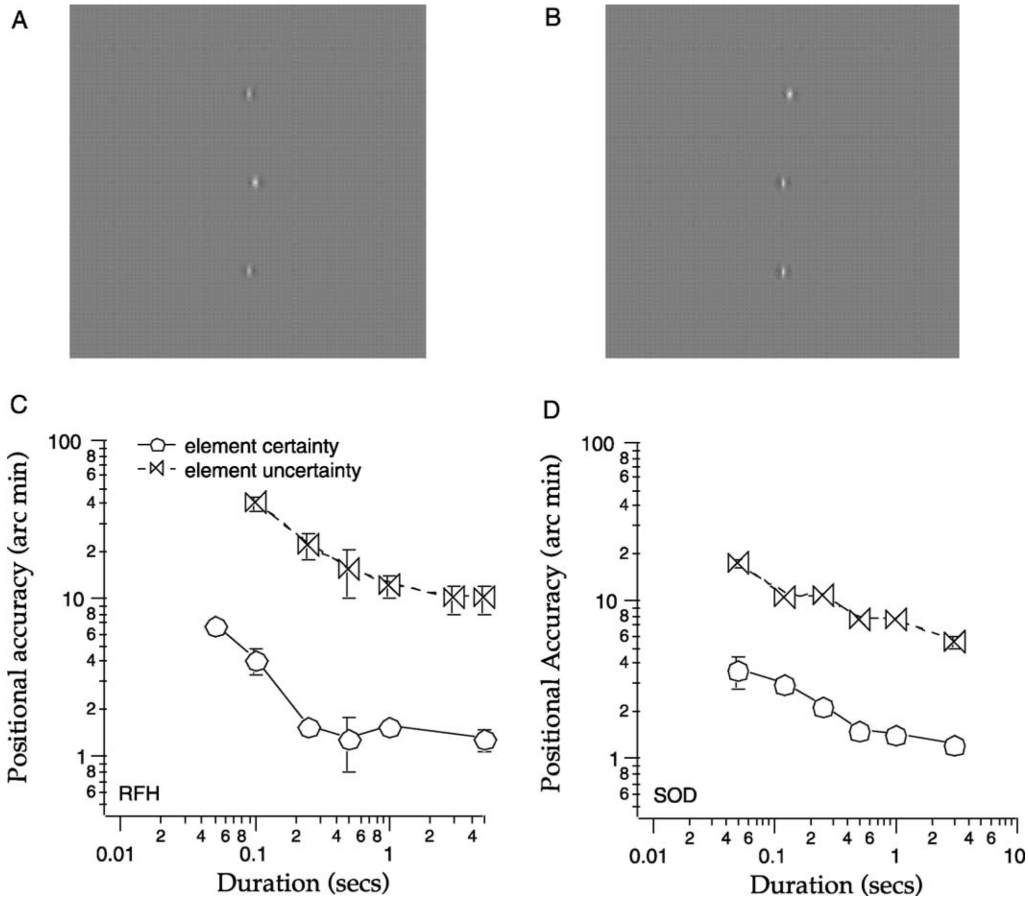

Fig. 3. Positional thresholds are compared for two subjects for the case where the identity of the misaligned element is known (element certainty case—circles) and where it is not known (element uncertainty case—bowties) as a function of exposure duration. Performance is between 6–10 times better in the former case, irrespective of exposure duration, this is not expected from an ideal observer model (see text).

so that the edges of the screen could not be used to solve the task. We manipulated the level of uncertainty in the following way: the subject was informed prior to a block of trials which elements provided the reference and which were possible signal elements (subjects knew that only one element was ever displaced). In the element certainty case where there was only one signal element, the subject knew which of the elements that was. In the extreme version of the element uncertainty case where any one of the four elements could in principle be the signal element, the subject was uncertain, on a trial by trial basis, which element contained the signal. A one interval, 2AFC procedure with the method of constant stimuli was used where the subject had to indicate the left/right misalignment of the displaced element. Thresholds were derived by fitting a Weibull function to frequency of seeing data (subjects were asked "is the misaligned element displaced to the left or right of the other two reference elements?"). Three threshold estimates were averaged, each of these was obtained from individual runs of 20 trials per stimulus level (11 levels). Thresholds were measured at a number of fixed stimulus duration from 50 ms to 3 s followed by a spatial mask (80% contrast 1-D spatial noise) to ensure that processing was limited to the stimulus duration.

3.2.2. Three-element stimuli and orientation task

Three Gabor elements (spatial frequency 2 c/deg; sigma 0.13°; separation 3.3°; contrast 80%) were arranged in a vertical line (see Fig. 5a). The absolute orientation of the two reference elements was randomized from trial to trial (±45°) so that the orientation judgement to be made was a relative one. In the element certainty case, the target element whose orientation was to be judged (relative to that of the other two reference elements) was known, whereas in the element uncertainty condition, any one of the three elements could be the target. Positional thresholds were derived by fitting a Weibull function to the frequency of seeing data obtained from a 2 AFC task using the method of constant stimuli (subjects were asked, "is the element with the different orientation rotated clockwise or counterclockwise from the other two reference elements?"). Three threshold estimates were averaged, each of these

