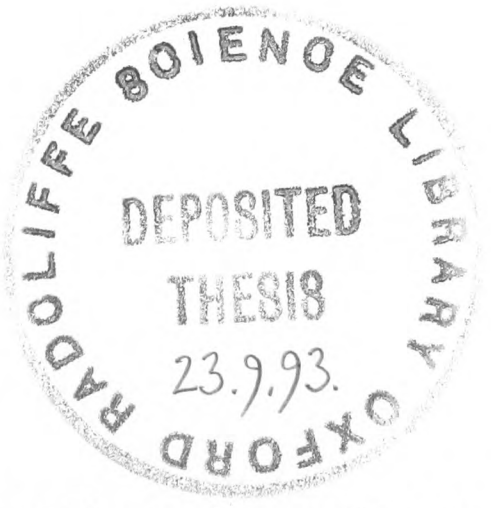
The Role of Spatial Scale in Binocular Stereopsis

Andrew Glennerster



Thesis submitted for the degree of Doctor of Philosophy,
University of Oxford
Hilary 1993

For Kate

Acknowledgements

I am indebted to many people who have helped me during the last three and a half years. Above all, I would like to thank my colleagues and friends in the "Vision Lab" in the Department of Experimental Psychology at Oxford.

In particular, I am very grateful to Dr. Brian Rogers, my supervisor, for his help throughout; to Dr. Michel Treisman for useful comments on a first draft of the thesis; to Mark Bradshaw for teaching me to program the Macintosh and for somehow managing to resolve any crisis; to Richard Eagle for three years of animated discussion; and to Billy Lee, Richard and Mark for acting as subjects. I would also like to thank my parents for their support and encouragement, and Roger Watt and Graeme Mitchison for many useful discussions.

Finally, I thank Kate for endless patience and encouragement; and Rose, for learning to sleep.

Abstract

The role of spatial scale in binocular stereopsis.

Andrew Glennerster, St. John's College, Oxford

Thesis submitted for the degree of Doctor of Philosophy,
Hilary Term, 1993

A model of stereopsis is proposed in which information from each eye's image is organised as a scale-based hierarchy before binocular comparison. The algorithm incorporates coarse-to-fine matching (like Marr and Poggio, 1979) but differs from previous models in that the position, and hence disparity, of features is defined relatively rather than by their retinal co-ordinate. Thus, fine scale disparities are measured and recorded relative to coarse scale disparities. Local surface slant and curvature is represented explicitly at a range of spatial scales. The theory is based on a hierarchical model of encoding position (Watt , 1988).

The first experiment investigates the time course of shape discrimination in random dot stereograms. The results are compatible with a model in which the scale of analysis changes from coarse to fine over the first second of viewing. The second experiment measures the magnitude of a new "3-D" Müller-Lyer illusion and compares it to that of the classical (2-D) illusion. Both these and the cyclopean Müller-Lyer illusion are consistent with a model in which hierarchical encoding of position is used by the visual system for 2-D (length comparison) and 3-D (slant) judgements. The third experiment compares the detection of large disparities and large displacements. "D_{max}" for the motion and stereo tasks is shown to be similar over a wide range of dot densities. The results are interpreted as evidence that similar spatial primitives are used in the correspondence process in both domains. The spacing of MIRAGE centroids (Watt and Morgan, 1985) fit the data well.

The proposed hierarchical model is similar to that put forward by Mitchison and McKee (1987), although their scheme was not based on spatial scale. The model bridges the gap between a primal and a 2 1/2-D sketch (Marr, 1982) and has important implications for many issues within stereopsis.

Extended Abstract

The role of spatial scale in binocular stereopsis.

Andrew Glennerster, St. John's College, Oxford

Thesis submitted for the degree of Doctor of Philosophy,
Hilary Term, 1993

Binocular stereopsis is a promising area of research in vision because the computational issues within it are well defined. The nature of the input to the process (the optic array at two slightly different vantage points) and of the output (information about the relative depth of objects in the scene) is relatively well understood. However, there is little agreement on the type of algorithm which best models human stereopsis.

The problem of understanding binocular stereopsis can be considered in two stages: (i) the comparison or correspondence process and (ii) the interpretation of disparities. This thesis is primarily concerned with the first of these issues (the correspondence problem), although the solution which is put forward has important implications about the way in which disparities may be interpreted.

The first chapter of this thesis examines different approaches to the correspondence problem. The coarse-to-fine matching algorithm proposed by Marr and Poggio (1979) is examined in detail. This is an elegant computational theory but the results of subsequent psychophysical experiments have failed to support it as a model of human stereopsis. In the second chapter a coarse-to-fine method of encoding the (2-D) position of features in an image is explored (Watt, 1988). In the third chapter, the main elements of these two theories are combined.

The synthesis provides a theoretical framework for the experiments described in this thesis. The principal feature of the model is that fine scale disparities are encoded relative to coarse scale disparities. The experiments in this thesis were designed to examine some of the predictions arising from this model.

The first set of experiments investigated the time course of shape discrimination in random dot stereograms. The target within the stereogram was a disparity-defined rectangle which subjects had to identify as either horizontally or vertically

elongated in a 2-alternative forced-choice paradigm. The threshold height:width ratio of the rectangle required for reliable discrimination of the target's orientation was measured as a function of exposure duration. Three types of stimulus were used in the main experiment: (i) the target was defined by a uniform disparity of 2 pixels; (ii) odd rows (raster lines) within the target region had a crossed disparity of 3 pixels, even rows a disparity of 1 pixel, i.e. the same average disparity as the first case; and (iii) odd rows had a crossed disparity of 1 pixel, even rows an uncrossed disparity of one pixel, i.e. the same average disparity as the background. The results show that threshold height:width ratios for the first two types of target are very similar and vary very little with exposure duration (from 1 second down to 60 ms). The pattern of thresholds for the third type of stimulus is entirely different. Thresholds for exposure durations less than about 80 ms were very high - at the shortest exposures too high to measure - and between 80 and 100 ms thresholds fell rapidly to values similar to those for targets of type (i) and (ii). For exposures of 1000 ms the thresholds for all three target types were similar.

The results from this experiment are interpreted as evidence of a coarse-to-fine process in stereoscopic matching. Several variations on the main experiment support this conclusion. In the first, the frequency of disparity modulation (strip width) for the type (iii) stimulus was varied. Subjects could perform the task at shorter exposures for lower frequency disparity modulations which is consistent with a coarse-to-fine theory. In the second, thresholds for a uniform target with a disparity of 1 pixel were found to be low even at the shortest exposure, illustrating that it is not simply the smaller disparity of the type (iii) target which makes it difficult to see at short exposures. Results were also collected for uncorrelated targets (the background is correlated), and other mixed disparity targets. Thresholds for all of these targets were lower than for the type (iii) target. In the final experiment in this chapter, low-pass filtered versions of type (i) and (iii) stimuli were displayed at a constant exposure duration. The results show that thresholds for the type (iii) stimuli changed sharply over a narrow range of blur (just as they did over a narrow range of exposure durations in the first experiment) whereas thresholds for the uniform disparity stimulus varied very little over the range of blur tested. The data from the experiments varying exposure duration and varying blur were sufficiently similar to allow them to be used to derive an approximate time course of the coarse-to-fine stereoscopic process.

The second set of experiments investigated a new "3-D" Müller-Lyer illusion in which the fins-in figure of the Müller-Lyer illusion was displayed to one eye and

the fins-out figure to the other eye. Most subjects perceived the binocularly fused shaft as slanted in depth although the shaft had the same physical length in both eyes' images. (The illusion is most striking when the fins are not too long (less than half a degree).) A 2-alternative forced-choice paradigm was used to measure the magnitude of the illusion. The subjects' task was to identify the slant of the shaft with respect to the fronto-parallel. The actual length difference of the shaft on the left and right eye's images was varied. The bias (50% correct point) of the psychometric function was defined as the extent of the illusion. This was compared, for a range of fin lengths, with the extent of the classical 2-D Müller-Lyer illusion. For fin lengths up to half a degree the agreement is very close.

These findings suggest that similar processing precedes binocular comparison of stimuli (the 3-D Müller-Lyer illusion) and comparison of stimuli across space or time (the 2-D Müller-Lyer illusion). In particular, the results are interpreted as consistent with a hierarchical model in which each eye's image is encoded hierarchically before binocular comparison for stereopsis.

The "cyclopean" (random dot stereogram) version of the Müller-Lyer illusion was studied using a similar paradigm. The extent of the cyclopean illusion was shown to be similar to that of the 2-D illusion over a wide range of fin lengths, as has been reported previously (Julesz, 1971). All three Müller-Lyer illusions (2-D, 3-D and cyclopean) are interpreted (and modelled) in terms of a hierarchical scheme, i.e. that length differences between the fins-in and fins-out figures found at a coarse scale fail to be corrected by fine scale information. It is proposed that the reason for this might be that the visual system is good at comparing information stored at one scale in the hierarchical description of an image but not good at combining information across spatial scales.

In the final set of experiments "stereo-d_{max}" was measured for random dot patterns in an analogous paradigm to that used for measuring the upper displacement limit (d_{max}) for motion. The stimuli consisted of large (16 by 21°) random dot stereograms in which all the dots were given a uniform disparity. The density of dots ranges from 0.006% (two dots) to 50%. The exposure duration of the stimuli is 150 ms. In a 2AFC paradigm, subjects were asked to identify the pattern as in front of or behind the screen. The disparity at which subjects made 20% errors was defined as "stereo-d_{max}".

Motion-d_{max} was determined for the same set of stimuli. That is, the stimuli were presented not as a stereo pair but as a two-frame apparent motion sequence. Exposure duration was 150 ms, with no inter-stimulus interval. In this case the task was to identify the direction of motion of the dots (left or right). The displacement at which subjects make 20% errors was defined as motion-d_{max}.

Eagle and Rogers (1991) have shown that, for a large patch size (25 by 25°), motion-d_{max} rises as dot density is reduced according to a power law (a slope of -0.2 when d_{max} is plotted against dot density on log-log axes). The results of the experiment reported in this thesis confirm their finding for motion-d_{max} and show that stereo-d_{max} follows the same pattern: the slope of the function and the absolute value of d_{max} for motion and stereo is similar over the whole range of densities tested. The results are interpreted as evidence that similar limitations apply to the motion and stereo correspondence processes (at least, under these conditions).

The same pattern of results was obtained for the stereo experiment for stimuli made up either of white dots on a black background or black dots on a white background. (Eagle (1992) has shown a similar result for motion.) This suggests that mean luminance, and measures of contrast which depend on mean luminance, do not account for the change in d_{max} with dot density. Rather, it is argued that d_{max} may reflect the density of false targets in the stimulus. If this is the case, then the similarity between the results for stereo and motion suggests that similar spatial primitives are used in the stereo and motion correspondence processes.

It is shown that the spacing of 1-D MIRAGE centroids (Watt and Morgan, 1985) varies with dot density at a rate which fits well with the experimental data. A simple matching algorithm is used to derive a "model-d_{max} " and this confirms that MIRAGE centroids are a suitable candidate primitive. The behaviour of the MIRAGE output at different densities, and the reason centroid spacing changes smoothly with dot density, are discussed.

In the last chapter, the evidence in favour of a hierarchical model is reviewed. Some of the best evidence that stereoscopic matching of fine scale features is affected by the coarse scale grouping in an image comes from experiments by Mitchison and McKee (1987, discussed in chapter 5). Despite being described in very different terms, the two models share many properties and both can explain their experimental results while other models, such as co-operative algorithms or Marr and Poggio's (1979) algorithm, cannot. Finally, other topics within binocular

stereopsis are examined in relation to a hierarchical model. Several areas which have traditionally been considered in terms of interactions *after* the correspondence process - for example, the anisotropy of slant perception, simultaneous depth contrast and the scaling of disparity gradients with viewing distance - are considered instead in terms of a hierarchical system of encoding position.

In summary, the proposal examined in this thesis is that the information within images is organised as a scale-based hierarchy. The theoretical advantage of this type of organisation is that large amounts of data can be handled efficiently. The results of experiments reported in this thesis suggest that a hierarchical model may be an appropriate description of human stereopsis.

Contents

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 1

1.1 What is binocular stereopsis?

We can see the three dimensional structure of the world around us with one eye closed, particularly when we move our heads. But an extra sensation of depth is present when we open both eyes. The process that yields this sensation is called binocular stereopsis (literally, seeing shape with two eyes). In some situations, stereopsis is the *only* cue to the presence of an object as, for example, when its texture is identical to the background and neither the observer nor the object are moving. Breaking camouflage may have been one of the evolutionary pressures for the development of stereopsis.

Marr (1982) has emphasised that in studying any visual process it is important to understand the constraints on the input and the purpose of the output of that process. In this respect, stereopsis is a promising area for research since the constraints on the input should theoretically be well understood. The principal

information on which binocular stereopsis is based is the difference between the left and right eyes images. This information arises entirely from the separate position in space of the two eyes. In addition, the structure of the disparity field is constrained by the structure of the physical world (e.g. the fact that most light in the optic array is reflected from relatively smooth, opaque surfaces). The most important output from binocular stereopsis is information about the relative depths of objects in the scene.

Many different algorithms have been developed that can derive this output from stereoscopic images (e.g. Grimson, 1981; Pollard, Mayhew and Frisby, 1985) and some can work in real time on natural images. However, despite radical differences between the algorithms and the constraints they use, there is no agreement on which one best models human stereopsis.

1.1.1 The correspondence problem

There are two central problems that any stereo algorithm must face: the correspondence problem and the interpretation of disparities. The correspondence problem is to determine which feature in the left eye's image matches which in the right eye's image. Marr and Poggio (1979) describe it saying:

> *"Three steps (S) are involved in measuring disparity: (S1) a particular location on a surface in the scene must be selected from one image; (S2) that same location must be identified in the other image; and (S3) the disparity of the two corresponding points must be measured.*
> *If one could identify a location beyond doubt in the two images, for example by illuminating it with a spot of light, steps S1 and S2 could be avoided and the problem would be easy. In practice one cannot do this..and the difficult part of the computation is solving the correspondence problem."*
> (Marr and Poggio, 1979, p301)

This thesis is primarily concerned with the first problem, that is, how correspondence is achieved.

1.1.2 The interpretation of disparities

The second problem, that of interpreting disparities, is much less well defined. There is no general agreement on the nature of the output of stereoscopic processing, over and above the detection of depth differences. It is possible that very little processing takes place after the detection of disparities. Morgan (1989) has suggested that the main roles of stereopsis are to break camouflage and to direct fine hand movements (e.g. picking fruit in a bush). Neither of these tasks require much more analysis beyond the detection of disparity and, in the case of

fruit picking, the ability to null a disparity with a hand movement. There are several phenomena of human depth perception that are often considered as evidence that further processing occurs after the measurement of disparities (e.g. the interpolation of depth (Collett, 1985); depth contrast (Graham and Rogers, 1983); the anisotropy of slant perception (Wallach and Bacon, 1976) etc.) but, as discussed in the final chapter, some of these may in fact be due to interactions at a stage before disparities are calculated.

1.2 The focus of this thesis

1.2.1 Why concentrate on the correspondence problem?

The central issue tackled in this thesis is the question of how the two eyes' images are compared for stereopsis. That is, what are the features in each eye's image that are "labelled" so that they can be matched with the corresponding feature in the other eye's image? Behind this question is a more fundamental one which goes beyond stereopsis: What aspects of the image are recorded explicitly in the visual system, not only for the purposes of comparing the two eyes' images for stereopsis, but for other visual tasks such as judgements of length or describing changes in the image over time?

Images contain an enormous quantity of information. The correspondence problem is one of comparing two very large data structures. The approach to the correspondence problem taken in this thesis is to ask first, how that data is organised, and only then how it can be compared with data in another image. Chapters 2 and 3 explore this idea in detail.

A second reason for studying the correspondence problem is that the input and output of the process are relatively well defined. As was pointed out in section 1.1, this is an important prerequisite for developing an algorithm to model any aspect of vision. But what happens *after* the correspondence process is less clear. One reason for this is that any subsequent process is dependent on the output of the correspondence process (i.e. the nature of the *input* to the next stage is not well defined). The other reason, as discussed in section 1.1.2, is that there is no general agreement on the goal of subsequent processing of disparities (i.e. the nature of the *output* of the next stage is not well defined).

1.2.2 Coarse-to-fine matching

The correspondence algorithm that is discussed in detail in this chapter (by Marr and Poggio, 1979) and also the one proposed in this thesis (chapter 3) are coarse-to-fine algorithms. That is, the two eyes' images are compared first at a very coarse scale, so that there are relatively few, widely spaced features in each image and they can be matched easily. The result of the coarse scale matches are then used to guide matching of finer scale features.

Coarse-to-fine matching is, in theory, a simple and elegant approach to the correspondence problem. However, psychophysical evidence does not support the implementations of coarse-to-fine models that have been proposed so far (e.g. Marr and Poggio, 1979). The details of the theory and the relevant evidence are discussed in section 1.4.

In contrast to coarse-to-fine matching algorithms, most other models consider many possible matches at an initial stage and then eliminate "false" matches according to a set of rules. This type of ("co-operative") algorithm is discussed in section 1.3. The principles behind co-operative algorithms are less relevant to the theory pursued in this thesis and so are discussed in less detail than Marr and Poggio's coarse-to-fine model. Co-operative algorithms have been implemented successfully (e.g. Pollard, Mayhew and Frisby, 1985) but they fail to account for some psychophysical evidence (e.g Mitchison and McKee, 1987a and b, discussed in section 1.3.3). The same evidence also presents problems for Marr and Poggio's theory, and is an important challenge for any model of stereo correspondence in human vision.

1.2.3 A new way of thinking about disparity

In chapter 2, a theory about how image data may be organised in the human visual system is examined (Watt, 1988). The key idea of the theory is that the position of a feature in the image is defined *relative* to other features rather than by recording its "local sign" (retinal co-ordinate) and that this information is organised as a hierarchical database, with coarse scale information at the top of the hierarchy. This theory has several important implications for a theory of stereopsis. First, it suggests a logical strategy for comparing information about the left and right eyes' images. Second, it fits very well with coarse-to-fine matching algorithms, although it differs in important respects from those described so far. Third, because it records information about the *relative* position of features, it exploits a useful property of the binocular optic arrays, which is that relative disparities of features in the image do not change with the vergence state of the eyes. This means that

disparities do not have to be re-computed (or stored in a memory buffer like the $2^1/_2$-D sketch (Marr and Poggio, 1979)) when fixation changes. Fourth, a hierarchical system for encoding the position of features emphasises some pieces of information about the image at the expense of others. That is, some information is recorded explicitly and other information only implicitly (an idea discussed more fully in section 5.1)

The experiments in this thesis are designed to test how far a hierarchical theory can model the correspondence process in human stereopsis.

1.3 Previous approaches to the correspondence problem

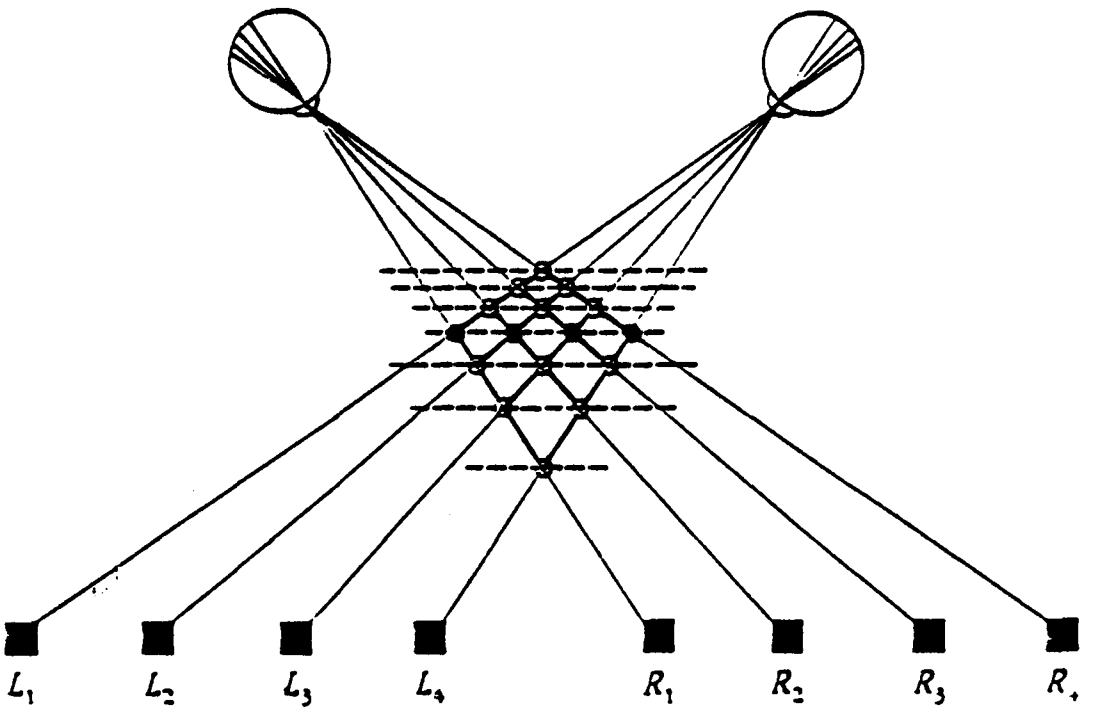
1.3.1 Constraining the problem

Figure 1.1 illustrates how difficult it can be to match the features in two images unless the problem is constrained. The images shown in this example contain only four dots and yet there are a very large number of possible configurations that would give rise to this pair of images, some of which are shown in figure 1.1. Images containing more than four dots would give rise to very many more possible matches. How is it, then, that human observers interpret most stereoscopic images, including that illustrated in figure 1.1, unambiguously?