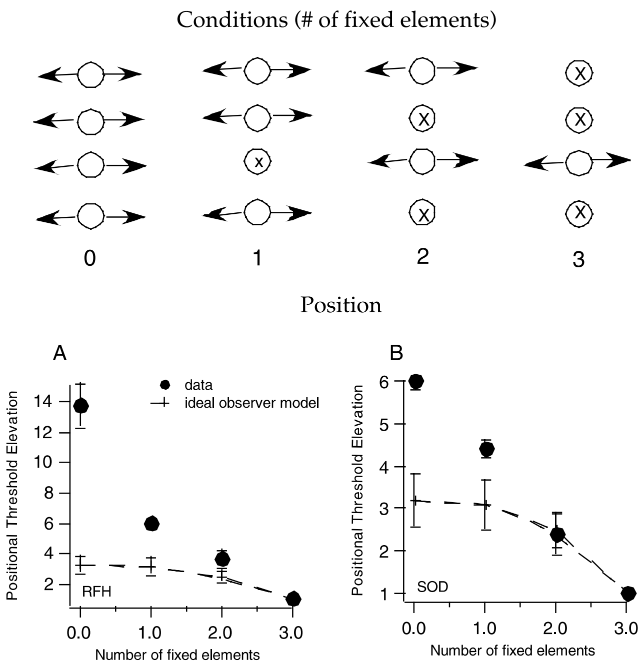

Fig. 4. Positional thresholds are compared for two subjects as a function of the uncertainty as to the identity of the misaligned element. Conditions 0 to 3 represent the number of elements that are fixed, i.e. are reference elements. Positional thresholds (symbols) are plotted for each of these conditions for a four-element display. Predictions from an ideal observer model (dashed lines) in which the positions of all elements are encoded with the same accuracy, are given for comparison. Human performance is not well predicted by the ideal observer.

was obtained from individual runs of 20 trials per stimulus level (11 levels). The stimulus duration was varied from 50 ms to 3 s followed by a spatial mask to ensure that processing was limited to the stimulus duration.

3.3. Results and discussion

The results for the two versions of the three-element alignment task (element certainty and element uncertainty) as a function of stimulus duration are shown in Fig. 3. In the element certainty case (circles), sensitivity is good to about 1 min (or 0.1 × the sigma of the 2-D Gaussian profile) and additional experiments showed that it does not depend on which element is the target (data not displayed). Performance varies with exposure duration as expected. The dynamics are slow, reaching an asymptote at around 500 ms (Waugh, 1998). In the element uncertainty case (bowties), performance is between a factor of 6–10 worse irrespective of stimulus duration.

If the visual system could encode two relative positions at the same time one would expect performance to have been unchanged in these two conditions. Furthermore, if the visual system could build up the number of estimates it makes of position over time then threshold performance for these two tasks should come together at longer stimulus durations which was not the case. This suggests that the visual system's encoding of position is limited. These results are consistent with the visual system being able to encode only one estimate of position at a time. Such an estimate would need to be relative to a known positional reference. However, if there is uncertainty about which element is the reference and which has been displaced, positional sensitivity is reduced. Furthermore, it would seem that more estimates cannot be accumulated rapidly over time, at least up to the 3–5 s limit investigated here. The mean ratio in performance between the certainty and uncertainty conditions for a group of eight subjects, six of whom were naive to the objectives of the experiment, was 8.5 ± 2.1 for a stimulus duration of 3 s.

An alternate explanation is that performance on this task, even in the certainty case, is limited by an intrinsic uncertainty of the absolute vertical. The further reduction in performance in the uncertainty case might then be the result of a further disruption to the frame of reference. Imagine the case where we randomize the absolute orientation defined by the two reference elements; this would not be expected to affect performance in the certainty condition but would render the uncertainty case impossible to do. Such an explanation would predict two things. First, in the element certainty case, removal of one of the two reference elements should make performance much worse (i.e. by further degrading the absolute vertical reference). Second, in the element uncertainty case, the addition of another reference element should make performance better (i.e. by better defining the absolute vertical reference). Performance was not found to be reduced in the element certainty case when we removed one of the two reference elements (e.g., two-element thresholds were $1.13'$ and $0.8'$ compared with three-element thresholds of $1.15'$ and $0.75'$ for RFH and SOD respectively). The addition of an extra element rather than improving performance in the uncertainty case as would be predicted from a better defined plane of reference, made performance worse. These results are shown in Fig. 4 where we systematically varied the number of possible elements that could be displaced for a four-element stimulus under unlimited viewing conditions. This experimental manipulation not only addresses the issue of the plane of reference (by having four elements) but also explores intermediate levels of uncertainty between the two extreme cases shown in Fig. 3.

The stimulus conditions are diagrammatically illustrated at the top of the figure. There are four Gabor elements, identical to those previously described in Fig. 3 and we vary the number of elements that are fixed and thus provide a reference. When this is 0, it corresponds

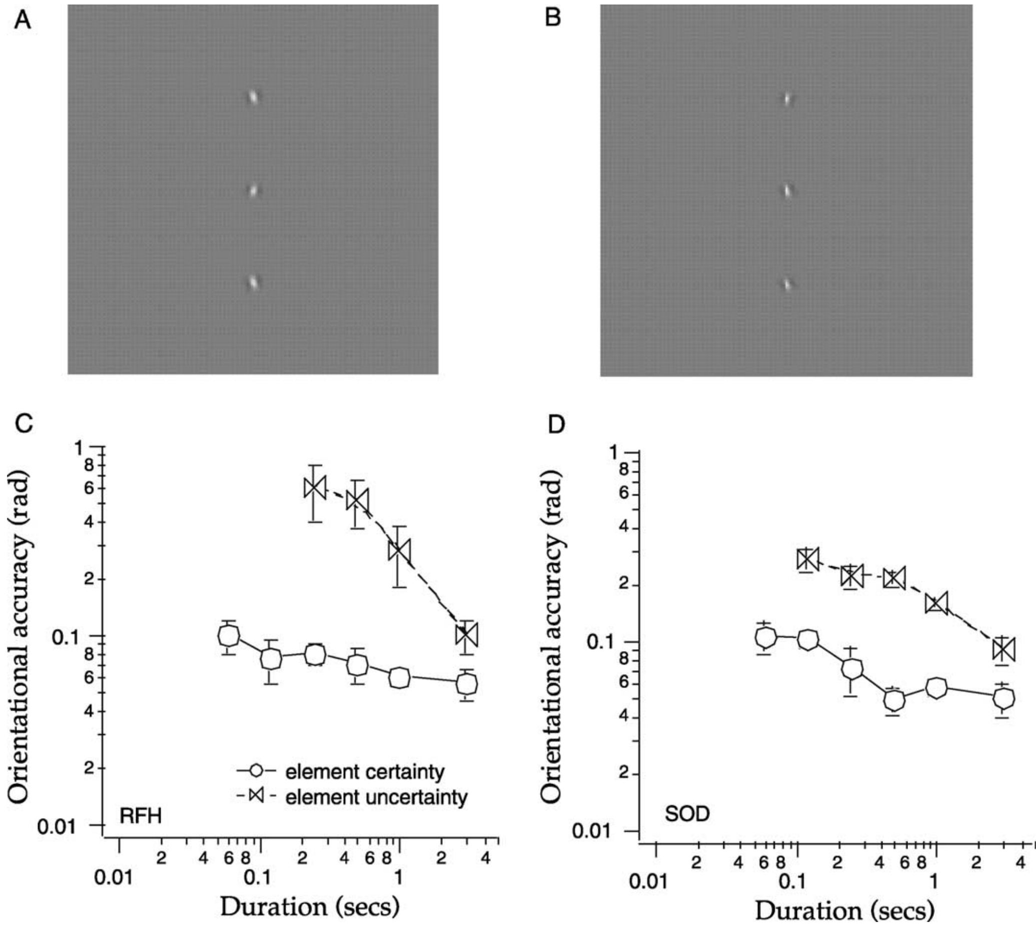

Fig. 5. Comparable conditions to those of Fig. 3 except that now relative orientation rather than relative position is being measured. Given enough time, performance is only a factor of 2 worse in the element uncertainty case (bowlies) compared with the element certainty case (circles), in line with predictions from our ideal observer model (see text).

to the element uncertainty case where anyone of the four elements can be displaced and the other three elements define the vertical reference plane. When this is 3, it corresponds to the element certainty case where the identity of the displaced element is known. Notice that when the number of fixed elements is 2, the subject has a central reference element to help anchor the vertical reference plane. The results are quite clear in that performance gets progressively worse as the level of uncertainty about the identity of the displaced element increases. The solid line represents the predictions of our ideal observer model in which all positions are encoded with equal accuracy and performance is determined within a relative position metric (see appendix). The model exhibits an initial loss of sensitivity going from 3 to 2 fixed elements but much less loss from 2 to 0 fixed elements. This initial loss of sensitivity may be due to the extra comparison stage required to decide which element was displaced when its identity is uncertain. This model does not account for the experimental results that show that sensitivity undergoes a progressive decline as uncertainty increases. We conclude that the reason for this extra loss of sensitivity is not related to an impoverished reference frame or the extra comparison required, but rather to the fact that the visual system does not independently encode more than one relative position.

That the threshold for the element uncertainty case (Figs. 3 and 4) is only 6–10 times higher than that in the element certainty case maybe partly because once an element is sufficiently displaced from the other two reference elements, the overall shape defined by the three elements is altered. This global shape change of the three-element stimulus is sufficiently elementary for the direction of the displaced element to be deduced secondarily using higher level cognitive processes.

It could be argued that there is another important difference between the element certainty and uncertainty cases discussed above, one that could account for the observed difference in performance but not lead to the conclusion that the visual system has access to only one estimate of relative position at a time. The element uncertainty case differs from the element certainty case in two ways. First, a knowledge of which element is to be displaced may result in a reduction in its positional noise, enough to account for a threshold elevation of 6–10-fold. Second, in the element uncertainty case, extra processing (for position) is required by the visual system to solve the task. It is possible that there is a substantial cost (i.e. increased noise, cognitive load, etc.) associated with this extra processing, adding enough additional noise to elevate threshold 6–10-fold. Such explanations would not be specific to relative position but would be equally applicable to any comparable visual task that requires multiple comparisons.