Marr and Poggio (1976) argue that constraints that apply to physical objects can be used to limit the possible solutions. In other words a set of rules can be used to distinguish between solutions that are more or less likely in the real world. The type of algorithm that implements this strategy is described in the next section.

1.3.2 Co-operative algorithms

In most co-operative algorithms there is, in theory, a node or "neuron" corresponding to every possible match at every retinal location in each eye's image. That is, in figure 1.1, there would be one "neuron" for each of the 16 intersections of the grid. (This applies to Marr and Poggio, 1976; Dev, 1975; Pollard et al., 1985; and Prazdny, 1985. Julesz's (1971) dipole model is an interesting exception, discussed below.) The final solution to the correspondence problem in these algorithms takes the form of a stable pattern of "active nodes" in the network that represent the disparity at each point on the surface. In the first stage of the algorithm, all possible matches are represented (there is activity in



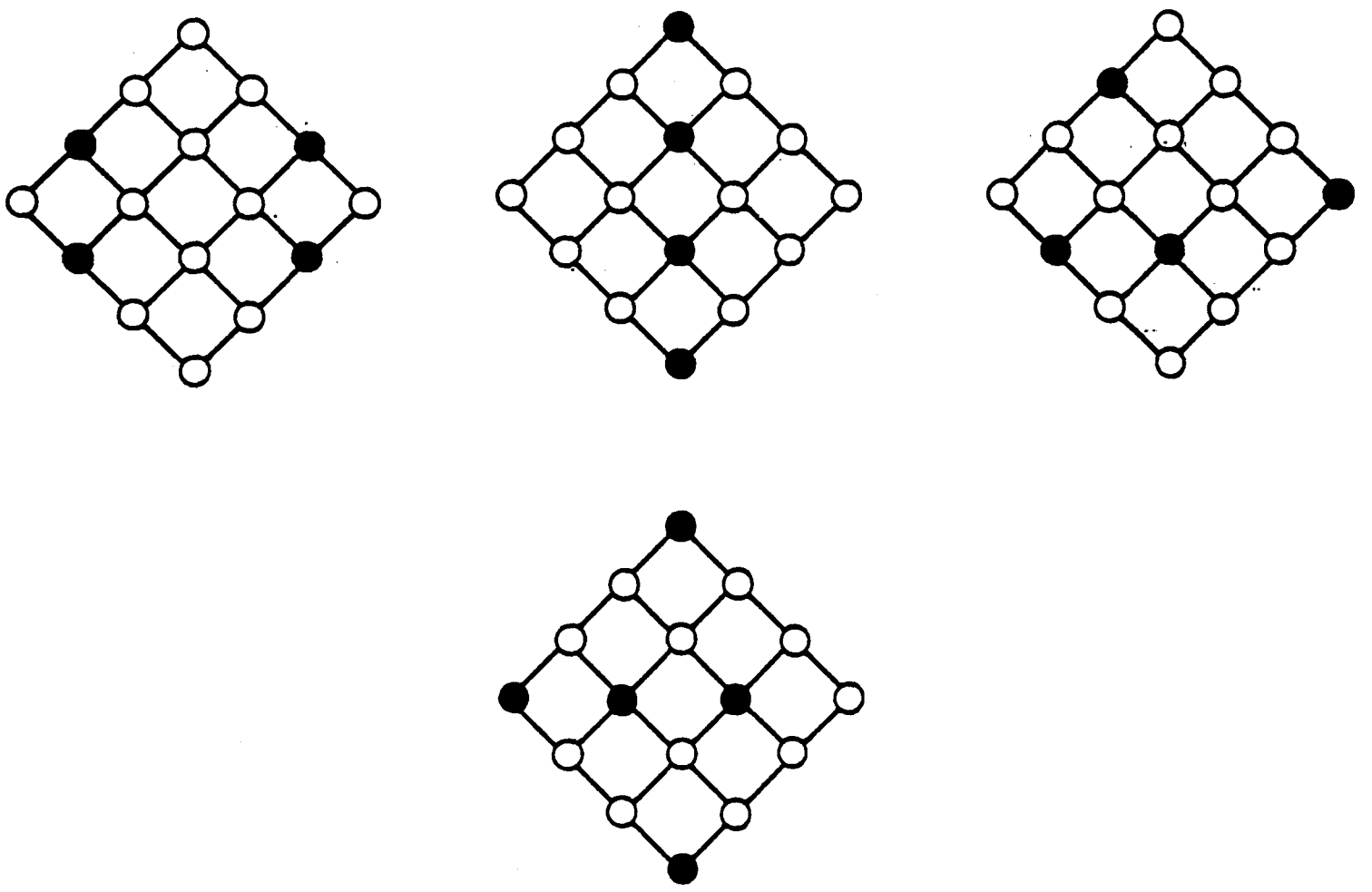


Fig 1.1

An illustration (taken from Marr and Poggio, 1976) of the correspondence problem given only four features in the left and right eyes' images. Each of the four points in one eye's view could match any of the four projections in the other eye's view. The four filled circles show the "correct" matchings. Many other configurations would produce the same pair of images, some of which are shown below. The three in the centre comply with the "uniqueness" constraint, i.e. any feature in the one eye's image matches only one feature in the other eye's image (there are another 21). A possible pattern which does not obey the uniqueness constraint is shown at the bottom.

many nodes, e.g. in figure 1.1 all 16 nodes). A series of rules about how nodes in the network interact ensures that, over several iterations, activity in most of the nodes decreases leaving only one pattern of activity.

The rules used differ slightly between models. In Marr and Poggio's (1976) algorithm three rules are used. First, possible matches are only sought between elements in the two images that could have arisen from the same physical object. (In figure 1.1 this means black dots are matched only with black dots. Marr and Poggio called this the *compatibility* constraint.) Second, only one match is sought for each element in the other eye's image. (In figure 1.1 this means that in the final solution there should be only one active node along each grid line (line of sight). Marr and Poggio called this the *uniqueness* constraint.) Third, it is assumed that, in general, disparity varies smoothly across the image. (In figure 1.1 this means that active nodes in horizontal rows should support one another. Marr and Poggio called this the *continuity* constraint.) A summary of the constraints used in many stereo algorithms is given in Mayhew (1983).

The rules in Dev's (1975) algorithm are essentially the same except that the uniqueness constraint is applied to lines through the cyclopean point rather than lines of sight from each eye. This has less validity than Marr and Poggio's version (because one point necessarily obscures another if, and only if, the two points lie on the same line of sight).

The algorithms by Pollard et al. (1985) and Prazdny (1985) build on Marr and Poggio's (1976) model by adding an extra constraint, a disparity gradient limit. A disparity gradient is defined as the disparity of two points divided by their cyclopean separation (Burt and Julesz, 1980). The theoretical limit of the disparity gradient limit, without violating the ordering constraint, is 2. Burt and Julesz found that, for stimuli consisting of two dots, diplopia occurred for disparity gradients above about 1. Prompted by this psychophysical finding, Pollard et al. (1985) introduced a disparity gradient limit of 1 in their algorithm. This means that support for a match is only sought from other nodes whose disparity gradient with respect to that node is less than 1. It corresponds to an assumption that real surfaces, in general, are slanted less than a certain amount with respect to the observer (the exact slant depends on the viewing distance). Prazdny's version, rather than having a sharp cut-off to the tolerated disparity gradient, uses a Gaussian function but otherwise the model is very similar.

Julesz's (1971) model is co-operative, too, although the co-operation is described in mechanical terms. It provided the inspiration for several of the algorithms described above. In his model there is only one node or "dipole" for each line of sight in each eye's image. The orientation of the dipole indicates the disparity at that point. The tips of the dipoles are joined by springs so that they tend to move together, i.e. this is a version of the continuity constraint. One interesting feature of the model is that

> *"Since stereopsis is the sensation of relative depth, we postulate that only differences between dipole rotations contribute to stereopsis. Thus, it is the same whether dipoles in the surround rotate while those in the centre* [i.e. in the region of a disparate square] *do not, or vice versa;"*
>
> (Julesz, 1971, p206)

This aspect is similar to the model proposed in chapter 3.

1.3.3 Mitchison and McKee

Surprisingly little work has been specifically directed at determining the nature of the correspondence process in human stereopsis. As Poggio and Poggio (1984) say in their review:

> *"The constraints that the matching process obeys are only indirectly explored with psychophysical techniques...On the whole, the question of the nature of the primitives and the constraints of matching waits for new psychophysical experiments."*
>
> (Poggio and Poggio, 1984, p390)

An important exception is the work by Mitchison and McKee (1987a and b, Mitchison, 1988). They set out specifically to discover the matching rules in human stereopsis. Their results are difficult to explain either within the framework of the co-operative algorithms discussed above or in terms of Marr and Poggio's (1979) algorithm. In fact, they challenge the accepted notion that the retinal or absolute disparity is the basis for matching. Mitchison and McKee's experiments form the basis of the in some detail.

In their basic paradigm, a grid of regularly spaced dots was presented to the left and right eyes. A regular pattern of repeating elements can be matched as a planar surface at several depth planes (as for the wallpaper illusion) in which case each element is matched with its neighbour to the right (or left) in the other eye's image. The edge dots in Mitchison and McKee's patterns, on the other hand, had an unambiguous match in the other eye's image (provided the ordering constraint was

not violated). (A schematic version of their stimuli is shown in figure 1.2.) Mitchison and McKee were interested in what factors influenced the choice of match. In particular, they manipulated the disparity of the dots at the edges of the grid and investigated the effect this had on the perceived depth of dots in the centre of the grid. They carried out their experiments both at short exposure durations, too brief for eye movements to take place, and with unlimited exposures.

They found that, for short exposures (150 ms), the internal dots were perceived at the same depth as the edge dots, even when the internal dots had no discrete match at this depth. For longer (unlimited) exposures, the depth of the internal dots almost always corresponded to a discrete match but the percept was bistable - a dot was sometimes seen in the forward plane, sometimes in the fixation plane - and the probability of these two outcomes was closely related to the short exposure data.

These and other results led Mitchison and McKee to propose that the initial stage in the matching process was to match unambiguous features, in this case the edge dots. Then, on the basis of these matches, an "interpolation plane" would be drawn through the figure at the depth of the edge dots. For short exposures, this initial approximation would be taken to indicate the depth of all the internal dots. But for longer exposures, this first depth estimate would influence which discrete match was chosen, the rule being to choose the match whose disparity was most similar to the disparity of the interpolation plane. They called this a "nearest disparity rule".

Mitchison and McKee (1987a) also investigated matching for a situation in which the edge dots had unequal disparities. In this case they found that the perceived depth of the internal dots (at long and short exposure durations) fits the predictions of a *slanted* interpolation plane (joining the edge dots) very well (e.g. Mitchison and McKee, 1987a, figure 7 and 9). The matching they observed could not be explained by a fronto-parallel interpolation plane or any nearest neighbour equivalent.

These results are difficult to explain in terms of a co-operative algorithm. This is because, for all the stimuli, there is such strong evidence of a match in the fixation plane. Mitchison (1988, Appendix A) examines in detail whether a co-operative algorithm could account for their results and concludes that one could not. What emerges from Mitchison and McKee's (1987a) results is the crucial importance of the edge dots in determining the pattern of matching.

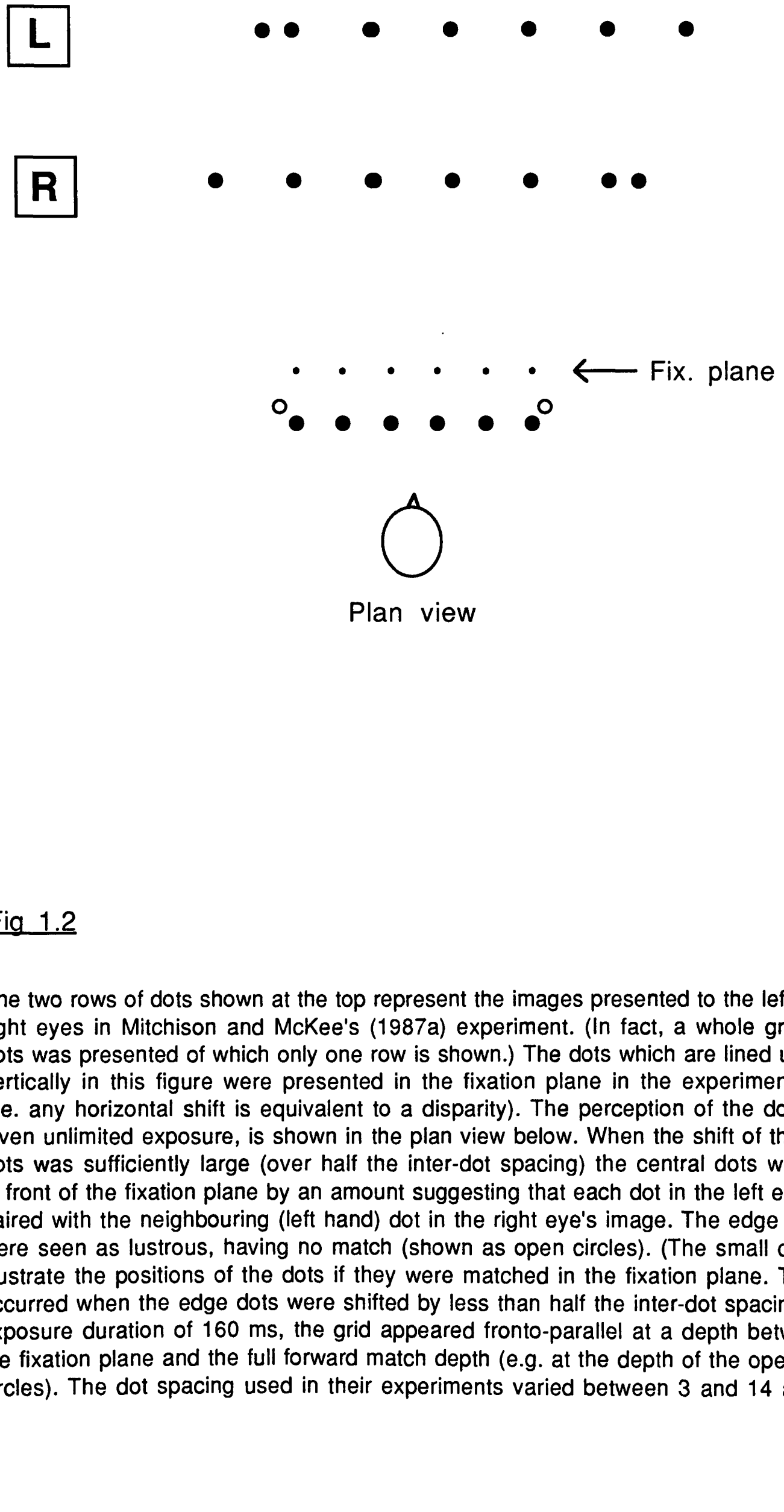


Fig 1.2

The two rows of dots shown at the top represent the images presented to the left and right eyes in Mitchison and McKee's (1987a) experiment. (In fact, a whole grid of dots was presented of which only one row is shown.) The dots which are lined up vertically in this figure were presented in the fixation plane in the experiment (i.e. any horizontal shift is equivalent to a disparity). The perception of the dots, given unlimited exposure, is shown in the plan view below. When the shift of the edge dots was sufficiently large (over half the inter-dot spacing) the central dots were seen in front of the fixation plane by an amount suggesting that each dot in the left eye was paired with the neighbouring (left hand) dot in the right eye's image. The edge dots were seen as lustrous, having no match (shown as open circles). (The small circles illustrate the positions of the dots if they were matched in the fixation plane. This occurred when the edge dots were shifted by less than half the inter-dot spacing.) For an exposure duration of 160 ms, the grid appeared fronto-parallel at a depth between the fixation plane and the full forward match depth (e.g. at the depth of the open circles). The dot spacing used in their experiments varied between 3 and 14 arcmin.

At first sight, this seems to favour a filtering model such as that proposed by Marr and Poggio (1979, discussed in section 1.4). But Mitchison and McKee (1987b) present other data which is very difficult to explain in terms of Marr and Poggio's theory, as indeed are the experiments in Mitchison and McKee (1987a) on matching in slanted surfaces. In summary, Mitchison and McKee's work poses an important challenge to all current theories of correspondence. A more detailed discussion of their arguments and of the model they propose is given in chapter 7.

1.3.4 Scale as a constraint.

If the image can be reduced to only a small number of widely separated features the correspondence problem is made considerably simpler. Mitchison and McKee use scale as a constraint by segmenting the image first (according to unambiguous matches) before applying their planar matching rule within each segment. The Pollard et al. (1985, "PMF") algorithm, at least in its current form (see Frisby and Pollard, 1991), uses a sparse set of seed points as its starting point which is a type of coarse scale analysis. But the main algorithm that uses scale to constrain the correspondence problem from a coarse scale down to the finest scale, is the one proposed by Marr and Poggio (1979).

1.4 Marr and Poggio

Marr and Poggio (1979) put forward an algorithm that tackled the correspondence problem by avoiding rather than solving it. Their theory is computationally elegant and, at least at the time of publication, seemed biologically plausible.

The crucial idea in the algorithm is that the two eyes' images are first compared at a coarse scale, in which case each image contains relatively few, widely separated features. This means that false matches are rarely made. The solution found at a coarse scale can then be used to guide the search for finer scale matches. The process can be repeated at progressively finer scales. If matches are sought within a region appropriate to the scale of analysis, the pairing of features in the two images should almost always be successful.

1.4.1 Filters

The theory was biologically plausible because it was based on filters or spatial frequency tuned channels. There is considerable evidence, both from

psychophysical experiments and physiological recording studies of the mammalian retina and visual cortex, that an important part of early visual processing can be modelled by assuming that the retinal image is processed by mechanisms sensitive to different bands of spatial frequencies (evidence reviewed by DeValois and DeValois, 1988).

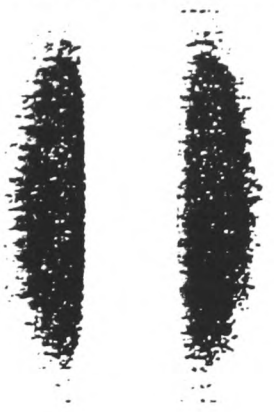
The substrate of these filters, or spatial frequency tuned channels, is assumed to be neurons with a centre-surround receptive field organisation. For instance, Enroth-Cugell and Robson (1966) showed that the responses of retinal ganglion cells in the cat displayed band-pass characteristics and modelled their results by assuming that the receptive field profile of the cells was made up of an excitatory and an inhibitory Gaussian centred at the same point. Marr (1982, figure 2.17) illustrates how closely the responses recorded from on-centre and off centre cells in the retina and lateral geniculate of the cat correspond to the positive and negative components of a Laplacian of Gaussian filter for edge and line stimuli (using data from Dreher and Sanderson (1973) and Rodieck and Stone (1965)). However, in the rest of this thesis reference will be made only to "mechanisms" or filters such as the Laplacian of Gaussian rather than to the properties of neurons in the mammalian visual system (except in the specific discussion of physiological papers in section 2.2.2 and 7.2.1). The aim is to discuss possible descriptions of the visual system at an algorithmic level rather than the possible implementation of the algorithm (Marr, 1982).

Details of the filters Marr and Poggio assumed for the purpose of their algorithm are given in Wilson and Giese (1977), who derived their model from contrast masking studies. An example of one of the filters is shown in figure 1.3. The ratio of excitatory to inhibitory field widths (each assumed to be a Gaussian distribution) is 1:1.5. The receptive field is elongated, being multiplied by a Gaussian in the orthogonal direction.