To test these alternate hypotheses, we measured performance using an identical certainty/uncertainty paradigm involving the same stimulus arrangement to that used in the relative position task but this time testing relative orientation (absolute orientation of the reference elements was randomized over the range ±45°). The task was to detect whether the middle element (i.e. the element certainty condition) was rotated clockwise or counterclockwise compared with the local orientation of the two outer reference elements. In the element uncertainty condition, one of the elements was rotated clockwise or counterclockwise compared with the local orientation of the two other elements. In this respect it was identical in principle (requiring the same number of comparisons and attentional/cognitive load) to that of the previously described three-element positional task. We measured performance, as we had done for the relative position task, as a function of the duration of stimulus presentation for two subjects. These results are displayed in Fig. 5c and d. Judgements of relative orientation, unlike those previously discussed for relative position, result in approximately a 2-fold reduction for the uncertainty condition at the longest exposure duration tested. The mean ratio in performance for a group of eight subjects, six of whom were naive to the objectives of the experiment was 2.6 ± 0.6 for a stimulus duration of 3 s. Our ideal observer supplied predictions for the general case where knowledge of the target element (i.e. whose parameter is to be altered) results in a reduction in the noise associated with its location. Furthermore, in the element uncertainty case, the visual system has to make more comparisons (an additional two). The model which is described in the appendix predicts a loss of a factor of 2 in the sensitivity for the uncertainty case, irrespective of the degree of noise reduction associated with a knowledge of the target element. This is much less than we had found for the positional task (i.e. 6–10-fold) but just what we had found in the comparable orientation task.

Finally, we compared performance for the same gradation of uncertainty (from 0 to 3 fixed elements) as we did for position in Fig. 4. These results are shown in Fig. 6 by the filled symbols. Performance in the case of orientation, unlike that for position (Fig. 4), does not progressively deteriorate as uncertainty increases. Performance is a factor of 2 worse in the three compared with the two fixed elements but does not deteriorate much from 2 to 0 fixed elements. The dashed curve represents the results of our ideal observer. This model, unlike that for the position case, proves to be a good predictor of orientation discrimination. This result for orientation is not surprising since we know that the visual system can encode orientation in parallel across the field (Dakin, 2001).

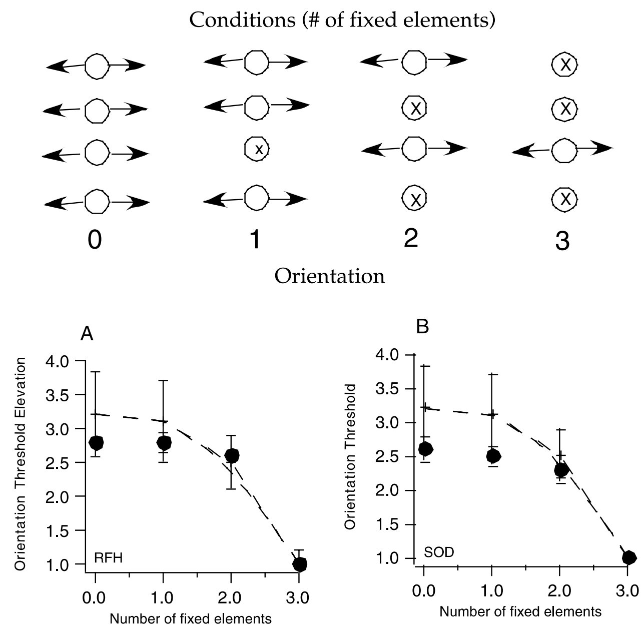

Fig. 6. Orientation thresholds are compared for two subjects as a function of the uncertainty as to the identity of the misoriented element. Conditions 0 to 3 represent the number of elements that are fixed, i.e. are reference elements. Orientation thresholds (symbols) are plotted for each of these conditions for a four-element display. Predictions from an ideal observer model (dashed lines) in which the orientations of all elements are encoded with the same accuracy, are given for comparison. Human performance in this orientation case, unlike its positional counterpart (Fig. 4) is well predicted by the ideal observer.

The fact that the main findings for the position task could not be replicated for a comparable orientation task suggests that any common features in the design of the positional task are not responsible for its poor performance in the element uncertainty case. For example, the fact that attentional/cognitive demands are different and that more comparisons are required cannot lie at the heart of an explanation for the poor positional sensitivity in the uncertainty case because this is equally true in the orientation task where the results are better and in line with our ideal observer predictions that assume a parallel encoding scheme. A parsimonious explanation is that the visual system is selectively deficient at making multiple position estimates, having access to possibly only one estimate of relative position at a time.

4. Experiment 3—Pooling of relative position

4.1. Introduction

Given that our position thresholds appear to be based on only one relative position in any one part of the visual field (experiments 1 and 2), can a number of such suprathreshold estimates be integrated across the visual field? To estimate how many samples of relative position can be pooled we asked subjects to estimate the mean positional offset of a set of four triplet alignment stimuli identical to that described above (element uncertainty case, in other words any one of the three elements could be misaligned). These were presented at different, but equi-eccentric, field locations (Fig. 8a). Each of the four, three alignment triplet represented one sample from a positional distribution whose mean was to be judged. To derive the number of positional samples pooled, we used the standard engineering approach of examining how performance deteriorates with the addition of noise (Barlow, 1956; Watt & Hess, 1987; Zeevi & Mangoubi, 1984).