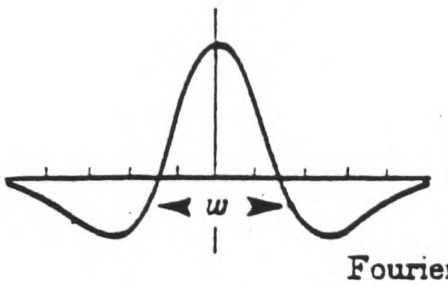
Fig 1.3 (overleaf)
At the top are shown (on the left and right respectively) the one and two dimensional profiles of the receptive fields derived by Wilson and Giese (1977). The excitatory width of the filter, *w*, is illustrated. Below these, the fourier transform of the same filter is shown, again in one and two dimensions.
The four images at the bottom illustrate the receptive field of a Laplacian of Gaussian filter in the same way.

two dimensional receptive field

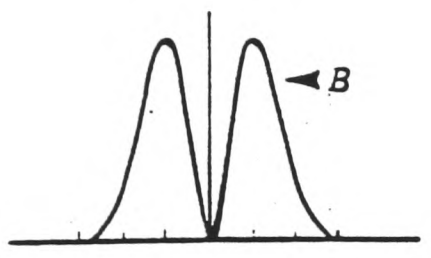
vertical orientation

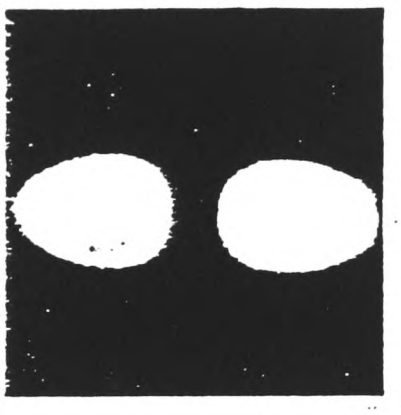


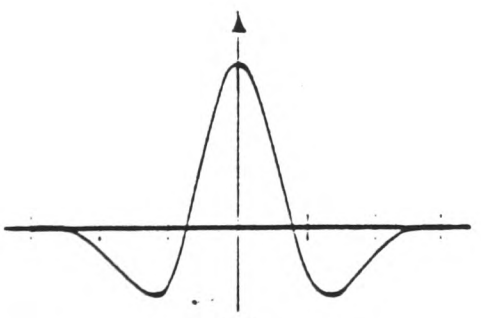
one dimensional profile

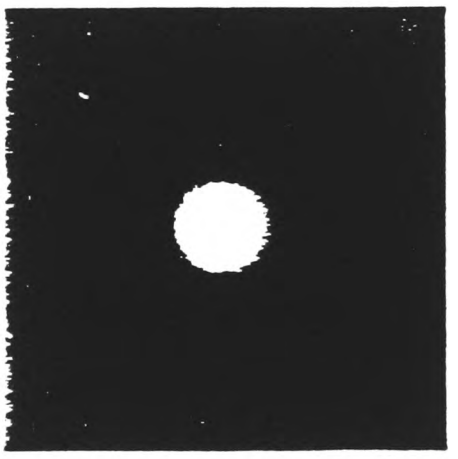


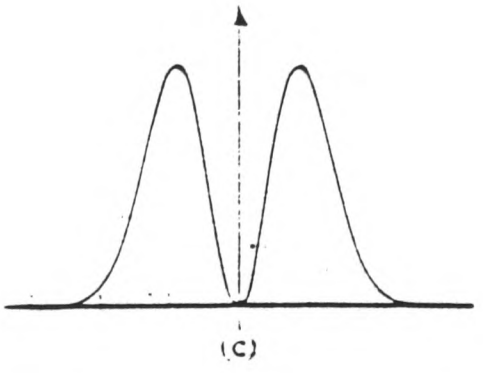
Fourier transforms











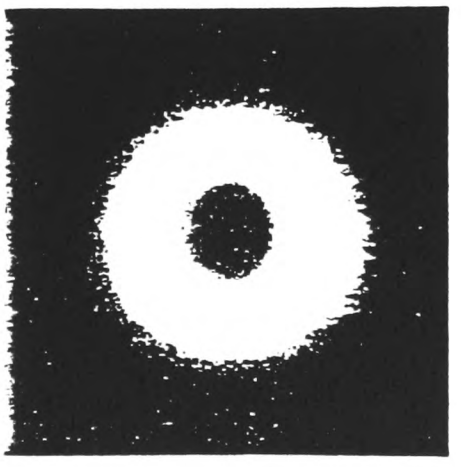


Fig 1.3 (legend on previous page)

The principles of Marr and Poggio's (1979) algorithm are not crucially dependent on the precise form of the filters. It is important that they are approximately balanced (i.e. that they do not respond to a uniform field of illumination) and therefore carry information mainly about *changes* in luminance across the image. Mechanisms with elongated receptive fields respond preferentially to changes in luminance orthogonal to the long axis of the filter. Mechanisms with circularly symmetric receptive fields, such as the Laplacian of Gaussian (LoG) filter shown in figure 1.3 (bottom) respond equally to changes in any direction. Filters of different sizes respond to changes in luminance at different scales. One way to consider the output of these filters is as the outcome of two separate operations carried out on the image. The first is to blur the image at a given scale (i.e. it is convolved with a Gaussian). Filters with large receptive fields blur the image at a coarse scale, small filters at a fine scale. The second operation is to differentiate the image twice with respect to space (in one dimension for an oriented filter, in two dimensions for a circularly symmetric one)*.

The outputs of these different filters can be considered as several independent images derived from the original retinal image. Examples of such images are shown in figure 1.4.

In Fourier terms, these images correspond to different "bands" of spatial frequencies in the original image. That is, each filtered image can be re-described in terms of a series of sine waves of different spatial frequencies, the coarse

---

Fig 1.4 (overleaf)
In the left hand column an image is shown and beneath it the images obtained by filtering it with a Laplacian of Gaussian filter with a space constant of 32, 16, 8, 4 and 2 pixels. (The space constant of a Laplacian is the standard deviation of the Gaussian from which it is derived. Details are given in appendix B. The original image is 256 by 256 pixels) The right hand column shows the result of applying the same set of filters to a 50% density random dot pattern. The images have all been scaled, for the purposes of illustration, to have the same Michelson contrast (or maximum peak to trough amplitude).

---

* The question of whether circularly symmetric or oriented filters best describe the processing which precedes stereo matching is disputed (Mayhew and Frisby,1978; Parker, Johnston, Mansfield and Yang, 1991). However, the spatial primitives, such as zero-crossings, derived from oriented filters perpendicular to epipolar lines (used by Marr and Poggio, 1979) are not dissimilar to primitives derived from a one dimensional analysis of the output of circularly symmetric filters (the approach taken in this thesis).
Strictly, the Laplacian operator is equal to the sum of the second differential of the image in the horizontal and vertical directions. The equation used in this thesis is given in appendix B.

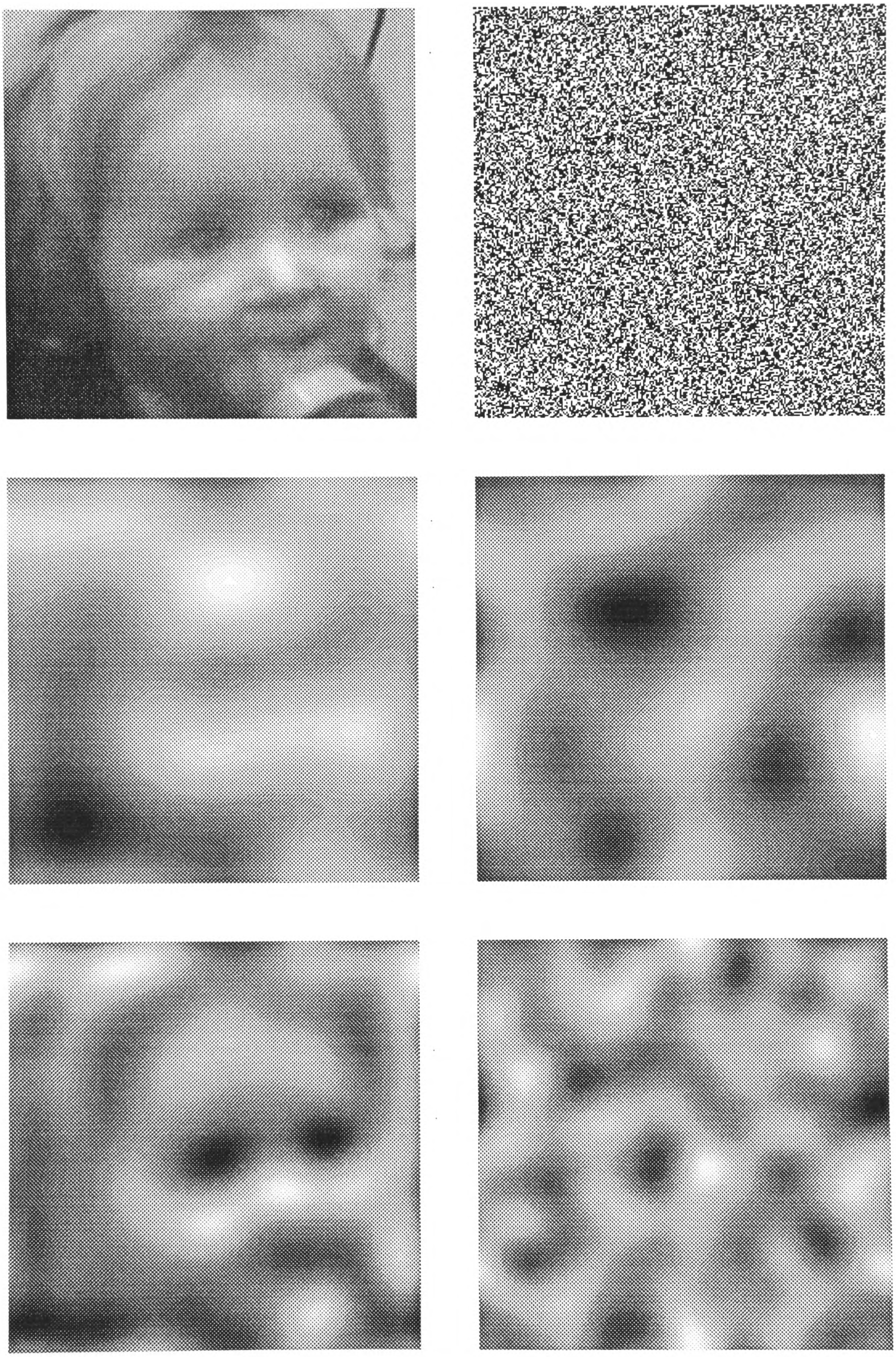


Fig 1.4 (legend on previous page)

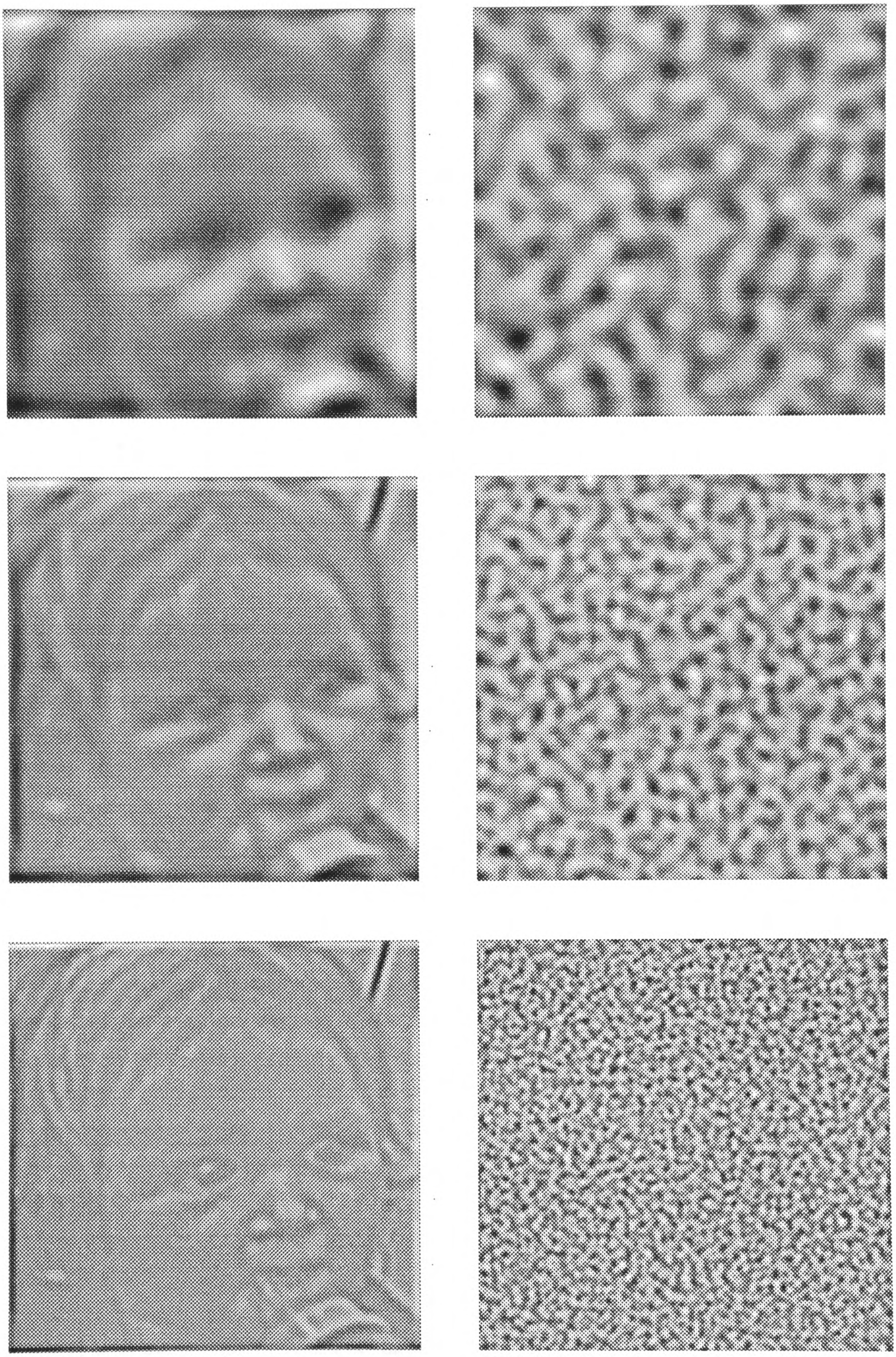


Fig 1.4(cont) (legend on previous page of text)

patterns in terms of low and the fine in terms of high spatial frequency sine waves. The filters let through, or pass, only certain spatial frequencies.

There are several important features of filtered images. First, their peak-to-trough amplitude is much smaller than the amplitude of the original luminance signal . The average signal is zero and the fluctuations around this value are much smaller than the variation of luminance that occur within and across images. This is because they carry a differentiated signal, they record *changes* in luminance. Watt, 1991, p71, illustrates a reduction in the range of values in a grey level image by two log units when the image is differentiated twice. (Temporal differentiation, through light adaptation, plays an even greater role in reducing the range of signal required to record luminance changes.) Second, the spatial characteristics of filtered images are quite regular. The average spacing of features in these images (such as the peaks, troughs or zero-crossings (i.e. the points at which the signal crosses zero)) is proportional to filter size.

The distribution of zero-crossings in a 1-D filtered noise pattern obeys regular statistics which were analysed by Longuet-Higgins (1962). Figure 1.5 shows the distribution of zero-crossing separations for two types of band-pass filter, including the Wilson and Giese (1977) filter. Marr and Poggio used this analysis to show that two similar zero-crossings (i.e. of the same sign, up-going or down-going), would, 95% of the time, be separated by more than a particular distance. This distance scales in proportion to filter size (in fact it is roughly equal to the excitatory width of the filter, w).

1.4.2 Avoiding false targets

Marr and Poggio supposed that matching occurred after the two eyes' images had been filtered and the positions of the zero-crossings had been determined. They proposed that this was carried out first at a coarse scale, in which case the zero-crossings would be widely spaced. The task would be to match each zero-crossing with the corresponding zero-crossing in the other eye's image. If the surface is in the fixation plane then the zero-crossings will be in corresponding positions, i.e. they will have the same retinal co-ordinate in both eyes. If there is no matching zero-crossing at this location then a match is sought at locations to the left or right of this point (i.e. corresponding to locations on the epipolar plane). The analysis described above means that, if the disparity of the surface is less than $\pm\ ^1/_2 w$, where w is the excitatory width of the filter, then any zero-crossing encountered within these bounds will be a correct match. (For instance, if the disparity is $+^1/_2$

w then, 95% of the time, the nearest false target zero-crossing will be at $-1/2\ w$ or further. For disparities less than $\pm\ 1/2\ w$ the probability that a false target is nearer than the correct one is even smaller.)

By adding a co-operative element to the scheme, Marr and Poggio suggested that the disparity range for successful matching could be increased to $\pm\ w$. At disparities up to this range, for 50% of zero-crossings, there will be no false targets (i.e. relaxing the criterion of 95% correct matches). The evidence for this claim comes again from the distribution of zero-crossing separations shown in figure 1.5 ($\int P_1$ is 0.5 at $2w$). For the other 50% of cases (where zero-crossing separation is less than $2w$), there will be two alternative matches and at least 95% of the time, one will be crossed and one uncrossed. Marr and Poggio suggest that these ambiguous matches could easily be biased by the unambiguous matches (i.e. the other 50% of matches) which would give the correct disparity. But beyond a range of $\pm\ w$ false matches become more common than correct ones (the $\int P_1$ curve is already very steep at $2w$) and no simple co-operative scheme could account for successful matching much beyond this range.

Marr and Poggio suggested that once the correspondence problem has been solved at a coarse scale, vergence movements could then be made to bring finer scale features into closer correspondence. This is illustrated in figure 1.6. The assumption this makes is that fine scale features will have a similar disparity to coarse scale ones. (If both arise from the same surface, it is a reasonable assumption. Failure of this assumption is discussed in section 6.1.2.) After a vergence eye movement has been made the finer scale zero-crossings on that surface should, in most circumstances, come within range and matching be carried out in the same way as before. The same cycle can be repeated until corresponding zero-crossings are matched down to the finest scale.

An important limitation of the model as described so far is that the coarse-to-fine guidance can only be carried out for one depth plane at a time (a vergence movement may take other features further away from correspondence). Marr and Poggio's solution to this was to propose a store or "dynamic buffer" to record the depths (or vergence movements or disparities) of features that had been successful matched. This they called the $2\ 1/2$-D sketch.

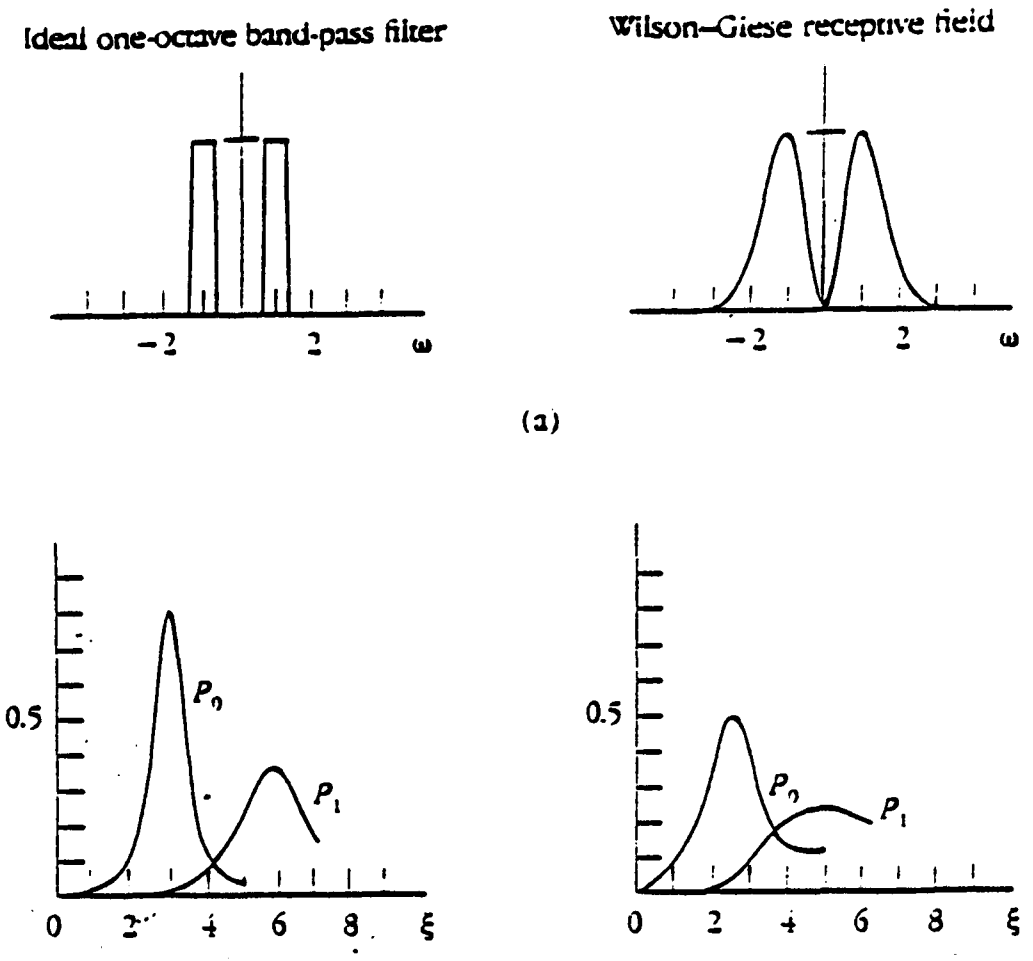


Fig 1.5

An illustration of the distribution of zero-crossing spacings in a random function after filtering with a band-pass filter. Analysis for two types of filter is shown, on the left for an ideal, one octave filter, on the right for a difference of gaussian filter as proposed by Wilson and Giese (1977). The fourier spectra of the filters are shown at the top (a). In the centre (b), is shown the expected distribution of intervals between adjacent zero-crossings (P_0) and between adjacent zero-crossings of the same sign, up-going or down-going, (P_1). At the bottom (c), the integral of P_1 is shown. This gives the frequency of zero-crossing intervals less than or equal to a particular interval (ε).

(from Marr and Poggio, 1979)

1.4.3 The 2 $^1/_2$-D sketch

The 2 $^1/_2$-D sketch was proposed by Marr and Nishihara (1978) and forms an important part of Marr and Poggio's (1979) model. It is a viewer centred representation of the surfaces in the image. It can be thought of as a 2-D image in which each point is labelled with the distance (*r*) of that point from the observer and the orientation of the surface normal at that point (*s*). This information could come from motion parallax, shading and texture information as well as from binocular stereopsis. That is, it would be a "gathering station" for information from many different cues before an object centred (i.e. fully 3-D) representation was formed.

For the purposes of Marr and Poggio's (1979) stereo algorithm, the role of the 2 $^1/_2$-D sketch was to act as a store of correspondence matches once they had been made. Accurate vergence movements over large disparities could then be driven from information in the 2 $^1/_2$-D sketch.

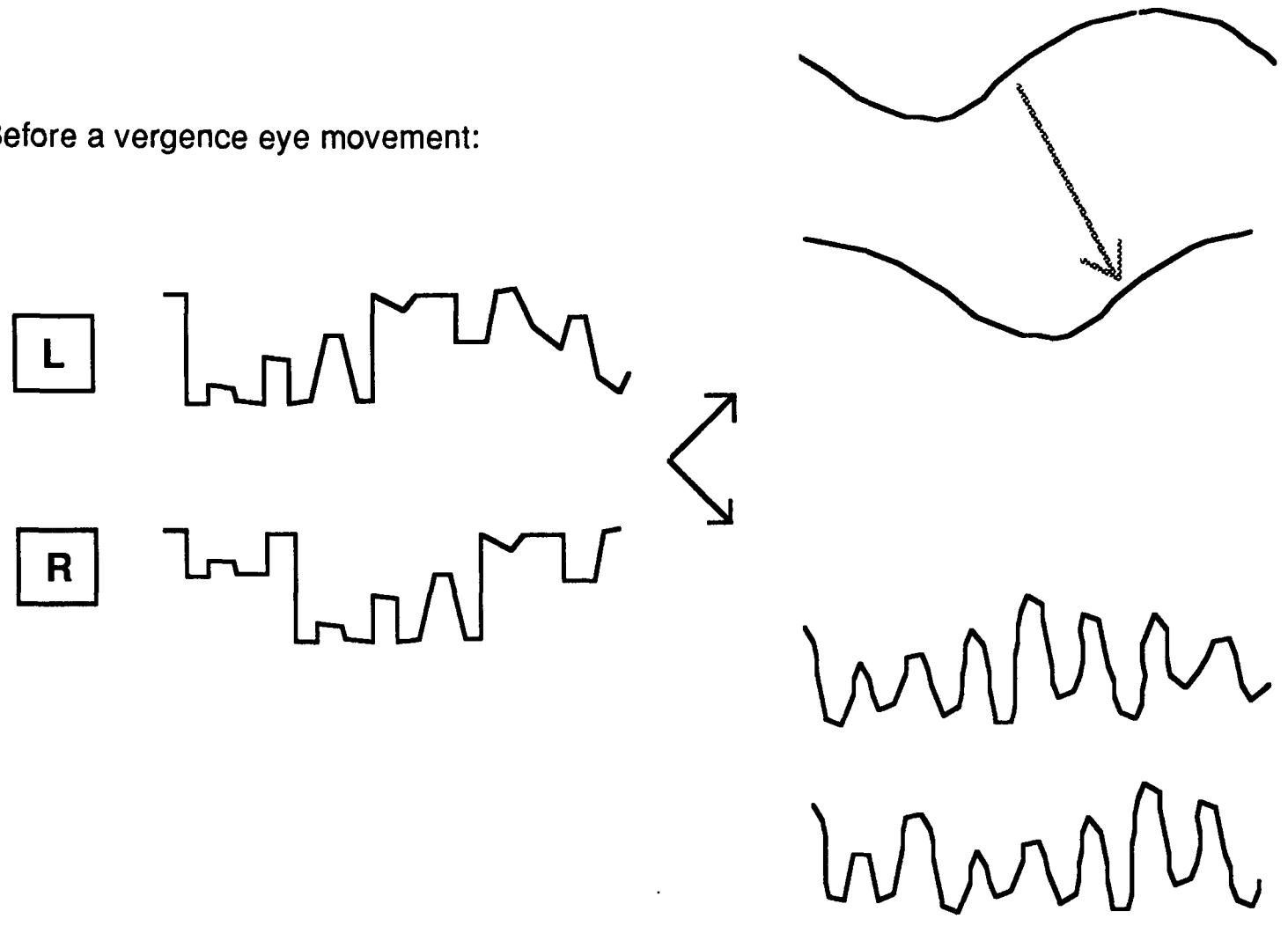
There are scant details about the implementation of this idea. The co-ordinate framework of the 2 $^1/_2$-D sketch is likely, Marr says (1982, p285), to be initially retinocentric and then converted to a headcentric one, perhaps centred at the cyclopean point. Grimson (1981) discusses the problem of a co-ordinate framework and admits that the solution he chose in implementing Marr and Poggio's algorithm may not be available to the human visual system:

Fig 1.6 (overleaf)

A schematic illustration of the correspondence problem and how it can be solved most easily at a coarse scale. At the top, a luminance profile in the left and right eyes' images is shown. The two profiles are drawn above one another with corresponding points on the two retinae vertically aligned. Features in the right eye's pattern are all shifted in one direction, i.e. there is a disparity between the two images. The output from a coarse filter (above right) and a fine filter (below) for each eye's image is shown. At a fine scale, the spacing between peaks or zero-crossings is small compared to the disparity so false matches would be made. At a coarse scale, the correct correspondence is easily found (arrow).

By moving the eyes so that the coarse scale features line up, the fine scale features are brought into much closer alignment and the correct matches can be made.

Before a vergence eye movement:



After a vergence eye movement:

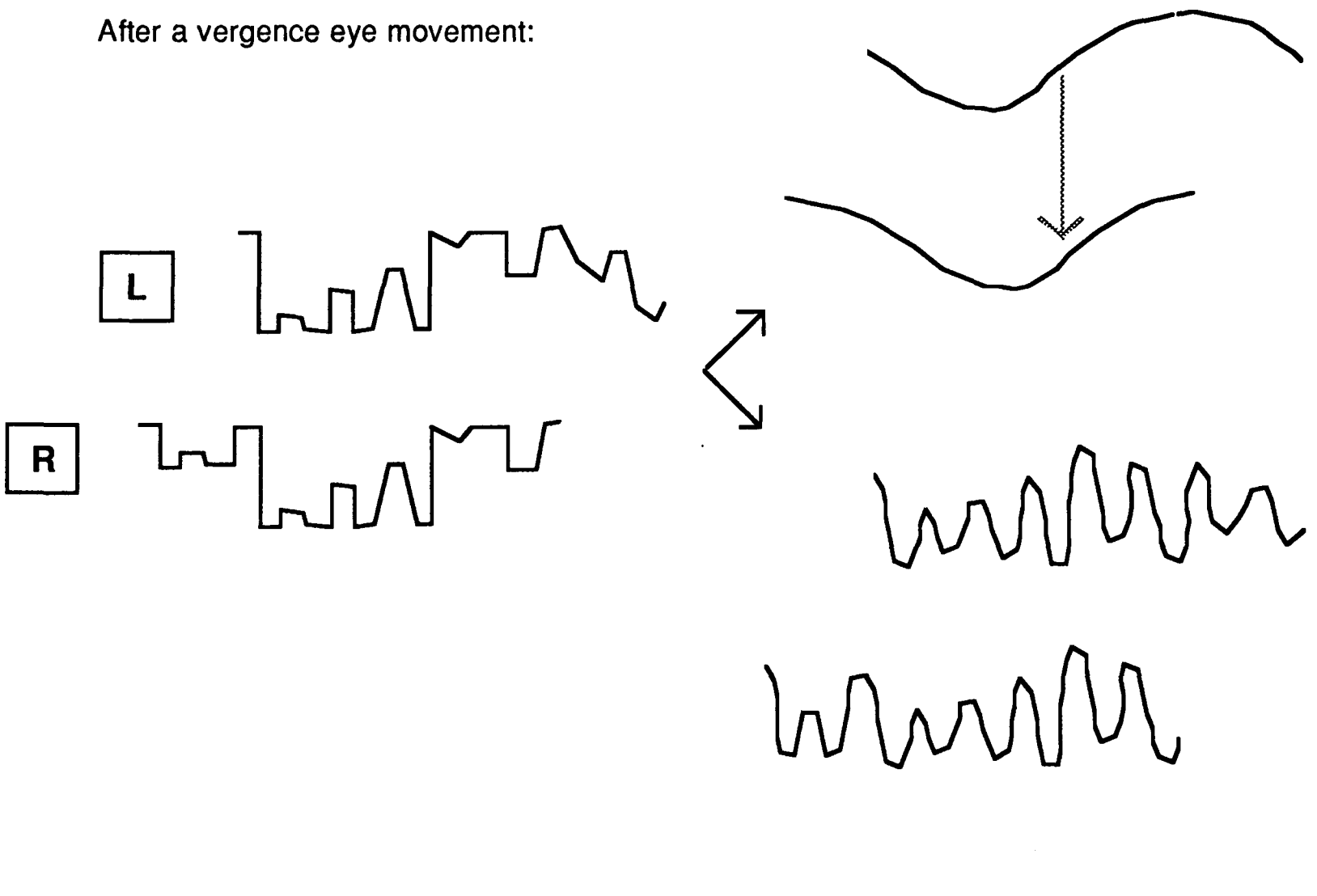


Fig 1.6 (legend on previous page)

> *"There are a number of critical questions concerning the form of the 2 $^1/_2$-D sketch, which have yet to be firmly answered...[One] critical question concerns whether the sketch uses the co-ordinates of the scene or the working arrays* [retinal co-ordinates, which change with eye movements]. *In the first case, the co-ordinates of the sketch would be directly related to the co-ordinates of the entire scene. The advantage of this is that, since disparity information about the scene is extracted from several eye positions, the representation is readily updated. However, this advantage also raises a difficulty. In order to store information into a buffer whose co-ordinate system is determined by the image of the entire scene, explicit information about the positions of the eyes is required. This is fine computationally, but in the human visual system this information may not be available to the stereo process."*
>
> (Grimson, 1981, p83-84)

There is a similar problem with the 2 $^1/_2$-D sketch which arises from the fact that in the human visual system the representation of the absolute distance of each point from the observer is likely to be crude. Marr admits this (Marr, 1982, p283) and concludes that the most precise information in the 2 $^1/_2$-D sketch is probably surface orientation. However, he does not address the problem that the disparity information specifying surface slant must be scaled according to the absolute distance of the surface (in fact, Marr gives an example illustrating that such scaling probably is carried out in the visual system, (Marr, 1982 p158)).

The main role of the 2 $^1/_2$-D sketch in Marr and Poggio's model is to direct eye movements. The most relevant information to record in the 2 $^1/_2$-D sketch for this purpose is the disparity of each point (with respect to some arbitrary origin).rather than its absolute distance from the observer.

Apart from their description of the 2 $^1/_2$-D sketch, Marr and Poggio's model was quite detailed and, since it was based on the filters derived by Wilson and Giese (1977), they were able to make quantitative predictions about psychophysical performance. These are discussed in the next section.

1.5 Psychophysical evidence

Marr and Poggio (1979) made a series of bold and precise predictions at the end of their paper. The predictions (P) were rated, with stars, according to the importance of the result for their model. Three stars meant that a prediction, if proved to be incorrect, would falsify their theory. Results that were already known at the time

(A) were also discussed in the context of their theory. In the following sections, Marr and Poggio's predictions are examined in the light of subsequent psychophysical findings.

1.5.1 "Channels" for stereopsis?

> *"1. (A,P***) Independent spatial-frequency-tuned channels are known to exist in binocular fusion and rivalry. The theory identifies these with the channels described from monocular experiments (Julesz and Miller, 1975; Mayhew and Frisby 1976;..Wilson and Giese 1977;..Felton, Richards and Smith 1972)."*
>
> (Marr and Poggio, 1979, p321)

It is now widely accepted that the retinal image is filtered at a range of spatial scales and that this is one of the earliest stages of visual processing. (The evidence for this is reviewed by DeValois and DeValois, 1988.) However, there is little agreement on how the outputs of these filters are used for stereo matching and further processing. For example, there is a debate as to whether the output of oriented or circularly symmetric filters are the starting point for stereopsis (Parker, Johnston, Mansfield and Yang, 1991; Mayhew and Frisby, 1978). Another debate concerns the spatial frequency tuning of filters involved in stereopsis (Julesz and Miller, 1975; Yang and Blake, 1991; Tyler and Barghout, 1992). All these studies depend on a masking paradigm, described originally on the stereo domain by Julesz and Miller.

Julesz and Miller (1975) demonstrated that a filtered random dot pattern (either high or low pass) could support stereopsis (a central square was visible) when masking noise of a very different spatial frequency was added to one eye's image, but that stereopsis was destroyed when mask and signal overlapped in spatial frequency. They concluded that the image must be analysed by spatial frequency tuned channels and that these must "reside before the stage of global stereopsis" p143.

Yang and Blake (1991) repeated Julesz and Miller's experiment in a more quantitative way, measuring the signal-to-noise ratio required for reliable detection of a disparate target (a rectangle either in the top or bottom half of the display). They used narrow band (0.4 octave), circularly symmetric filtered random dot patterns both for the signal and noise. For each signal frequency they measured the effect of masks with a range of centre frequencies. They found that the most effective mask was not always of the same frequency as the signal. Instead the

"peak" masking spatial frequency tended to be closer to the medium spatial frequencies (3-5 c/deg) than the signal, i.e. for low spatial frequency stimuli the most effective mask was of a slightly higher spatial frequency than the signal and vice versa at high spatial frequencies. The height of the curves (i.e. mean signal-to-noise ratio for each signal frequency) varied systematically with signal frequency. The lowest signal-to-noise threshold ratios were obtained for the medium (3-5 c/deg.) frequencies. They interpreted their results as evidence for a narrow range of spatial frequency tuned channels subserving stereopsis and plotted a "spatial sensitivity function for stereopsis" (their figure 13). Within this "envelope" they postulated two spatial frequency tuned channels for stereopsis, while Tyler and Barghout (1992), on the basis of the same data, suggest rather more.

An alternative interpretation of these results can be given. Adding noise at the same spatial frequency as the signal has the effect of completely disrupting the spatial structure of the image - it looks entirely different. This is referred to as "pattern masking". It would affect any stereo matching process and does not, by itself, imply independent stereo "channels". This fact may be illustrated by considering the cross correlation of two band-pass images. Noise added far away from the signal frequencies will not affect the cross correlation. Noise overlapping in spatial frequency with the signal will have a large disruptive effect on the correlation. Thus Julesz and Miller's results are not surprising.

Yang and Blake's (1991) result, in particular the fact that the peak masking effect occurs away from the signal spatial frequency, requires a further explanation. What they have not considered is the effect of different contrast sensitivities at different spatial frequencies on their signal-to-noise ratios. If the signal is of high or low spatial frequency, away from the peak of the (luminance) contrast sensitivity function (c.s.f.) while the mask spatial frequency is nearer the peak (3-5 c/deg) then the *effective* signal-to-noise ratio will be much lower than it appears to be. Without taking this into account the peaks of the masking functions would all be expected to be shifted towards 3-5 c/deg, as Yang and Blake found.

Another aspect of their results, the change in height of the tuning curves for different signal frequencies, lends support to the above interpretation. Yang and Blake deduce from this rise and fall a "spatial sensitivity function for stereopsis" (their figure 13) which turns out to be almost identical to the contrast sensitivity function (Campbell and Robson, 1968). However, this may be more than a fortuitous coincidence: when the signal is at either extreme of the c.s.f., it might be

expected that less noise would be required to mask it than when the signal spatial frequency is at the peak of the c.s.f.

So pattern masking, rather than a "stereo channel" model, may be adequate to explain these results, but only after the effect of the c.s.f. is taken into account. In other words, it is necessary to assume that filtering of the image at a range of spatial scales takes place but the results do not rule out a model in which the output of these filters are combined before stereo matching occurs.

1.5.2 The primitives used for matching

> *"2. (P***) Terminations and signed, roughly oriented zero-crossings in the filtered image are used as input to the matching process."*
>
> (Marr and Poggio, 1979, p321)

In most of the rest of their paper Marr and Poggio (1979) only consider zero-crossings as the primitives used in the matching process. Their reference to terminations as primitives was prompted by demonstrations by Julesz (Julesz and Spivack, 1967; Julesz, 1971) that stereograms made up of lines containing minute breaks can be fused successfully. The breaks in the lines can be as little as 16 arc sec (monocularly), i.e. in the hyperacuity range and stereopsis still achieved. Julesz and Spivack argued that the vernier breaks must be detected before stereoscopic matching occurred. Nishihara and Poggio (1982) have questioned this argument. They showed that the output of a relatively coarse DOG filter (excitatory width of 4 arcmin, i.e. coarse compared to the line breaks) can be used as the input to a matching algorithm that can successfully solve the correspondence problem and determine the depth of the central square. This filter is well within the range of sizes proposed by Marr and Poggio (1979) and Wilson and Bergen (1979). Nishihara and Poggio pointed out that the variation in line spacing in Julesz and Spivack's patterns was relatively large (the ratio of the smallest to the largest spacing was 2:3) and showed that it was this variation that provided the coarse scale cue. When the average line spacing was increased but the size of the vernier cue kept constant, Nishihara and Poggio found that the computer algorithm failed (using the same sized DOG filter) and that subjects could not fuse the stereograms as easily (despite the fact that larger spacings should improve vernier acuity (Westheimer and McKee, 1977))

Experiments that have addressed the nature of the primitives used in spatial vision (e.g. Watt and Morgan, 1984) and for stereopsis (e.g. Legge and Gu, 1989) do

not support zero-crossings as a candidate. Both these experiments measured location accuracy (either for a spatial or a stereo vernier task) for low contrast stimuli. These experiments are discussed in greater detail in section 2.3, but, in summary, both groups found that the variability of zero-crossing locations changed much too rapidly as signal-to-noise ratios were reduced to account for the psychophysical results. The authors differed in their choice of primitive: Legge and Gu chose peaks, Watt and Morgan centroids, and there is little to distinguish these two. (In fact, after a more recent publication, Gu and Legge (1991), both groups agree on centroids as the most appropriate primitive.)

1.5.3 Disparity discrimination at large pedestal disparities

> *"3. (P**) In the absence of eye movements, discrimination between two disparities in a random dot stereogram is only possible within a range of* +/- w *of the largest active channel."*
>
> (Marr and Poggio, 1979, p321)

It is not clear, if Marr and Poggio refer to *unfiltered* random dot patterns, how the largest active filter would be determined, unless the values of filter sizes are taken from studies such as Wilson and Giese (1977). Mayhew and Frisby (1979) have addressed this prediction using random dot stereograms, both filtered and unfiltered. The task in their experiment was to discriminate which of two patches had the greater disparity (always convergent). The disparity difference was always $2\ ^1/_2$ arcmin with different pedestal disparities. For the unfiltered patterns, subjects could perform this task reliably at exposure durations too short for eye movements to be initiated (less than 160 ms) for pedestal disparities up to 16 arcmin. This is within the range predicted by Marr and Poggio (see section 1.5.6). However, Mayhew and Frisby's results for filtered patterns do not tally with Marr and Poggio's predictions. Depth discriminations were reliably made for disparities sometimes as much as four times, and invariably more than twice, the limit Marr and Poggio predicted. The results are qualitatively in agreement with Marr and Poggio's theory, but then almost any scheme that was affected by the number of false targets would show this trend, and Marr and Poggio had put great emphasis on the quantitative nature of their predictions. The main problem with this experiment is that all the stimuli were convergent so that, as the authors admit, anticipatory eye movements could be made, considerably weakening any quantitative claims made in the paper.