4.2. Methods

4.2.1. Alignment pooling stimuli and task

The stimulus consisted of four three-element alignment stimuli (Gabor spatial frequency 2 c/deg: sigma 0.25°; contrast 80%) positioned around the circumference of a circle of radius 2.5°, centered on fixation (Fig. 8a). Each alignment triplet was identical to that already described in the element uncertainty case, except that because of space constraints the elements were now separated by 1.5°. To render local carrier alignment ineffective, the local phase of the carrier frequency within each patch was randomized from trial to trial. Thresholds were measured for the mean positional offset as a function of the variance of a 1-D Gaussian distribution from which the four samples were derived. Thus four different left/right displacement samples were drawn from a Gaussian distribution with a mean equal to the cued position (i.e. at fixation $\pm$ the cue generated by APE, an adaptive method of constant stimuli) and a variable bandwidth. The subjects' task was a single-interval, binary forced choice that involved the estimation of the mean displacement (was it to the left or right of the central fixation cross?) of the stimulus as a whole. This involved integrating the four independent left/right positional samples to determine the mean left/right displacement of the four triplets as a whole. An APE adaptive method of constant stimuli (Watt & Andrews, 1981) was used to sample a range of positional offsets. Each threshold was obtained from 64 trials, and four thresholds were averaged for each condition. Data were pooled over different runs with a particular stimulus configuration, and a bootstrapping procedure used to fit a cumulative Gaussian function to the results (Dakin, 2001) and derive confidence limits for the model parameters. Given that thresholds are estimates of response variance, the non-ideal behaviour of observers with noiseless stimuli can be expressed as an additive internal noise. The level of internal noise is simply measured by increasing the amount of external noise in the stimulus and determining the point at which observers' performance begins to deteriorate. If the task requires integration then observers' robustness to increasing amounts of external noise will depend decreasingly on internal noise but instead on how many samples are averaged. The form of the variance–summation model is

$$\sigma_{\text{obs}} = \sqrt{\sigma_{\text{int}}^2 + \frac{\sigma_{\text{ext}}^2}{n}}$$

where σ_{obs} is the observed threshold, σ_{ext} the external noise, σ_{int} the equivalent intrinsic or internal noise and n the number of samples being employed. In terms of the positional discrimination task, σ_{obs} corresponds to the threshold position discrimination, σ_{ext}^2 to the variance of the distribution from which the positional samples are derived, σ_{int}^2 to the noise associated with the measurement of each positional sample and their combination and n to the estimated number of positional samples being combined by the visual system.

4.3. Results and discussion

By varying the variance of the distribution of suprathreshold position estimates we measured the positional offset for an array of four, three-element alignment stimuli and derived estimates of the parameter n (the number of positional samples pooled by the visual system) in the above equation by fitting the standard summation–variance model. An example of the data (symbols) and model fits (solid curves) are shown in Fig. 7. We undertook this analysis at a number of different exposure durations to gauge its effect on pooling. The derived sampling parameter from the model fits shown in Fig. 7 is displayed in Fig. 8 where the estimated number of samples are plotted against exposure duration (error bars represent $\pm$1 SD). For the shortest exposure duration (i.e. 125 ms) only one sample was used, whereas at the longest exposure duration (1 s) two samples were used (Fig. 8). The fact that samples could only be accumulated at such a slow rate (1 sample/s) suggests a serial process, possibly driven by changes in focal attention. A similar investigation of orientation coding has shown that the visual system can integrate hundreds of local orientation samples across the field within a 100 ms exposure, suggesting parallel processing (Dakin, 2001).