Accurate control of eye movements is crucial for the results of this sort of experiment to be interpreted correctly. This point is illustrated by another experiment carried out by Frisby and Mayhew (1978). They measured contrast sensitivity for the perception of depth in filtered random dot patterns as a function of the spatial frequency and disparity of the stimulus. (Subjects gradually increased the contrast of the stimulus until they could confidently determine whether it lay in front or behind the fixation plane.) They found no effect of disparity. When Smallman and MacLeod (1992) repeated their experiment with proper control of eye movements the results were entirely different. Smallman and MacLeod's results are shown in figure 1.7 along side those of Frisby and Mayhew.

Smallman and MacLeod come to the opposite conclusion, i.e. that spatial frequency and disparity are very closely linked, just as Marr and Poggio suggested* .

1.5.4 Stereoacuity

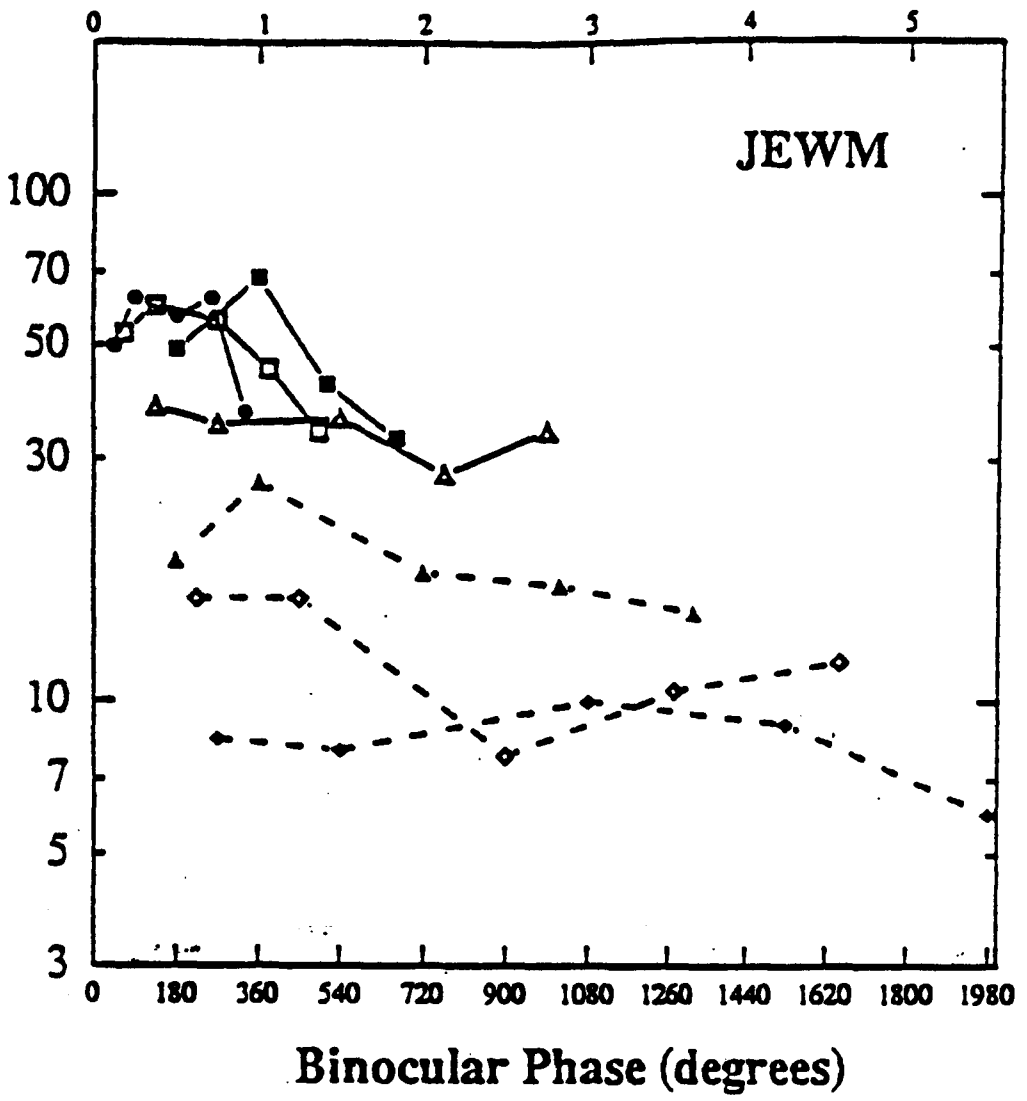
> *"3. (cont.) Stereo acuity should scale with the width* w *of the smallest active matched channels (i.e. about 10" for the smallest and 40" for the largest foveal channels)."*
> (Marr and Poggio, 1979,p321)

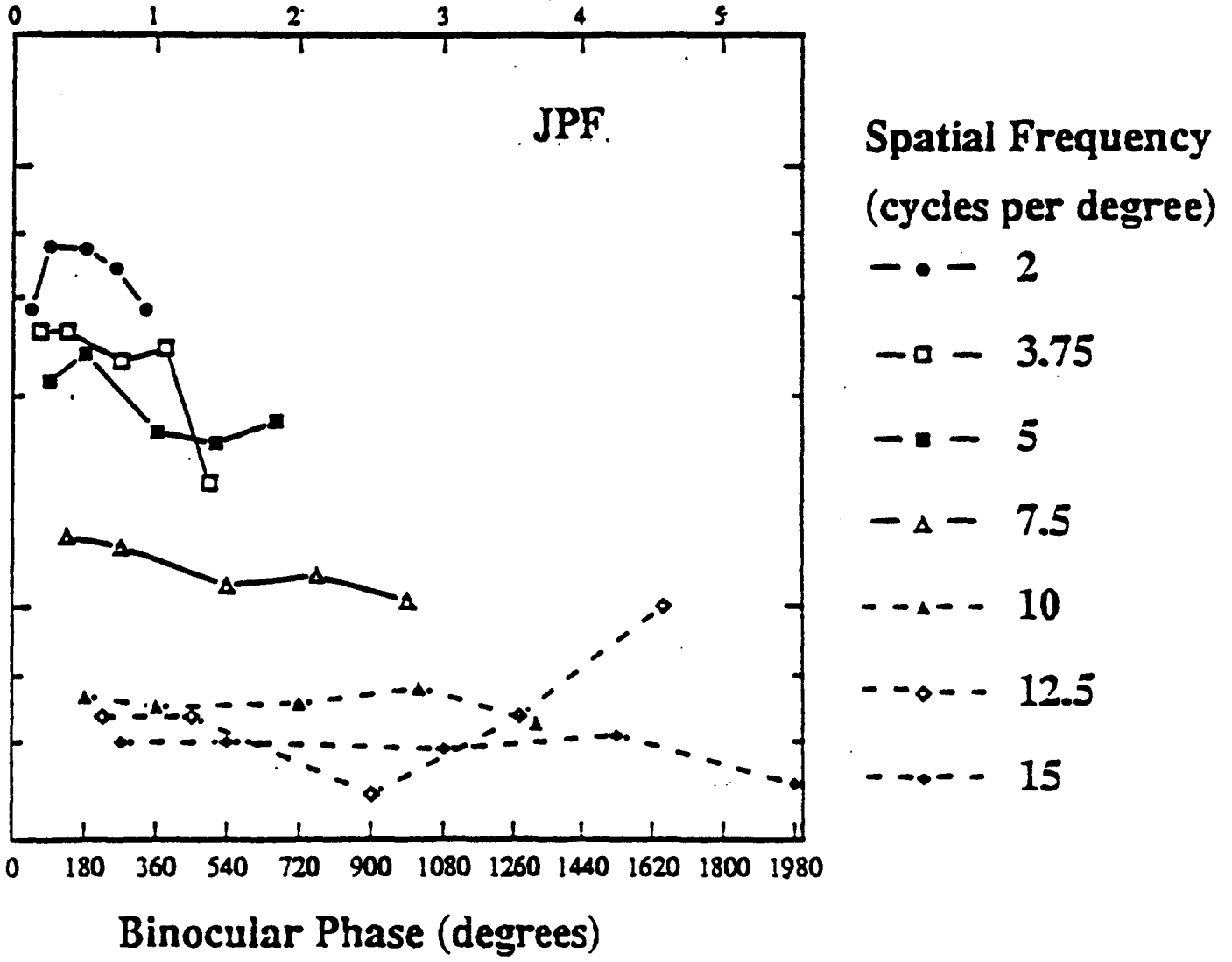
There are two ways, according to Marr and Poggio's theory, in which the smallest active channel can be manipulated. One is to filter out high spatial frequencies from the stimulus. The other is to increase the disparity so that the smallest channels give no coherent signal. Stereoacuity has been measured under both these conditions.

Fig 1.7 (overleaf)
(a) The results from Mayhew and Frisby's (1978) experiment are shown for two subjects. Contrast sensitivity for stereopsis is shown as a function of disparity. (Disparity is plotted as a phase shift of the centre frequency of the stimulus.) There is no clear effect of phase. (b) The results of Smallman and MacLeod's experiment (manuscript in preparation) are plotted on the same axes. The pattern of results is very different. Many of the functions peak near 90° phase disparity.

---

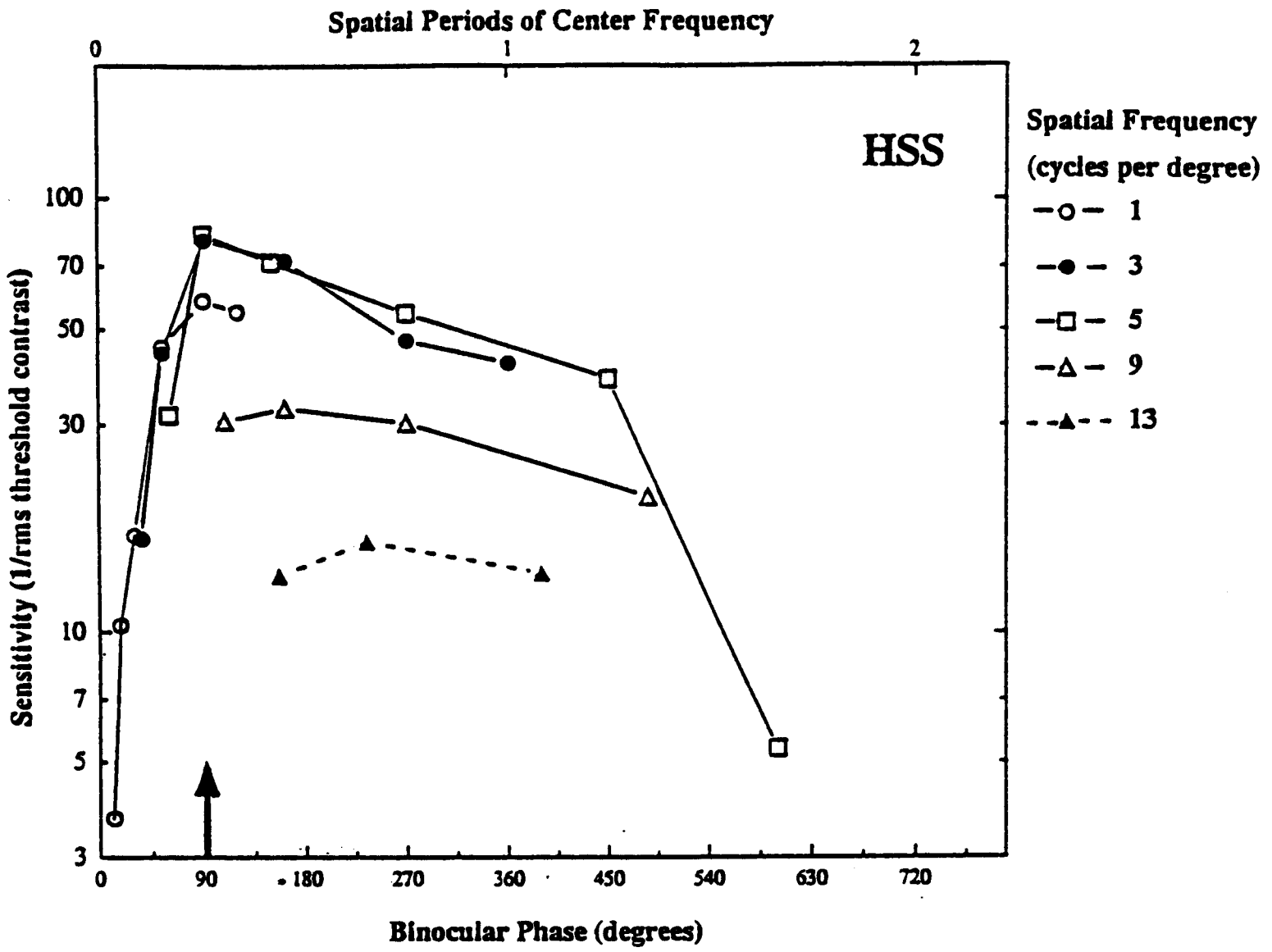
* However, the assumption behind both these experiments, that increasing contrast can overcome the correspondence problem, is not easy to account for within the framework of Marr and Poggio's theory. One possibility is that non-linearities in the filters are exaggerated at high contrast, thus introducing spurious low spatial frequencies.





Spatial Periods of Center Frequency

Fig 1.7a (legend on previous page)



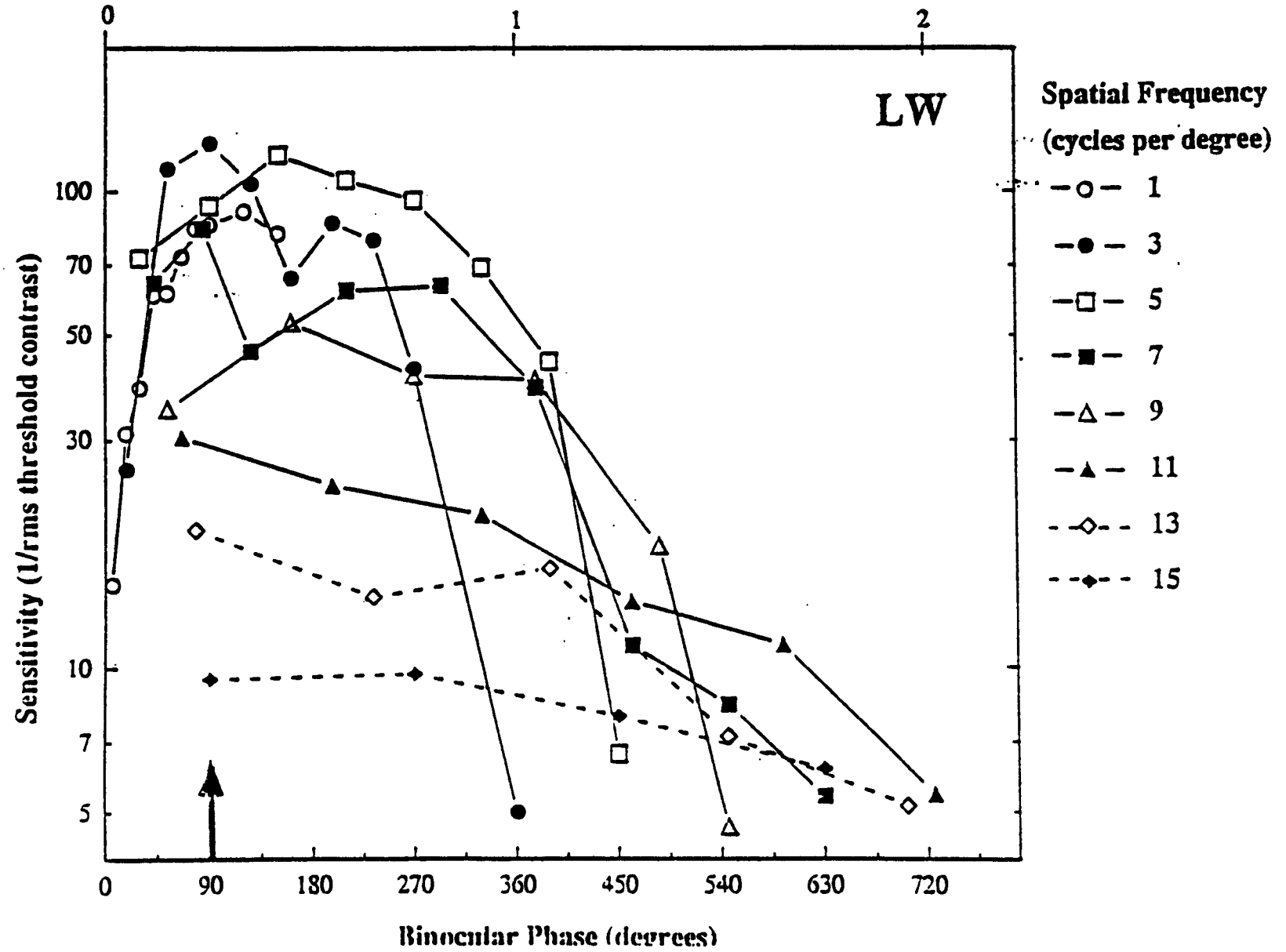


Fig 1.7b (legend on previous page of text)

Several studies (e.g. Ogle, 1953; Blakemore, 1970b) have measured stereoacuity for unfiltered stimuli at a range of pedestal disparities. Marr and Poggio's theory predicts that stereoacuity should scale in proportion to the pedestal disparity, because the size of the smallest filter that can detect the disparity is proportional to the disparity, and the stereoacuity of each channel is proportional to the filter width (an argument explored in more detail in chapter 4). This is not the experimental finding. Both Ogle and Blakemore found an "exponential" increase in stereo thresholds with pedestal disparity (i.e an exponent greater than one). In Blakemore's experiment stimuli were presented too briefly for eye movements to take place (100 ms). In both experiments line stimuli were used with disparities up to 120 arcmin. Badcock and Schor (1985) found similar results for disparities up to 20 arcmin but little or no increase in thresholds for stimuli up to 80 arcmin pedestal disparity. (Thresholds remained between 1 and 2 arcmin).

The reason for the different findings of Badcock and Schor is not clear. A significant difference between the procedures is the exposure duration used (100 ms for Blakemore, 200 ms for Ogle but 750 ms for Badcock and Schor). However, Badcock and Schor repeated their experiment using a 100 ms exposure and claimed to obtain a very similar pattern of results. Badcock and Schor discuss whether diplopia might account for the difference in the results (since for the larger disparities all their subjects perceived the stimuli as diplopic and not easily seen in depth while Blakemore's subjects were reported as "not noticing" any diplopia). One reason to doubt this hypothesis is that Badcock and Schor obtained a similar pattern of results (i.e. a relatively slow rise in thresholds for disparities above 20 arcmin) using wide difference of Gaussian (DOG) stimuli which never gave rise to diplopia. Also, it is difficult to see why the perception of diplopia should improve performance.

It is unusual to find (in any psychophysical comparison task) that increasing the size of the standard (in this case the pedestal disparity) has no effect on thresholds: a Weber or proportional relationship is much more common. Badcock and Schor's results for unfiltered stimuli show a reduction in the Weber fraction for stereo-depth comparison between disparities of 20 and 80 arcmin, reaching about 1% at 80 arcmin, which is very low indeed. A pattern of results like this raises the question of whether some other cue that is constant for different pedestal disparities could account for subjects' performance. For example, unless the position of the standard is jittered between trials there is a possibility that subjects could use a monocular vernier or orientation cue to perform the task. The standard

was set to the right of fixation in Badcock and Schor's experiment "in order to avoid vernier alignment cues" (p1212) but this does not eliminate a constant orientation cue. The task was a two-interval forced choice procedure in which the stimulus had the same disparity as the standard in one interval and a slightly different disparity (pedestal-plus-test) in another. Subjects indicated in which of the two intervals they perceived the relative depth offset and were provided with feedback for incorrect responses. It is possible that under these conditions the task could be performed monocularly and that performance might not vary with pedestal disparity above some lower limit. Blakemore jittered the position of the standard and provided no error feedback. He does discuss the possibility that subjects could use the relative separation of the diplopic images of test (above) and standard (below) as a cue but in a control experiment in which subjects were asked explicitly to use this cue, the pattern of results, in particular the bias in subjects' responses, was very different. This, he argues, makes it unlikely that subjects were using the relative separation cue to judge relative depth.

Schumer and Julesz (1984) used a slightly different paradigm (the detection of sinusoidal depth modulations in a random dot pattern).to test stereoacuity for a range of pedestal disparities up to 50 arcmin. They obtained similar results to those found by Ogle (1953) and Blakemore (1970b), i.e an exponential rise in stereoacuity thresholds with increasing pedestal disparity.