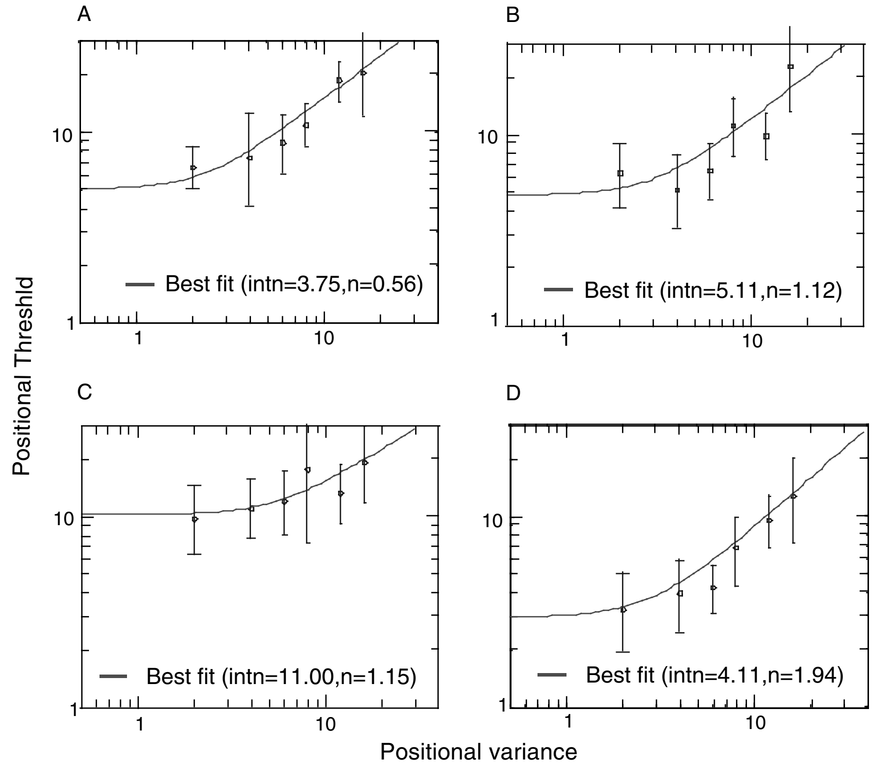

Fig. 7. Sample data from one subject (RFH) for four exposure durations for the positional pooling task. The stimulus which is depicted in Fig. 6a consists of four independent samples from a positional distribution. Here we plot positional thresholds in pixels (1 pixel = 1.1 min at 1 m) against the variance of the distribution from which the samples have been chosen. The data are fit by a standard summation–variance model (see text) from which we derive estimates of intrinsic noise and sampling. The derived sampling parameter, *n* (see text) is given in the figure inset.

5. General discussion

Although we are highly accurate at discriminating the relative positions of two features in the image, the present results suggest that the visual system has problems in doing this in more than one position in the visual field at the same instant. Our ability to utilize multiple relative positions to derive the orientation of a Gaussian texture bar defined by relative position is poor (experiment 1). Over the range of element disarray investigated here, perceived density was not a factor in determining performance. Not only do we not benefit from having multiple estimates of relative position across the field, such information only makes performance worse. This suggests that if the processing of relative position is a low level process that operates in parallel across the visual field the output of these calculations are not made available to higher levels of processing, where possibly attention–driven process are limited to a relatively few estimates of relation position.

This conclusion is also supported by the results using the three- and four-element alignment task. Performance is substantially worse when there is uncertainty as to which element is displaced and much worse than predicted by an ideal observer model that encodes all elements with equal sensitivity. Furthermore, performance for the positional task is much worse than that found for a comparable task involving orientation, which we can process in parallel across the field (Dakin, 2001). This argues for the special status of relative position and is consistent with the conclusion that at the level of perception only one relative position (i.e. relative to a known reference) can be judged. Additional estimates of relative position can only be accumulated over extended periods of time.

Finally, we are also limited in how many separate suprathreshold estimates of relative position that can be integrated at any one visual field eccentricity. Within a typical perceptual processing time of 500 ms, only one estimate can be used, additional estimates can be pooled but only over a time scale of seconds, suggesting serial search and a more cognitive, higher level strategy. This finding for position coding is very different to that for orientation coding where there is evidence that the visual system can integrate multiple samples in parallel across the field within a 100 ms exposure (Dakin, 2001).

There is a general belief that position is extracted at multiple points across the visual field and used in low level visual processes (Marr, 1982; Watt, 1988). As-

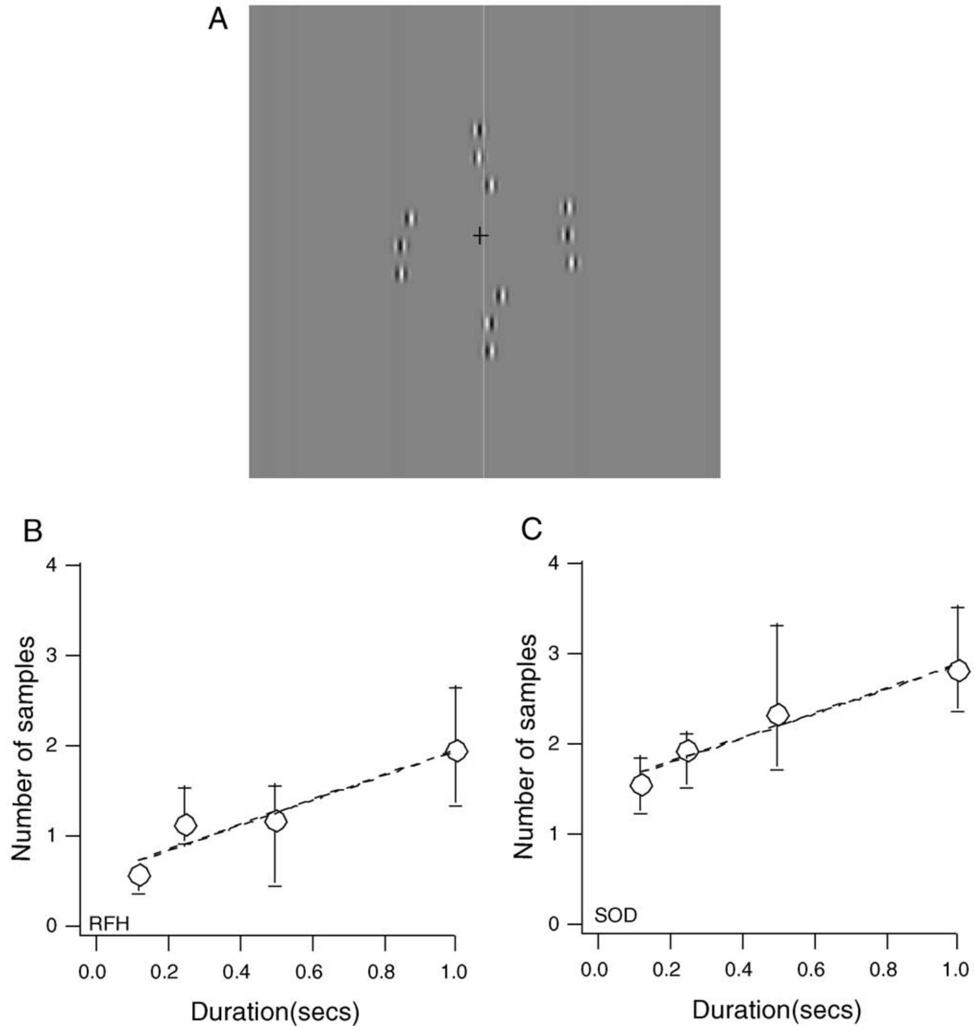

Fig. 8. In A, the stimulus for the positional pooling task is illustrated. It is composed of 4, triplets. To remove any possible shape cue to position, the phases of the individual elements were randomized as was the position of the elements that carried the positional signal within any triplet. In B, the sampling parameter derived from the fits in Fig. 7 are plotted (together with ±1 SD) as a function of exposure duration for two subjects. For brief exposure durations, approximately 1 sample can be utilized, additional samples take time.

suming that this is so, the present results argue for only a limited capacity for encoding position at the perceptual level. The picture that emerges from this study is not of the encoding of relative position in a parallel fashion across the central field at the level of conscious perception but one where relatively few, and possibly only one relative position is encoded at a time. Additional positional samples can only be accumulated slowly suggesting that the process is cognitive in nature and involves the role of attention. This apparent limitation represents an important specialization, our positional coding excels in one location rather than being mediocre in many. Thus it may be hardly surprising that just before brief saccadic eye-movements, when attentional processes are momentarily interrupted, that our whole positional framework is so easily disrupted (Ross, Morrone, & Burr, 1997).

Acknowledgements

This work was supported by a CIHR grant (MT 108-18) to RFH. We thank Harriett Allen and Dave Ellemberg for helpful comments.

Appendix A

The ideal observer encodes all positions. This estimate of position is not absolute (i.e. relative to some fixed screen coordinate) but relative to the mean position of others items. Zero-mean Gaussian noise is added to each positional estimate. The ideal observer's judgments are based on these positional estimates. This process is repeated ($n = 1000$) to get a percent correct measure. A threshold is then determined at 78% correct.

In experiment 1 the display consists of a square grid of Gabors where positional variance is added to the elements within a Gaussian bar. This bar is either oriented vertically or horizontally. The ideal observer determines the orientation of the bar by comparing the variance in the positional estimates in the vertical and horizontal directions. This comparison is based on actual element positions and is independent of the original grid positions. More specifically for a vertical positional variance estimate, the display is segregated into columns corresponding to the original, unjittered, grid positions. The standard deviation of the orthogonal (horizontal) vectors relative to the column's mean is computed. The standard deviation of the central 10 columns is then compared to the same estimate in the horizontal direction. The largest estimate is taken as the orientation of the Gaussian noise bar.

The five-element display in experiment 1 is constructed in an identical fashion as the multi-element display described above except that only the central five elements are visible. In this case we consider only two bars, each consisting of three (or two, if the middle element is ignored) elements. The standard deviations of the positions orthogonal to the mean of the bars is computed. The bar with the larger standard deviation indicates the orientation of the perturbation. This procedure is identical to that described above for the multi-element display, except that the middle element was ignored. Ignoring the middle element improves the ideal observer's performance compared to when the middle element is included in the analysis. This is because since the middle element is always perturbed it adds noise to both the horizontal and vertical dimensions. Furthermore this procedure corresponds to the reported strategy followed by the subjects who reported that they tried to ignore the middle element.

The performance of the ideal observer increased with increasing amplitude of the noise bar (Fig. 9). For both strategies, performance for the five-element display is worse than the performance in the multi-element case. We chose the most conservative case, namely ignoring the central element which in fact matched with what subjects reported they were using to solve the task. In this case, performance is only nine times worse in the five-element case compared with the multi-element case.

In experiment 2, the ideal observer determines the displacement (left or right) of a particular element or target relative to the mean location of all elements. To model a beneficial attentional effect (element certainty case), the noise added to the targets location is a variable fraction of that added to the other reference elements' locations.

For the element certainty case, the target element is known. For each true target displacement, the ideal observer makes an estimate of this displacement (left or right) relative to the mean position of all elements. For an uncertainty case, an identical procedure is followed but treating each element as the target, resulting in two or more direction–displacement judgments. When two judgments are made the largest displacement is taken as the correct one. When more than two judgments are made, the majority of these judgments are identical and assumed to correspond to the flankers, the odd one is taken to be that of the target.