Stereoacuity has also been measured in several studies using filtered stimuli. In these experiments exposure duration and eye movements were not limited. Schor and Wood (1983) used difference of Gaussian (DOG) stimuli and a method of adjustment. They found that, for DOG widths greater than about $^1/_2$ ° (i.e. centre frequencies below 2.4 c/deg.), stereo acuity thresholds were proportional to DOG width, as Marr and Poggio would predict. Legge and Gu (1989) found the same relationship for sine wave stimuli: below 3 c/deg thresholds rose in proportion to the spatial period. Thresholds can be expressed as a constant phase shift of 3-6° for Legge and Gu's data, and, for Schor and Wood, a phase shift of about 3° of the stimulus centre frequency. Although these results are compatible with Marr and Poggio's theory, it may be that they reflect instead an "informational limit" rather than the properties of different mechanisms stimulated by sine waves of different frequencies. This possibility is discussed in chapter 3.

One paper combines both approaches, i.e. manipulating both pedestal disparity and spatial frequency content. Badcock and Schor (1985) measured stereoacuity over a

range of disparities using DOG stimuli of different widths. The main finding, which, as they point out, is contrary to Marr and Poggio's predictions, is that large disparities (e.g. 80 arcmin) can be signalled with high precision (1-2 arcmin) for very narrow DOG stimuli (9.6 c/deg). In fact the precision is much greater than for low spatial frequency DOGs.

In summary, the results show that high spatial frequency information is important for good stereoacuity but they do not support Marr and Poggio's quantitative predictions with respect to the effects of pedestal disparity.

1.5.5 Diplopia and maximum perceived depth

> *"4. (P***) In the absence of eye movements, the magnitude of perceived depth* in non-diplopic conditions *is limited by the lowest spatial frequency channel stimulated."* [emphasis added]
>
> (Marr and Poggio, 1979, p321)

Two separate issues are combined in this prediction. First, how is the maximum perceived depth related to the lowest spatial frequency in the image? Second, how is the maximum disparity without diplopia being perceived related to the lowest spatial frequency in the image? The second issue arises because, for most stimuli, diplopia occurs at smaller disparities than that which gives rise to the greatest perceived depth (e.g. Ogle, 1952; Richards, 1971).

A paper by Schor, Wood and Ogawa (1984) gives a clear answer to the second question. Diplopia is limited by the *highest* spatial frequency in the stimulus. They used unfiltered bars or DOG stimuli with a range of widths. Fixation was maintained using nonius lines and the stimulus disparity was gradually increased until diplopia was observed (i.e. a "slight doubling, an increase in width or a lateral displacement" of the stimulus, p662). Figure 1.8 illustrates their results. Diplopia thresholds varied very little with bar width (unfiltered) and the slight rise in thresholds for large bar widths might be accounted for by blurring of the bar edges as they fell on more eccentric parts of the retina. A bar is a broad band stimulus. Removing low spatial frequencies (results for a narrow DOG) did not change diplopia thresholds, i.e. the curves for the DOG and bar overlap at this point. Removing high spatial frequencies (results for wide DOGs) has, in contrast, a dramatic effect on thresholds which rise in proportion to DOG width (in fact they

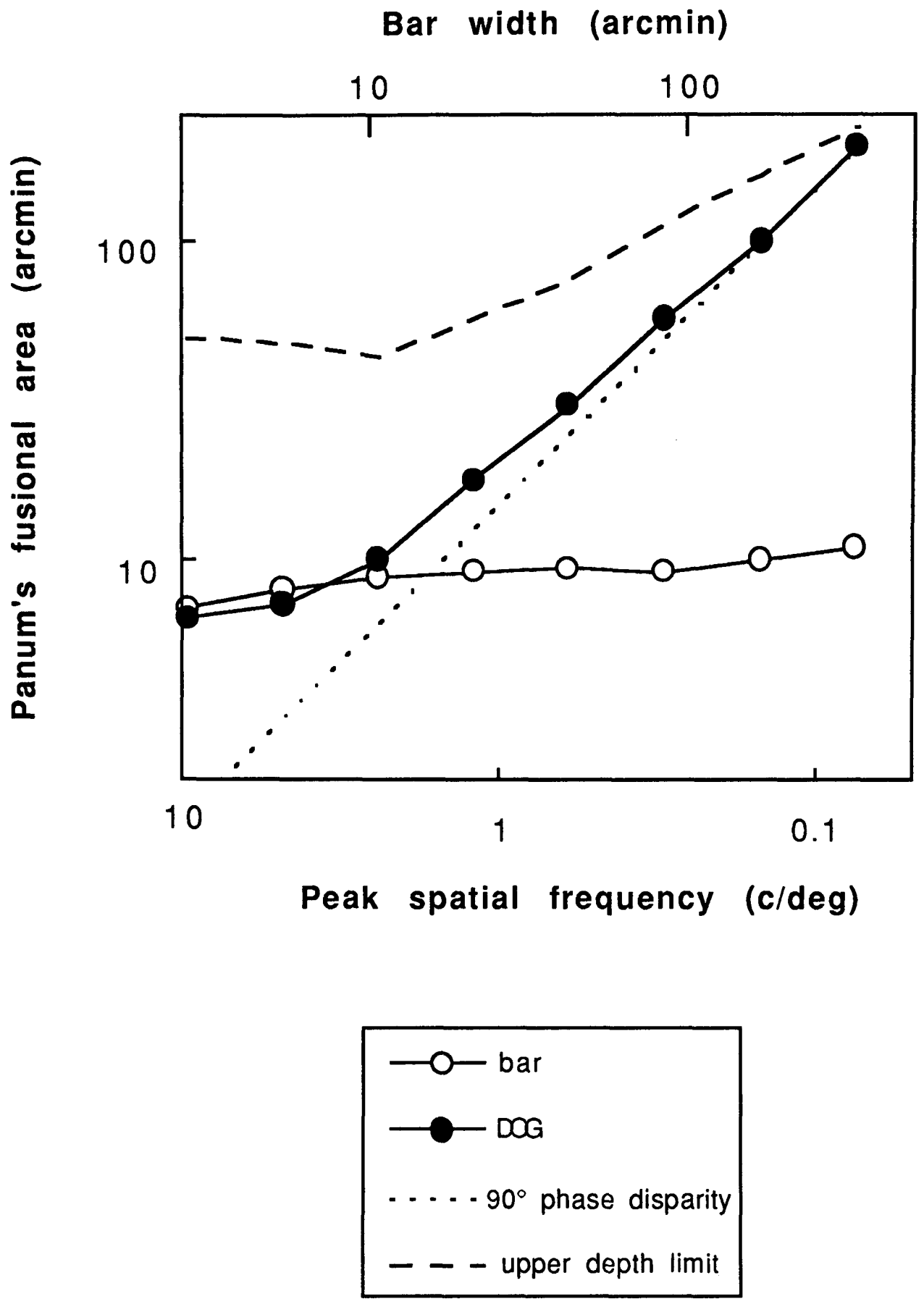


Fig 1.8

This figure illustrates how diplopia thresholds vary with the spatial frequency content of the stimulus. Results are re-plotted from Schor, Wood and Ogawa (1984) for one subject. The abscissa shows the width of the unfiltered bar or, for the DOG stimulus, the width of the excitatory centre (above). For DOG stimuli the centre frequency is also given (below). The disparity which gives rise to diplopia for bar stimuli (open circles) and for DOG stimuli (filled circles) is shown. The upper limits of depth perception (defined in text) are also shown for DOG stimuli (dashed line). The diagonal dotted line shows the disparity for a constant phase angle of 90°.

closely approximate 90° phase disparity when expressed in cycles of the stimulus centre frequency). Together these results provide good evidence for the conclusion that high spatial frequencies determine the disparity at which diplopia occurs (known as Panum's fusional area)* .

The question Marr and Poggio ask becomes a different, but still valid one, without the caveat of "non-diplopic conditions": i.e. what aspect of the stimulus limits the magnitude of perceived depth, given fixation is maintained? In an experiment described by Schor and Wood (1983), subjects matched the depth of an unfiltered bar by manipulating the disparity of a DOG stimulus (fixation was controlled with nonius lines). Schor and Wood only tested disparities for which a match could be made with the narrowest DOG stimulus so a direct test of Marr and Poggio's prediction cannot be made on the basis of these results. However, Schor and Wood found that the relationship between disparity and perceived depth was not constant across scale (much larger disparities had to be added to wide DOG stimuli for them to be perceived at the same depth as the standard) and there is a suggestion in these data (their figure 2) that, even though larger *disparities* may be detected in stimuli containing low spatial frequencies, the magnitude of *perceived depth* may be no greater.

In summary, there is again no clear support for Marr and Poggio's predictions. In particular, the assumption they appeared to make, that diplopia reflects the limitations of the correspondence process, has proved incorrect.

1.5.6 Panum's fusional range

> *"5. (P***) In the absence of eye movements, the minimum fusible disparity range (Panum's fusional range) is +/- 3.1' in the fovea, and +/- 5.3' at 4 deg. eccentricity. This requires that only the smallest channels be active.*
> *6. (P***) In the absence of eye movements, the maximum fusible disparity range is +/- 12' (possibly up to +/- 20') in the fovea, and about +/- 34' at 4 deg. eccentricity. This requires that the largest channels be active, for example by using bars or other large bandwidth stimuli."*
>
> (Marr and Poggio, 1979, p321)

---

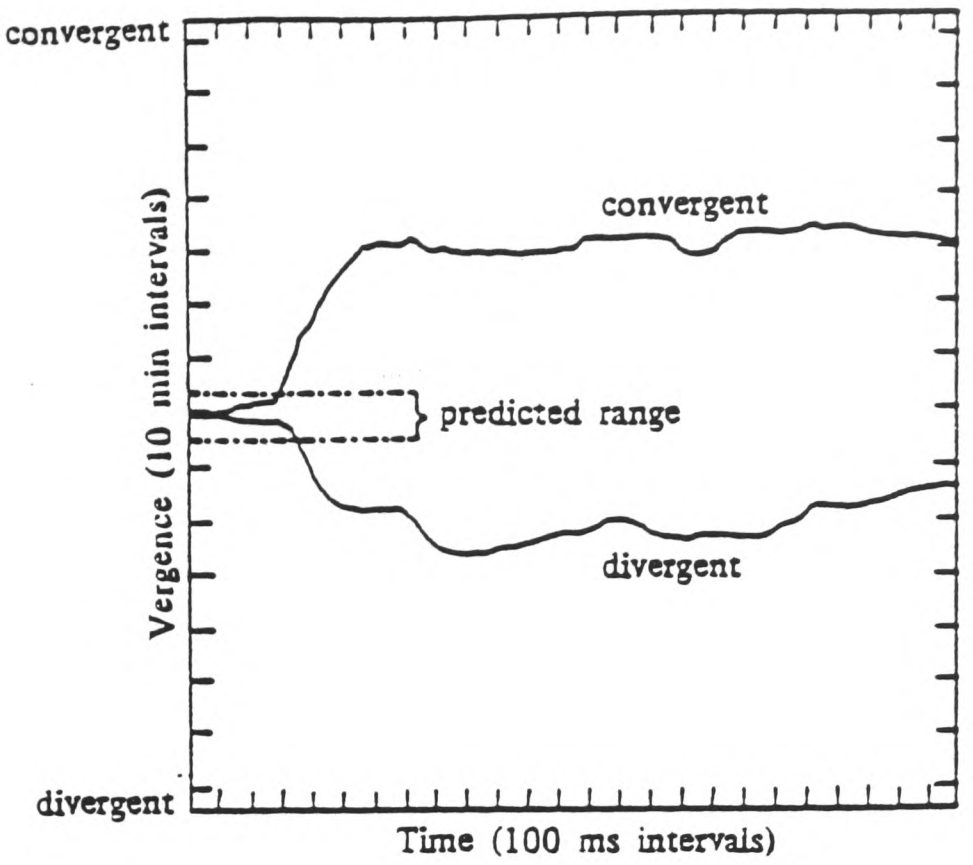
* This is not unreasonable since stimuli far from the fixation plane are likely to be blurred on the retina. The exact relationship between disparity and accommodative blur depends on the fixation distance, but certainly it can be said that the sharply focussed bar stimulus presented at large disparities is an unnatural one. It is unclear exactly what diplopia reflects. It is a subjective measure and does not correspond to the maximum disparity for which depth is perceived or even the disparity which yields the maximum depth. Schor et. al's (1984) results suggest that, if in natural scenes there is a statistical relationship between stimulus disparity and blur (a distribution of blur:disparity ratios), then diplopia may be perceived for stimuli at the extreme of that distribution.

These predictions refer to Panum's fusional area, i.e. they concentrate on diplopia as the signal that correspondence has failed. As discussed above, this is not necessarily the case. The examples of experiments that they claim agree quantitatively with their predictions all use stimuli containing high spatial frequencies. As Schor et al. (1984) showed, when these are removed Panum's fusional area can increase by a factor of ten.

Clearly, in these two predictions Marr and Poggio are concerned with the maximum disparity that can be detected and how this depends on the spatial frequency content of the stimulus. Two studies address this question directly (Mowforth, Mayhew and Frisby, 1981; Frisby and Mayhew, 1980). They measured the maximum detectable disparity by recording eye movements made in response to large disparity stimuli. The patterns used in these experiments were random dot patterns filtered with circularly symmetric, narrow band filters ranging in centre frequency from 2 to 7.5 c/deg. The texture of the entire stimulus area was given a disparity (except the uncorrelated areas caused by the disparity shift, so the "window" remained in the fixation plane).

In Frisby and Mayhew's (1980) experiment, the stimuli were presented for 1 second and nonius lines were flashed on briefly after a given exposure duration. Subjects had to remember and match the offset of the nonius lines after each trial as a measure of their vergence state. Frisby and Mayhew's results showed that vergence movements were made successfully, and without "hunting" eye movements, for targets with a convergent or divergent disparity of 28 arcmin. This was true for all stimulus types including the highest spatial frequency filtered patterns (7.5 c/deg), although vergence was rather slower for the highest frequency patterns.

Mowforth et al.. (1981) used similar stimuli and measured eye movements with an eye tracker. They found a similar pattern of results, i.e. that vergence movements could be made in response to large disparity stimuli carried by high spatial frequencies. Figure 1.9 (top) shows eye movements made in response to one convergent and one divergent stimulus, together with the range predicted by Marr and Poggio. The disparity of these stimuli is *seven times* larger than that which should, in Marr and Poggio's theory, be detectable for a stimulus of 7 c/deg. This



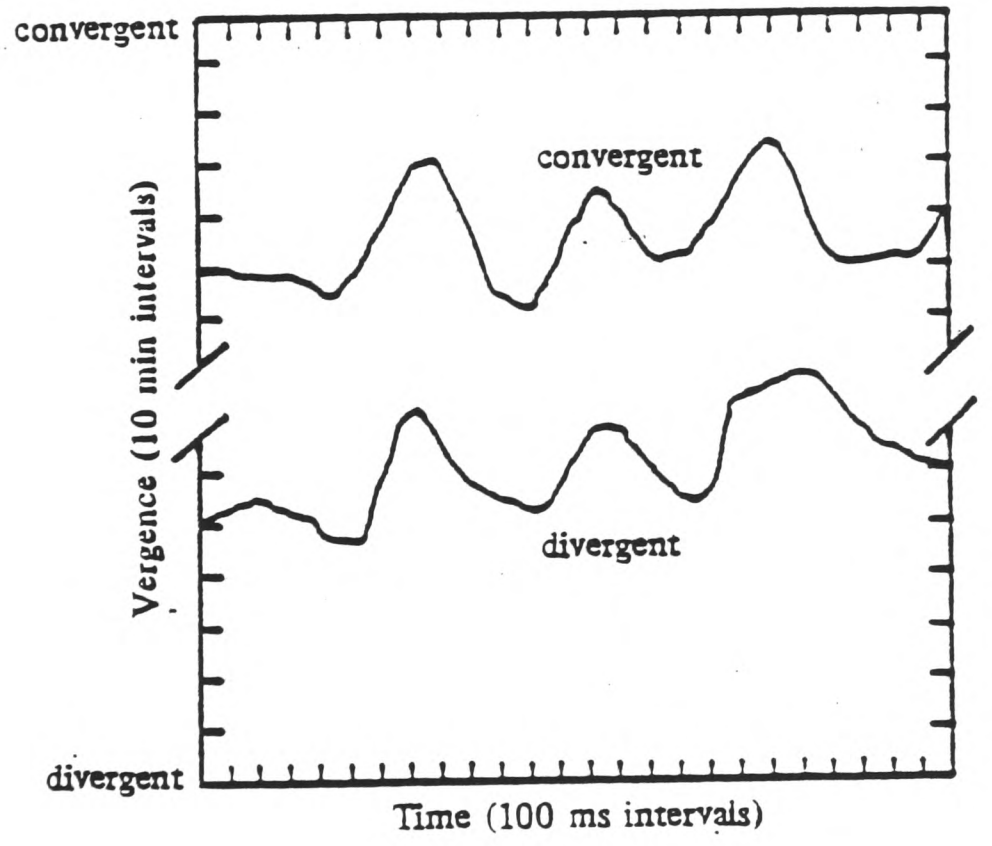


Fig 1.9

This figure (from Mowforth et. al, 1981) shows eye traces made in response to filtered random dot patterns containing a large disparity. Exposure duration is shown on the abscissa, vergence state on the ordinate. In the plot shown at the top, two traces are shown, one in response to a pattern with 28 arcmin convergent disparity, the other to an equal divergent disparity. The centre frequency of the pattern was 7.5 c/deg. and the range of eye movements predicted by Marr and Poggio (1979) is shown on the trace. Below, results for the same pattern but a larger disparity (56 arcmin) are shown. In this case, "hunting" vergence eye movements take place.

is a surprising result. It is, for instance, much larger than the results found for comparable studies in the motion domain, discussed in chapter 6. Much emphasis has been placed on this paper, particularly in discussions of the evidence against Marr and Poggio's theory. It is unfortunate that no quantitative study has repeated their observations (Mowforth, Mayhew and Frisby publish only one "representative" eye trace).

In addition to individual problems with these two experiments, it is possible to criticise the whole approach of using filtered stimuli to answer the questions Marr and Poggio raise. For example, as discussed earlier, it is not easy to account for the experimental finding that band-pass stimuli presented with a large disparity and at medium contrast may not support depth judgements but will do so at a higher contrast (Frisby and Mayhew, 1978; Smallman and MacLeod, 1992). At least, this result does not fit very well with Marr and Poggio's idea of correspondence being limited by zero-crossing spacing. One possible explanation is that high contrasts exaggerate non-linearities in the filters and so introduce spurious low spatial frequencies. If this were the case, the results of experiments using relatively high contrast band-pass stimuli would become rather difficult to interpret al.though, at first sight, band-pass filtered stimuli appear be suitable for studying mechanisms tuned to different spatial frequencies, paradoxically it is often difficult to determine which "mechanism" is responsible for the observed results. Such criticism does not apply to the use of band-pass stimuli presented contrasts close to their detection threshold and the principal value of using band-pass stimuli has been in connection with contrast threshold measurements. A different approach is taken in chapter 6 to the question that Marr and Poggio ask, that is, what limits the maximum disparity that can be detected?

1.5.7 Adaptation

> *"8 (A) As measured by disparity specific adaptation effects, the optimum stimulus for a small disparity is a high spatial frequency grating, whereas for large disparities, the most effective stimulus is a low spatial frequency grating. Furthermore, the adaptation effect specific to disparity is greatest for gratings whose periods are twice the disparity (Felton, Richards and Smith 1972). (In our terms, in fact, λ is approximately 2.2* w *where λ is the centre frequency of the channel.)"*
>
> (Marr and Poggio, 1979, p322)

Felton, Richards and Smith's (1972) experiment was based on Blakemore and Campbell's (1969) demonstration that contrast elevation after adaptation to a

luminance sine wave grating was spatial frequency specific. Felton et al.. compared contrast elevation after adaptation in the fixation plane to contrast elevation after adapting (and testing) at a convergent disparity. For most spatial frequencies the amount of contrast elevation was the same in both cases. However, for gratings whose spatial period was twice the disparity (so the gratings were 180° out of phase in the two eyes) Felton et al.. observed particularly marked adaptation. (They expressed their results as the ratio of contrast elevation after adaptation in the disparate plane to contrast elevation after adaptation in the fixation plane and, for a range of different spatial frequency gratings, this ratio peaked when the phase difference of the grating in the left and right eyes was 180°.) A disparity of 180° is just *outside* the range ± w for the channel whose centre frequency is $1/\lambda$ ($w = \lambda /2.2$) so one might expect, in Marr and Poggio's model, the *peak* sensitivity to be at a lower disparity. In other words, this experiment does suggest a link between size and disparity, indeed it was one of the first papers to do so, but the quantitative fit with Marr and Poggio's model is not quite as close as they imply.

1.5.8 Slant

> *"10 (P***). In the absence of eye movements, the perception of tilt in stereoscopically viewed grating pairs of different spatial frequencies is limited by 4,5, and 6 above."*
>
> (Marr and Poggio, 1979, p322)

Marr and Poggio's prediction relates to an experiment reported by Blakemore (1970a) showing that slant could be perceived in vertically oriented gratings of slightly different spatial frequencies. He emphasised that the whole surface appeared fused even for quite large slants (up to 30°) and for high spatial frequency patterns. In these cases, any algorithm working on local sign would inevitably make false matches since only one part of the grating can be near the fixation plane at a time. Marr and Poggio suggest that if eye movements were restricted then this false match limit (i.e. disparities of $\pm w$ over the whole surface) would mean that only relatively small slants could be perceived, particularly for the high spatial frequency gratings. The fact that Blakemore observed very much larger slants, according to Marr and Poggio, is because eye movements were allowed. They imply that the percept of the whole grating as fused is a result of the $2\ ^1/_2$-D sketch which is built up from local matches made during a series of eye movements. The $2\ ^1/_2$-D sketch can store disparities up to about 2° which, over a 3° patch, fits quite well with the magnitudes of slant Blakemore reported.

The same issue is addressed in an experiment by Mitchison and McKee (1987a) who used patterns of regularly spaced dots that were expanded (horizontally) in one eye relative to the other. Again the question of local matches arises - whatever the vergence of the eyes some of the dots will always be incorrectly matched by a "nearest neighbour" rule, and yet for human observers, at exposures too brief for eye movements, the "correct" (slanted) surface is perceived. The results of their experiment (as discussed in section 1.3.3) do not support Marr and Poggio's prediction.

1.5.9 "Hunting" eye movements

> *"13 (P***). For a novel two planar stereogram vergence movements should exhibit a random-search-like structure. The three star status holds when the disparity range exceeds the size of the largest masks activated by the pattern."*
>
> (Marr and Poggio, 1979, p322)

Mowforth, Mayhew and Frisby (1981) give an illustration of "random search-like eye movements" made in response to a high frequency, large disparity stimulus. These are shown in figure 1.9 (bottom trace). However, the experimental results of Mowforth et al.. agree only qualitatively with Marr and Poggio's theory and, as discussed in section 1.5.6, there are problems determining which is the "largest mask activated by the pattern.".

1.6 Summary

In this chapter the two main problems within stereopsis have been described: the correspondence problem and the problem of interpreting disparities. Of these, the correspondence problem is the more fundamental, since any distortions resulting from this first stage will affect the interpretation of disparities, a theme taken up in chapters 5 and 7. It is the correspondence problem and how it is solved in human vision that forms the focus of this thesis.

Radically different approaches to the correspondence problem have been put forward and it is surprising that there is, as yet, no agreement on which one best models the matching process in human vision. In section 1.3 the most successful correspondence algorithms were summarised and the advantages of a coarse-to-fine strategy, which avoids false matches altogether, were discussed. The best

known coarse-to-fine algorithm, put forward by Marr and Poggio in 1979, was examined in detail. This has many theoretical advantages and has been successfully implemented in machine vision (Grimson, 1981) for natural images and random dot stereograms.

Much of the psychophysical evidence, discussed in section 1.5, supports the spirit of Marr and Poggio's theory. In particular one fundamental aspect of their theory, the relationship between size and disparity (i.e. between coarse scale features in the image and large disparities, fine scale features and small disparities) recurs many times. For instance, larger disparities are fusible for coarse scale than fine scale images (Badcock and Schor, 1985; Schor, Wood and Ogawa, 1984); larger disparities are detectable for coarse than fine scale images, as demonstrated by the recording of eye movements (Mowforth , Mayhew and Frisby, 1981); adaptation to a gratings off the fixation plane suggests a proportional relationship between scale and disparity (Felton, Richards and Smith, 1972); and contrast sensitivity to band-pass filtered stimuli off the fixation plane shows a similar pattern (Smallman and MacLeod, 1992).

The details of the theory are less well supported. Marr and Poggio gave an unprecedented list of precise, quantitative predictions by which their theory could be tested and, by now, all of the main predictions have been challenged. The primitives they suggested, zero-crossings, are probably not used in human stereopsis (Legge and Gu, 1989). Vergence eye movements are probably not the only means of guiding fine scale matches. Matching of features on slanted planes cannot be explained within their framework (Mitchison and McKee, 1987a). There is little direct evidence for a $2\ ^1/_2$ -D sketch, particularly for the idea that the distance from the observer to each surface in the image is recorded explicitly.

This does not necessarily mean that a coarse-to-fine model of human stereopsis should be abandoned altogether. Various attempts have been made to modify Marr and Poggio's basic idea to make it more compatible with psychophysical data. The next chapter begins with a discussion of these.

CHAPTER 2

2.1 Modifications to Marr and Poggio

In the last chapter, the idea of a coarse-to-fine matching algorithm was introduced as a computationally efficient solution to the correspondence problem. One algorithm in particular, by Marr and Poggio (1979), was discussed in detail. Although this was an elegant theory, many of their psychophysical predictions have not been supported by subsequent findings. Partly as a result of this failure, most current models of the matching process in human vision take a completely different approach (e.g. Prazdny, 1985; Pollard et al.., 1985), using as their starting point information at the finest scale.

Nevertheless, interest in the coarse-to-fine approach continues. One reason for this is that evidence of a size-disparity correlation in psychophysical experiments (summarised at the end of the last chapter) is not explained by co-operative ("local-to-global") algorithms. Also, co-operative algorithms fail to account successfully for matching in some situations (e.g. Mitchison and McKee, 1987a & 1987b, discussed in section 1.3). The question arises, rather than abandoning Marr and Poggio's theory altogether, is it possible to modify it to accommodate the more recent psychophysical evidence?

2.1.1 "Neural shifts"

A few attempts have been made to modify Marr and Poggio's (1979) algorithm. Mainly these have been concerned with the method of guiding fine scale matching once a coarse scale match has been found. The reason was that Marr and Poggio's proposal (of using vergence eye movements to guide fine scale matches), appeared to be the weakest parts of their theory (it was directly criticised by, for example, Mowforth et al.., 1981; Mayhew and Frisby, 1979; and Mitchison and McKee, 1987b, as discussed in chapter 1).

Several alternative proposals for guiding fine matching without the use of eye movements have been published (e.g Nishihara, 1984; Quam, 1984; Anderson and Van Essen, 1987)). The general solution is to use some sort of "neural" shifting instead of shifting the whole retina physically with an eye movement. The theoretical advantage of this strategy would be that a different shift could be applied to different parts of the image.

Nishihara (1984) and Quam (1984) suggest the shift takes place in disparity space. This means that, as for co-operative models (e.g. Marr and Poggio (1976) or Pollard et al.. (1985)), there must theoretically be a neuron, or group of neurons, that code for every possible disparity at every retinal location (i.e. there is a one-to-one relationship between a point in disparity space and the neurons or "nodes" that represent that point). Figure 2.1 illustrates this schematically. Exactly how densely the neurons must be spaced along the disparity axis depends both on stereoacuity at different pedestal disparities and on the degree to which rate-coding and interpolation are used (discussed in Lehky and Sejnowski, 1991). The search for matches within this disparity space is guided by coarse scale matches, in the way Marr and Poggio (1979) suggested. In physiological terms this means that there must be neurons with very small receptive fields at widely disparate locations on the two retinas that together can stimulate a fine-scale, large-disparity unit. In fact, there must be a neuron, or group of neurons, for every possible disparity at every location. However, only those neurons sensitive to disparities similar to the coarse scale disparity are allowed to be active.

Anderson and Van Essen's (1987) model proposes a shift in the image domain, i.e. before the matching process (e.g. a coarse scale match could be used to "shift" the x-location of all the fine scale features in that area). Although it is described differently, a shift in the image domain and disparity domain are formally equivalent.

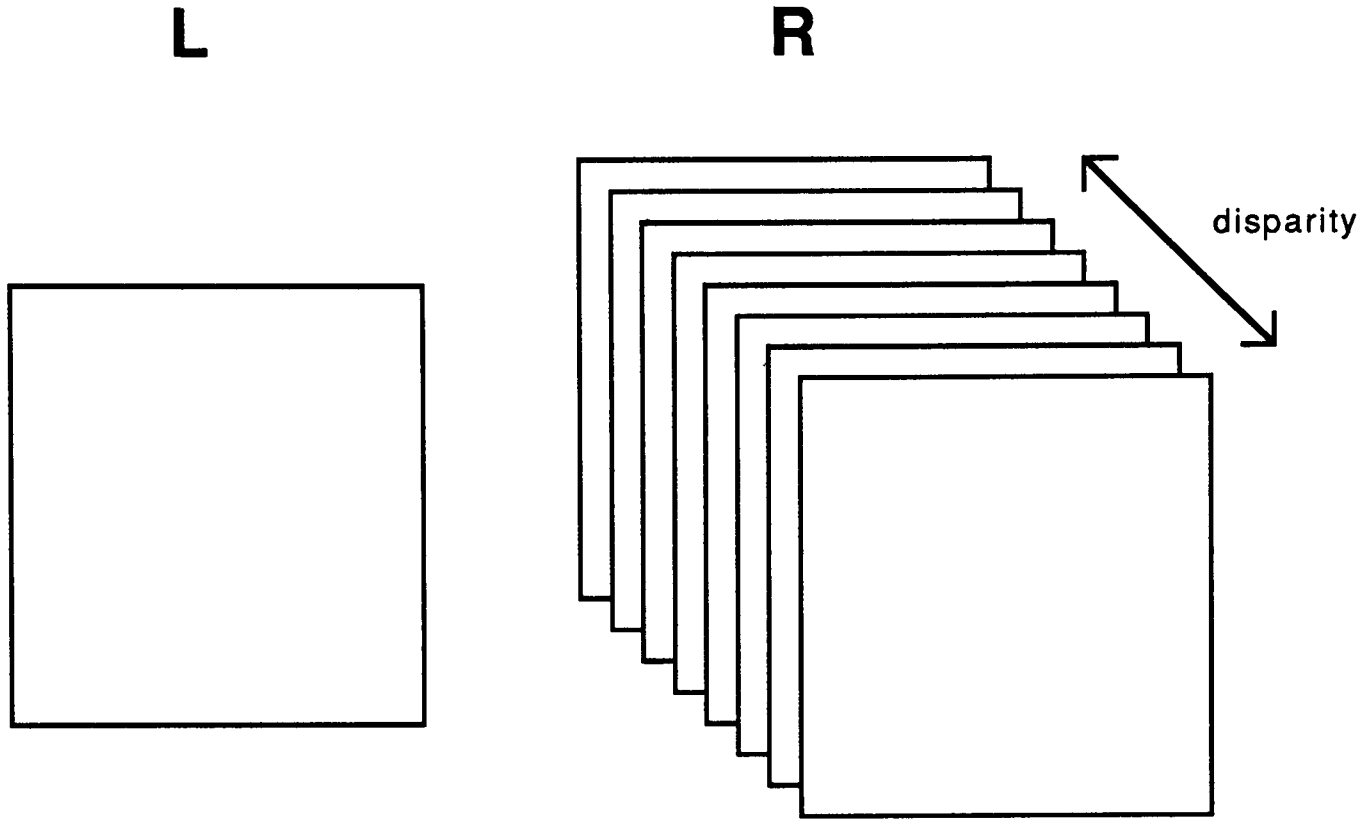


Fig 2.1

If a coarse-to-fine strategy is to be implemented using a "neural shift" (as suggested by Nishihara (1984), Quam (1984) and Anderson and Van Essen (1987)) then, for each patch of the left eye's image there must a set of neurons "looking at" a correspondingarea in the right eye's image (zero disparity) and at every other possible disparity, i.e. a three-dimensional "volume" of neurons is required for each retinal location.

The role of fine scale mechanisms is quite different in a neural shift model to that in Marr and Poggio's theory, at least when the signalling of disparities across the whole scene (and several eye movements) is considered. In Marr and Poggio's scheme fine scale disparities can only be determined close to the fixation plane. To determine fine scale disparities over a wide range of depths, eye movements are required. In effect, the fine scale mechanisms signal a fine scale *adjustment* to a coarse scale disparity which has already been measured and recorded (in the $2^1/_2$-D sketch). In a neural shift model fine scale mechanisms signal the absolute disparity of an object (i.e. with respect to the fixation plane), hence there must be neurons that respond to every point in disparity space. The principal question addressed in this chapter is whether a version of Marr and Poggio's model is possible in which fine scale mechanisms signal the disparity of a feature relative to the coarse scale disparity (as in the $2^1/_2$-D sketch) but without the need for eye movements.

Two other issues must be addressed when considering a neural shift algorithm. One is segmentation, i.e. how each fine scale feature is "ascribed" to a coarse scale feature and hence shifted with it. The second is that any shift nulls the coarse scale disparity of a fronto-parallel surface but not of a surface slanted in depth (the same is true of a vergence movement). Slanted surfaces can at best be approximated by a series of fronto-parallel planes.

The first issue (segmentation) is considered later in this chapter. The second (fine scale disparities on slanted surfaces) is discussed in chapter 3. First, the concept of relative disparity is discussed with particular reference to a coarse-to-fine matching scheme.

2.2 Relative disparity

2.2.1 Psychophysical evidence.

The absolute disparity of a target is equal to the difference between its binocular parallax and the ocular vergence angle (illustrated in figure 2.2). It can also be described as the difference in local sign between the image of the target in the left and right eye. The relative disparity of a feature, on the other hand, requires there to be at least two visible points in the image (see figure 2.2). Absolute disparity depends on the angle of vergence, relative disparity does not. There are several different proposals about how relative disparity might be derived in the visual

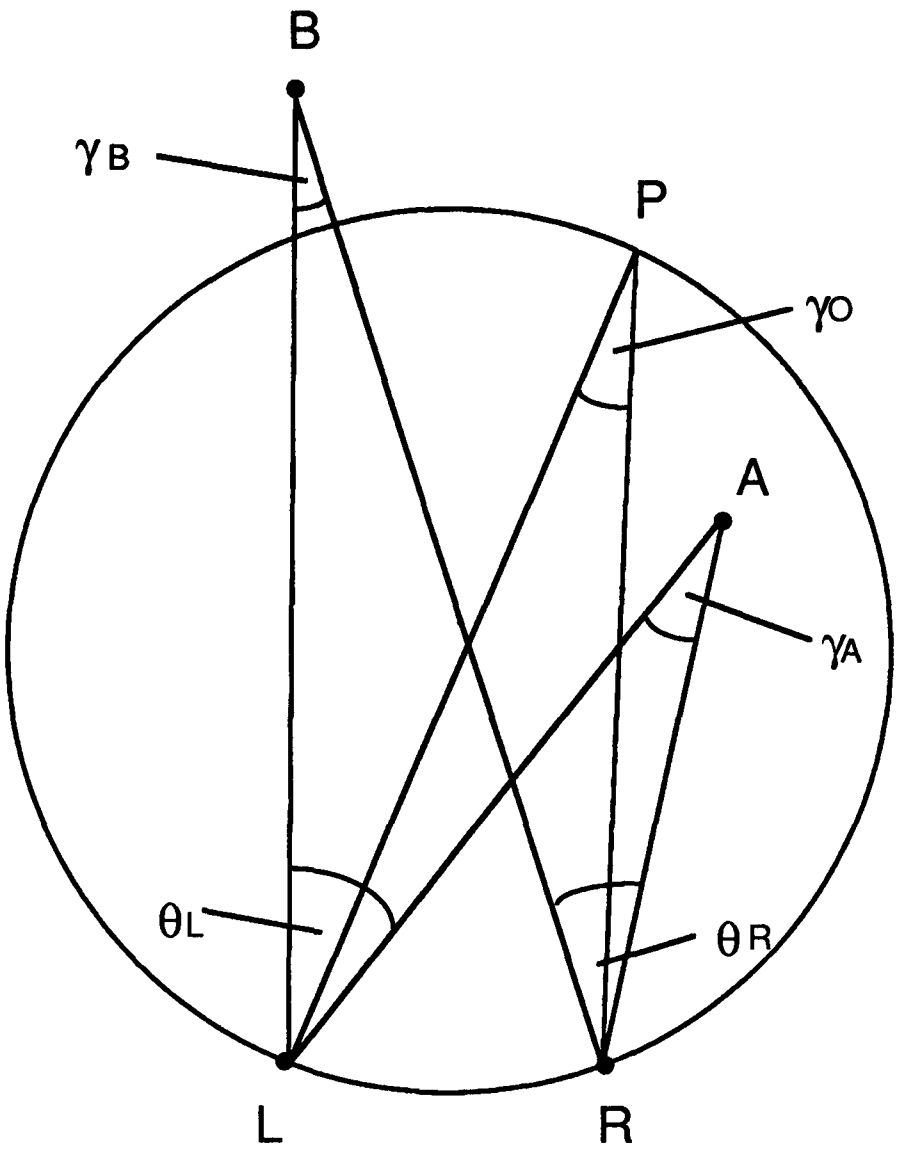


Fig 2.2

The angles of ocular vergence (γ_O) and binocular parallax (γ_A γ_B) at two points, A and B. The Vieth-Müller circle is shown passing through the optical centres of the left and right eyes (L and R) and the intersection of the lines of sight (P). The absolute disparities of points A and B are:

$$\delta_A = \gamma_A - \gamma_O \quad \text{and} \quad \delta_B = \gamma_B - \gamma_O .$$

These relations hold irrespective of whether there is a visible target at P.
The relative disparity of points A and B is:

$$\delta_{AB} = \delta_A - \delta_B = \gamma_A - \gamma_B$$

or alternatively,

$$\delta_{AB} = \theta_L - \theta_R .$$

The latter equation emphasises that relative horizontal disparity can be expressed as the difference in angular separation of points A and B at the two eyes, i.e. width disparity. The geometry illustrated in this figure is a simplified approximation because it considers only horizontal disparities.

The relationships described here are independent of where the eyes are looking (they apply to the optic array at two points). Strictly, the width disparity of points A and B does not remain invariant with eye movements when described in terms of positions on the retina since the centre of rotation of the eye is not at its optical centre.

system (e.g. Blakemore, 1970a; Rogers and Cagenello, 1989; Ohzawa et al.., 1990; Motter and Poggio, 1990, discussed below). Most methods involve a monocular measurement of the relative positions of features (e.g orientation, separation or curvature) followed by a binocular comparison of these measurements. The debate as to whether absolute (retinal) disparity or relative disparity form the basis of binocular depth perception is long-standing and as yet has no definitive answer.

Several psychophysical studies suggest that absolute disparity on its own is a poor cue to depth. Westheimer (1979) compared stereoacuity thresholds for ordinary, simultaneously presented line pair stereograms with those for successively presented lines at different absolute disparities. The time interval between the presentations was zero, with a single binocular line acting as a self-reference. Thresholds for successive discrimination were about 1'; they were at least ten times larger than a subject's best stereoacuity threshold. The result suggests, as Westheimer pointed out, that some mechanism based on relative rather than absolute disparities is responsible for good stereoacuity. He noted that such a mechanism would make stereoacuity insensitive to a variety of naturally occurring disturbances such as movement of the eyes or stimulus.

Erkelens and Collewijn (1985) reached a similar conclusion on the basis of a very different experiment. They studied the perception of random dot stereograms in which the half-images were moving sinusoidally in opposite lateral directions. If no fixation target or other fixed reference was present the stereogram was perceived as fused and stationary. The target motion evoked sinusoidal ocular vergence movements with a gain smaller than unity and a certain delay, from which it was deduced that the retinal disparity of the half-images was changing during the oscillations. However, only when a visible fixation marker was present did the stimulus appear to change in depth (loom and recede). They point out the implications this finding has for theories about the use of absolute and relative disparity information in the visual system:

> *"The .. conclusion of this study is that neither convergence nor absolute disparity are cues for motion in depth. This makes binocular perception of the visual space to a certain extent independent of the eye positions. For the perception of an object in three-dimensional space, the locus of the two retinal images is less important than the relationship of these retinal images to other parts of the retinal images. This limited independence of eye positions creates the advantage that imperfections of oculomotor control which have been reported to exist* [e.g. Steinman et al., 1982] *do not have perceptual consequences."*

(Erkelens and Collewijn, 1985, p587)

Erkelens and Collewijn found that fusion and the perception of depth in random dot stereograms was possible when the vergence error (and hence retinal disparity) was between 1 and 2 degrees of arc. This value is similar to that found by Fender and Julesz (1967) using stabilised images, and in many ways the experiments are closely related. Fender and Julesz used an apparatus that could stabilise images on the retina of each eye. This meant that vergence movements would not alter the retinal disparity of any part of the image. However, the disparity of the image could be manipulated by the experimenter. Fender and Julesz measured the disparity at which subjects perceived stimuli as diplopic when a progressively larger divergent disparity was applied (the stimuli were moved symmetrically temporalward in the visual field) and also the disparity at which stimuli were perceived as fused when the disparity was reduced. The two values differed, "break-away" occurred at a larger disparity than re-fusion, a phenomenon that Fender and Julesz named hysteresis. Both line targets and random dot stereograms were used. Hysteresis was most pronounced for the random dot stereograms. Under conditions of stabilised vision the retinal disparity of the pattern could be increased up to about 2 degrees without diplopia or loss of depth perception, but, once lost, the disparity had to be reduced to about 10 arcmin before fusion was re-gained.