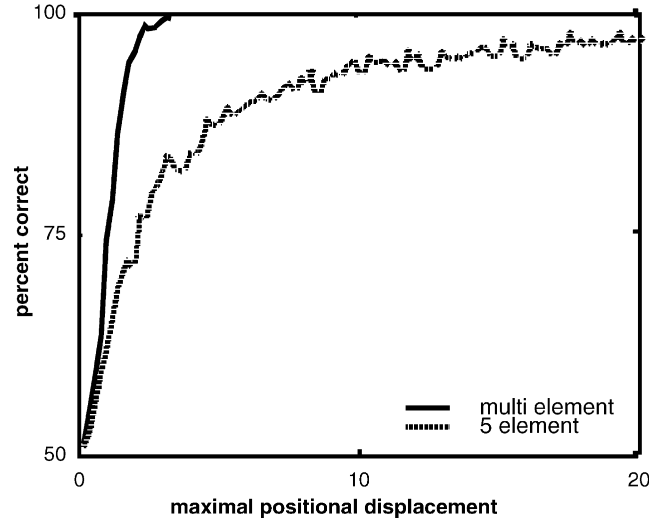

Fig. 9. Percent correct as a function of the 2-D element disarray for the orientation discrimination of a 1-D Gaussian profile. Two cases are compared; the multi-element display and the five-element display.

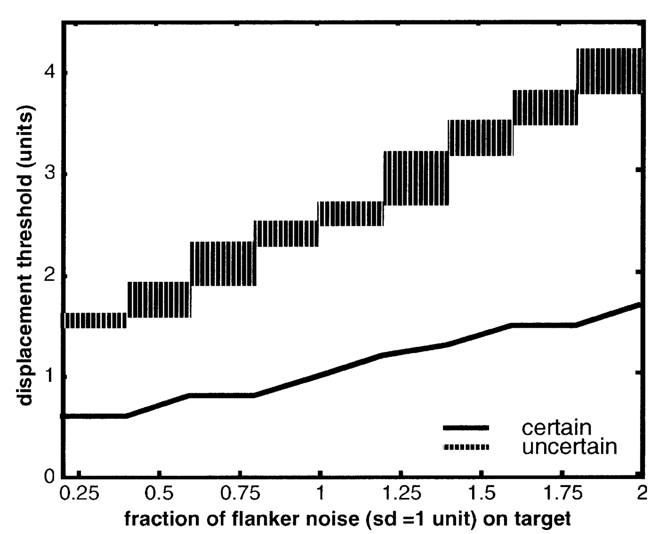

Fig. 10. Threshold displacement plotted as a function of the proportional noise reduction associated with the positional coding of an attended element in a *n*-element alignment task. The certain case refers to when the identity of the misaligned element is known to the observer whereas in the uncertain case the misaligned element could be any one of the *n*-elements in the target. Ideal performance is about a factor of 2.5 worse in the uncertain case irrespective of fraction of noise reduction associated with a knowledge of the identity of the misaligned target.

In both cases the fraction of noise added to the target element, as compared to the other elements, was varied to determine the beneficial influence of a local noise reduction due to attention. That is, in the element uncertainty condition, each of the elements considered to be the target was subject to a different amount of displacement-noise.

The absolute thresholds increased as a function of the relative amount of noise added to the target element (Fig. 10). The three-element uncertainty condition showed consistently higher thresholds than those for the element certainty condition. The ratio of the thresholds of element uncertainty to certainty conditions however remained constant at 2.5. This implies that the weighting we give to attentional factors has no effect on the relative performance between certainty and uncertainty conditions.

References

Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perceprion and Psychphysics, 49*, 303–314.

Barlow, H. B. (1956). Retinal noise and absolute threshold. *Journal of Optical Society of America, 46*, 634–638.

Brainard, D. H. (1997). The Psychophysics Toolbox. *10,* 433–446.

Carney, T., & Klein, S. A. (1999). Optimal spatial localization is limited by contrast sensitivity. *Vision Research, 39*, 503–511.

Dakin, S. C. (2001). An information limit on the spatial integration of local orientation signals. *Journal of the Optical society of America A, 18*, 1016–1026.

Hess, R. F., & Holliday, I. E. (1996). Primitives used in the spatial localization of non-abutting stimuli: Peaks or Centroids? *Vision Research, 36*, 3821–3826.

Marr, D. (1982). *Vision*. San Francisco, California: Freeman.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research, 31*, 1337–1347.

Ross, J., Morrone, M., & Burr, D. C. (1997). Compression of visual space before saccades. *Nature, 384*, 598–601.

Toet, A., & Koenderink, J. J. (1988). Differential spatial displacement discriminations for Gabor patches. *Vision Research, 28*, 133–143.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science, 255*, 419–423.

Watt, R. J. (1988). *Visual processing: computational, psychophysical and cognitive research*. London: Lawrence Erlbaum Associates.

Watt, R. J., & Andrews, D. (1981). APE: Adaptive probit estimation of psychometric functions. *Current Psychological Review, 1*, 205–214.

Watt, R. J., & Hess, R. F. (1987). Spatial information and uncertainty in anisometropic amblyopia. *Vision Research, 27*, 661–674.

Watt, R. J., & Morgan, M. J. (1983). Mechanisms responsible for the assessment of visual location: Theory and evidence. *Vision Research, 23*, 97–109.

Waugh, S. J. (1998). Masks reveal temporal processing of position. *Investigative Ophthalmology and Visual Science, 39*(suppl), s622.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics, 18*, 292–297.

Zeevi, Y. Y., & Mangoubi, S. S. (1984). Vernier acuity with noisy lines: Estimation of relative positional uncertainty. *Biological Cybernetics, 50*, 371–376.