Fender and Julesz interpreted their results as evidence for some cortical registration process similar to that described by Anderson and van Essen (1987) in which, provided the retinal image is shifted within certain limits (up to about 2 degrees) and at a sufficiently slow rate, correspondences that were determined when the stimulus was within Panum's fusional range can be maintained. Marr and Poggio (1979) also suppose that correspondences, once achieved, can be maintained at large retinal disparities, through the 2 1/2-D sketch. However, the idea is rather different from a neural shift or cortical registration model since the 2 1/2-D sketch is a buffer or memory. This would lead to different predictions. For instance, Fender and Julesz's model is compatible with the detection of a small change in the stimulus when it is stabilised at a large disparity whereas Marr and Poggio's model, being based only on a memory, is not.

A neural shift model, or a hierarchical model as discussed in chapter 3, is compatible with Fender and Julesz's finding that a large retinal disparity can be added to a random dot stereogram without destroying the perception of depth. The

emphasis Fender and Julesz give to the perception of a fused versus a diplopic image as a measure of the limits of stereoscopic mechanisms may not be justified. As discussed in chapter 1, Schor, Wood and Ogawa (1984) have distinguished diplopia, which may relate to the high frequency content of an image, from the ability to discriminate depth. This distinction is examined in more detail in chapter 6. It is possible that depth differences could be discriminated for line targets in stabilised vision at large retinal disparities (e.g 2 degrees) despite diplopia occurring at a disparity of about 1 degree. Fender and Julesz do not provide data that bear on this point.

A coarse-to-fine neural shift model (e.g. Nishihara (1984) or Quam(1984) and also the hierarchical model discussed in chapter 3) seem to imply a co-ordinate system that is entirely "free-floating" which makes it difficult to explain why a random dot stereogram must be brought into such close retinal correspondence (6 arcmin) before fusion and a perception of depth can be re-gained after the images have been pulled apart. There are two factors that may "tie down" a relative co-ordinate system so that, at least as a default, it has similar characteristics to a retinal co-ordinate system. The first is the surround. Fender and Julesz used an apparatus in which the image was projected through a telescope into the eye. Light arriving through the telescope was stabilised with respect to the retina but Fender and Julesz do not give details of whether any of the rest of the scene was visible in the periphery. The second is the fovea. It is possible that the reason fusion is obtained for random dot stereograms with near zero disparity is that the image in each eye lies on the fovea in each eye, rather than because of its disparity *per se*. Blakemore makes a similar point when discussing the perception of large disparity, diplopic images:

> *"...fusional eye movements (Rashbass and Westheimer, 1961), which bring objects on to the horopter, and therefore into fusion, also bring them into the range of maximum stereo-acuity. A possible reason for the complicated system of retinal correspondence thus emerges. Although relative disparity is analysed over a large range of absolute disparity, it is useful to bring objects of interest to the region of maximum resolution. This is achieved by making images with inappropriate absolute disparity the cue to a vergence movement to bring them to that region. The subjective consequence of this system is the fusion of images on corresponding points."*
>
> (Blakemore, 1970b, p616)

A related demonstration by Julesz (1971) raises several interesting issues. He showed that random dot stereograms in which one eye's image was enlarged relative to the other by a factor of 10 or 15% (in both horizontal and vertical

dimensions) could still be fused and a central square seen in depth. First, the demonstration raises issues about how the correspondence problem is solved. When the retinal disparity of some points in the stereogram is zero, the horizontal and vertical disparity of many other points in the stereogram is large (how large depends on the size of the stimulus) and yet the correspondence problem can be solved. This would appear to suggest that the search for possible matches for each dot is carried out over quite a wide extent, both vertically and horizontally. Such a wide search would add considerably to the computational load of any co-operative algorithm (e.g. Marr and Poggio, 1976; Pollard, Mayhew and Frisby, 1985). An alternative possibility (discussed in section 3.2 and 7.2) is that the fine scale co-ordinates are "re-scaled" according to the overall width (for horizontal co-ordinates) and height (for vertical co-ordinates) of the random dot pattern in each eye. After re-scaling, the positions of the background dots in the left and right eye's images would be the same.

The second issue raised by Julesz's demonstration is the way in which the vertical expansion of one eye's image "nulls" the effect of the horizontal expansion (on its own, a horizontal size difference would result in the stereogram being perceived as slanted about a vertical axis). The observation is closely related to Ogle's "induced effect" (Ogle, 1950). He showed that a vertical expansion of one eye's image, without any change in horizontal width, resulted in the perception of a stereogram as slanted. Rogers and Koenderink (1986) have interpreted the induced effect as evidence that the visual system might use the deformation component of the disparity field (Koenderink and van Doorn, 1976) to determine surface slant. This is equivalent to saying that the visual system is sensitive to the horizontal width difference *after* the images in the left and right eyes have been uniformly expanded or contracted to account for vertical size differences. The analysis of Mayhew and Longuet-Higgins (1982) accounts for the effect in a similar way, although it depends on the calculation of global parameters whereas the measurement of differential invariants that Rogers and Koenderink (1986) propose could be carried out locally and independently in different parts of the image.

Both the algorithm of Mayhew and Longuet-Higgins (1982) and Rogers and Koenderink (1986) assume that the correspondence problem has already been solved. An alternative, as discussed above, is that the whole random dot pattern is in some way expanded or "re-scaled" before correspondence of the fine scale dots is determined. An algorithm on these lines could be designed to have properties very similar to those proposed by Koenderink and van Doorn (1976). Suppose

that an isotropic expansion or re-scaling of one eye's image was carried out first so that the vertical dimensions of the pattern in the left and right eyes were matched. Then any "residual" horizontal difference in width would signal the slant of the pattern. In chapter 7 this and other possible manipulations are discussed (section 7.2). In summary, a "re-scaling" model might help explain both how correspondence is achieved and how the slant of the surface is perceived in Julesz's demonstration.

The idea that relative disparities might be important in human stereopsis was discussed by Blakemore (1970a), although he supposed that these would be used in addition to absolute disparities rather than replacing them altogether. He says in his introduction:

> *"Consider the hypothesis that it is not only the relative positions but also the separation or spatial periodicity of the retinal images of disparate objects that can be analysed to produce the sensation of relative depth. It may be the gap between two contours on the retina that is measured and compared with the distance between the corresponding contours in the other eye, as well as the absolute positions of the contours themselves."*
>
> (Blakemore, 1970a, p1181)

Tyler and Sutter (1979) criticised some of the techniques Blakemore used but, in a series of further experiments, they reached a similar conclusion. Both Blakemore and Tyler and Sutter considered only the detection of width or spatial frequency disparity, which are usually produced by surfaces slanted about a vertical axis.

Surfaces slanted about a horizontal axis result in differences in the *orientation* of features in two eyes. Orientation disparity, like spatial frequency disparity, provides a signal about surface slant that does not vary with the vergence angle of the eyes. Cagenello and Rogers (1988) provided some evidence in favour of the hypothesis that orientation disparities are used in the human visual system. They showed that thresholds for perceiving slant in surfaces rotated about a vertical or horizontal axis depended on the orientation of the surface markings. Thresholds were highest for the condition in which there were no orientation differences between the line elements in the left and right eye. Mitchison and McKee (1990) showed that orientation disparity alone probably cannot explain the anisotropy of slant perception although, interestingly, there is a suggestion from their results that orientation disparity at a coarse scale may be more important for low thresholds than fine scale orientation disparities (their figure 7).

Another (higher order) relative disparity is curvature disparity. Rogers and Cagenello (1989) proposed that the visual system might use the difference in curvature of elements in the left and right eye's images to derive information about the local curvature of surfaces. Like orientation and spatial frequency disparity, curvature disparity does not vary with vergence angle.

2.2.2 Relative and absolute disparity: physiological evidence

Two physiological papers are described in this section which present evidence that some neurons in the visual cortex may be sensitive to relative disparity rather than absolute (retinal) disparity.

A recent paper by Motter and Poggio (1990) noted that the very sharp disparity tuning curves of some binocular neurons were difficult to explain given the known variability of an animal's fixation (they reported that reliably different responses are usually obtained for disparities differing by 3 arc min. whereas the 2-D scatter of the monkey's fixation pauses (monocularly) has a standard deviation of 6 - 8 arc min. (e.g. Motter and Poggio, 1984), and estimates of fluctuations in vergence range from about 10' (e.g. St-Cyr and Fender, 1969) to 3' (in a review by Steinman, Cushman and Martins, 1982)). This observation suggests that the neuron may be responding to the relative disparity of the stimulus and its surround (e.g. the fixation target) rather than to the fixation or retinal disparity of the stimulus. Motter and Poggio presented direct evidence in favour of this hypothesis by measuring eye movements in awake, behaving monkeys while recording from visually driven neurons. First, in a series of experiments using only monocular stimulation, they demonstrated neurons whose response was highly correlated with the position of the stimulus with respect to the fixation target but only poorly correlated with calculated retinal position of the stimulus (i.e. when the monkey's actual fixation was taken into account). In other words, the retinal location of the neuron's receptive field appeared to shift with the monkey's eye movements while its receptive field "in real space" was "essentially invariant" (p39), at least during an attentive fixation. Second, they demonstrated that the response of a binocularly driven neuron was better correlated with the relative disparity of the stimulus with respect to the fixation target than with the actual (retinal) disparity of the stimulus.

The model Motter and Poggio propose to account for the behaviour of these neurons is that "visual information is dynamically re-routed or gated in its passage from retina to visual cortex." (p42) and they quote the modelling of Anderson and van Essen (1987, see section 2.1) as an example of how this might be achieved.

However, as they admit, such a model requires information to control the shifting or gating and in this paper they do not propose a way in which this might be done.

A paper by Ohzawa, DeAngelis and Freeman (1990) raises very similar issues. The response of the neurons from which they recorded (in the visual cortex of the cat) depended on the the disparity of the stimulus but was independent of the position of the stimulus over a wide area within the receptive field. Thus, these neurons display a similar property to the binocular neuron described by Motter and Poggio, that is, a response to relative rather than retinal disparity. Ohzawa et al.. observed this type of response only in complex cells and their model reflects this finding. They propose that such a complex cell receives input from several pairs of simple cells, each sensitive to a similar disparity but at different positions or phases within the receptive field. Motter and Poggio do not report whether the binocular neuron they describe was a simple or complex cell but, of the neurons whose type they did determine, all of which were tested monocularly, the majority were simple cells.

The model Motter and Poggio use to account for their results is different in one important respect from that of Ohzawa et al.. Motter and Poggio propose that the binocular "dynamic stabilisation" that they observed could be explained if binocular neurons received their input from cells with monocularly "stabilised" receptive fields such as those described in the first part of their part of their paper. In other words, a monocular signal of the relative position of the stimulus, independent of its retinal location, may be responsible for the behaviour of the binocular neuron. The model of Ohzawa et al.. cannot account for Motter and Poggio's results so well because a "position-invariant" response only emerges at a binocular stage.

The theory developed in this thesis is related to both these models in emphasising the importance of relative rather than absolute (or retinal) disparity. It is closer to the model described by Motter and Poggio than that of Ohzawa et al.. because it assumes that before binocular combination takes place a monocular measurement is made that is independent of retinal location. The model is described in rather different terms and does not include the idea of a dynamically shifting receptive field (see chapter 3). It consistent with both the monocular and binocular results of Motter and Poggio and those of Ohzawa et al..

2.2.3 The requirements for a model based only on relative disparities

There are several different routes by which relative disparities can be calculated. In the models described in the previous two sections, most included the measurement of absolute disparities at some stage. For example, the model Ohzawa et al. (1990) put forward assumes that the receptive field of the neurons they describe is at a corresponding location on the left and right retina. Those neurons that are sensitive to high spatial frequencies have, they assume, small receptive fields and can only respond over a small range of absolute disparities centred on the fixation plane (DeAngelis, Ohzawa and Freeman, 1991). This raises all the problems that have been discussed in relation to Marr and Poggio's (1979) algorithm and that the neural shift models (Nishihara, 1984; Quam, 1984) were designed to avoid.

The primary measurement in the model of Ohzawa et al. (1990) is an absolute disparity measurement (phase disparity). The same is true of Westheimer's model (1979). He proposed that a mechanism that measured the difference in absolute disparity of two points was responsible for good stereoacuity.

On the other hand, orientation, spatial frequency and curvature disparity (Cagenello, 1990; Blakemore, 1970a; Rogers and Cagenello, 1989) rely on measurements of *relative position* in each monocular image and do not involve an explicit calculation of absolute disparity. However, the result is a local measure of the slant or curvature of the surface. None of the models has proposed that absolute disparities were not used at all, for example in judgements of relative depth of objects in widely separated parts of the image. Blakemore (1970a) makes this distinction explicitly (see section 2.2.1).

Neural shift models (Andersen and van Essen, 1987; Nishihara, 1984; Quam, 1984) are based on retinal co-ordinates, with each node or cell in the system sensitive to a given absolute disparity (or range of disparities).

The model that is developed in the rest of this chapter and in chapter 3 combines ideas from the relative position models (e.g. Cagenello, 1990; Blakemore, 1970a; Rogers and Cagenello, 1989) with the hierarchical scale-based concept used in neural shift models (e.g. Nishihara, 1984; and Quam, 1984).

The central idea is that if the positions of features in the whole image were described in relative co-ordinate framework (i.e. in terms of polar co-ordinates,

orientation and separation) then orientation and width disparities could replace the need to use absolute disparities not just for local judgements of surface shape but for all depth judgements. In order to do this it would seem that the orientation and separation of every possible pair of points in the image would have to be measured. One way to avoid such a large amount of data being required is to order positional information hierarchically.

The models of Nishihara (1984) and Quam (1984) use a coarse-to-fine approach but they are not truly hierarchical. The large number of nodes (or "neurons") required to implement a "neural shift" algorithm reflects the fact that each node can represent only one disparity, or a small range of disparities, at one retinal location in each eye. The extent of the space which must be represented is large (in the two spatial dimensions and in the disparity dimension), and there must be fine scale mechanisms or nodes filling the whole of this "cube". However, if the position of fine scale features in the image were defined *relative to coarse scale features* rather than labelled with a retinal "place code", then it would be natural to measure fine scale disparity, as Marr and Poggio did, as an adjustment of coarse scale disparity. The difference would be that no eye movement would be necessary to "bring fine scale features into alignment".

Thus a hierarchical model can be seen as a natural extension both of the relative disparity models and of the coarse-to-fine models which have already been proposed. In the rest of this chapter, the rationale for representing position hierarchically (i.e. where the position of fine scale features is defined relative to coarse scale features) is examined and some of the properties of this sort of representation discussed.

2.3 A "hierarchical" representation of position

The idea of a hierarchical representation of position was put forward recently by Watt (1987, 1988). The main hypothesis of this thesis is that disparity may be encoded in the same way. That is, if the position of features in the left and right eyes' images is defined hierarchically as Watt proposes, then disparity, the difference in position of a feature in the left and right eye's image, may have the same hierarchical structure. In the rest of this chapter Watt's theory is examined in detail.

The argument for representing position hierarchically may be broken into the following parts:
(i) the rationale for creating a new map or co-ordinate framework for position rather than relying on retinal co-ordinates (i.e. "local sign");
(ii) the rationale for grouping features and for doing so according to proximity;
(iii) the grouping of features which results when the MIRAGE operation (Watt, 1988) is applied to an image;
(iv) the hierarchical representation of position which follows from such a grouping process;
(v) whether the calculation of position needs to be carried out in the same detail right across the image and the time course of this process.

The rationale for each step is theoretical and ought to apply to any visual system. At the end of the chapter, some of the evidence in favour of this theory is discussed.

2.3.1 A new map

The argument for a new map is part of a wider discussion about the metric the visual system might use to represent different attributes of the image. Andrews (1964) proposed that the internal metric by which visual attributes are encoded may be plastic and depend on the visual environment of the organism. For instance, he suggests that the metric for contour curvature may be organised such that the contour which is perceived as straight depends on the statistical distribution of contours in images.

> *"For most of the time the metric of visual space corresponds rather well with that of (E) [the environment]. Kohler (1951) has shown some startling visual phenomena of adaptation. For instance, if we assume distorting spectacles, (R) [the ideal receiver, or the eye] gives a distorted output at first reflected in (I) [the private visual space]. This distortion gradually goes and despite the distorting lenses we see straight lines as straight once more. It follows that the correspondence of (E) and (I) is not the result of a constant transfer function. and this applied equally to the case before the distortion was introduced. How therefore are straight lines perceived as straight?"*
>
> (Andrews, 1964, p110)

His answer is that both the origin and the scale of the metric for curvature may be "self-organising" or "self-correcting": this would be done according to the visual experience of the system, i.e. according to the statistical distribution of the image attribute in the input. Figure 2.3 (a) shows a plausible frequency distribution for curvature in images. The mean of this distribution could be taken to define zero

curvature. Given this definition, it is not surprising that a straight line which projects to a curved image on the peripheral retina is seen as straight - it is the statistics of the input which defines straightness. Similarly for blur, which might have a frequency distribution like that in Figure 2.3 (b), the origin, "a sharp edge", is simply one end of the range of received blurs, it is the smallest blur experienced. If someone had gradually deteriorating vision from a cataract but the metric for blur was self-correcting in this way, they would not necessarily notice the loss of acuity. Two image quantities, position and luminance, do not have reliable or suitable frequency distributions to form the basis of a metric. But it is vitally important that we have a reliable metric for position as we must act on it.

Although the distribution of edges does not have reliable statistics, its derivative, separation, does. More specifically, it is the separation of adjacent contours which has a statistically stable distribution* and it is these (together with orientations) which must be used to build up a co-ordinate framework. Watt (1988) considers the problem of deriving a metric for position:

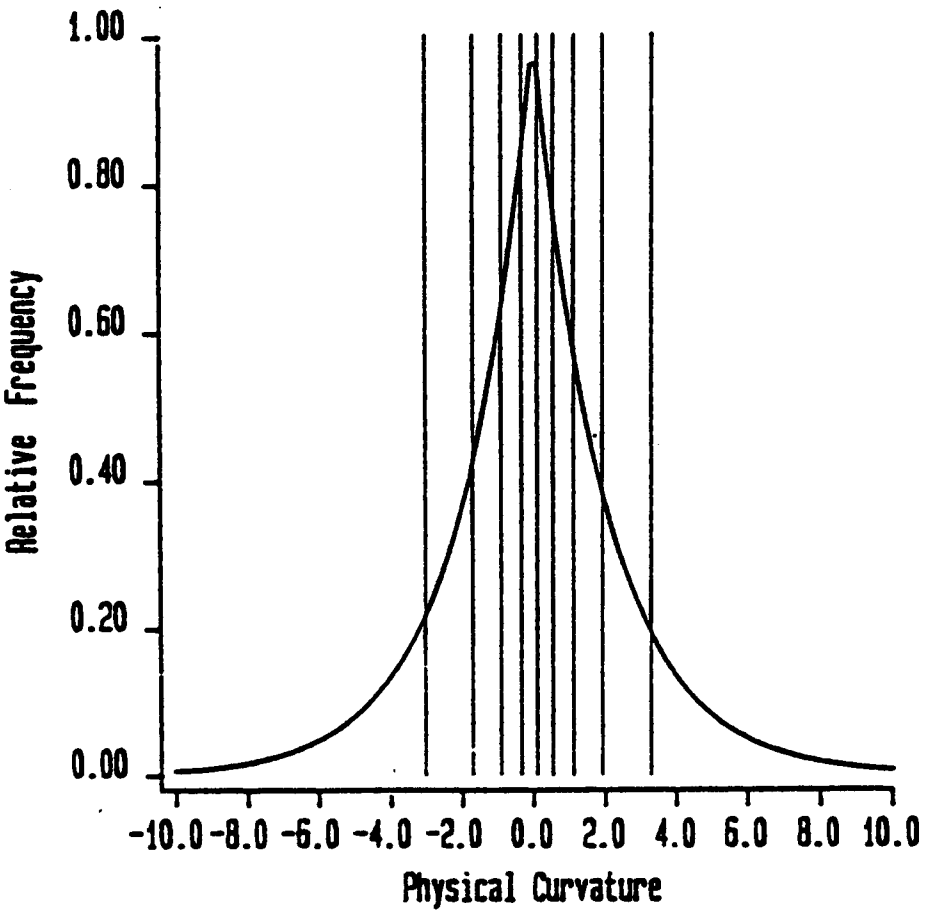
> *"The separation between one image contour and any other is not explicitly represented because it is only the separation between one image contour and its neighbour that is drawn from a statistically stable distribution. Any visual task which requires information about separations across several contours must use a representation in which luminance-edge locations are explicit or all possible pairwise separations are explicit."*
>
> (Watt, 1988, p 82)

The idea is that local polar measurements, separation and orientation, are used to build a cartesian co-ordinate system.

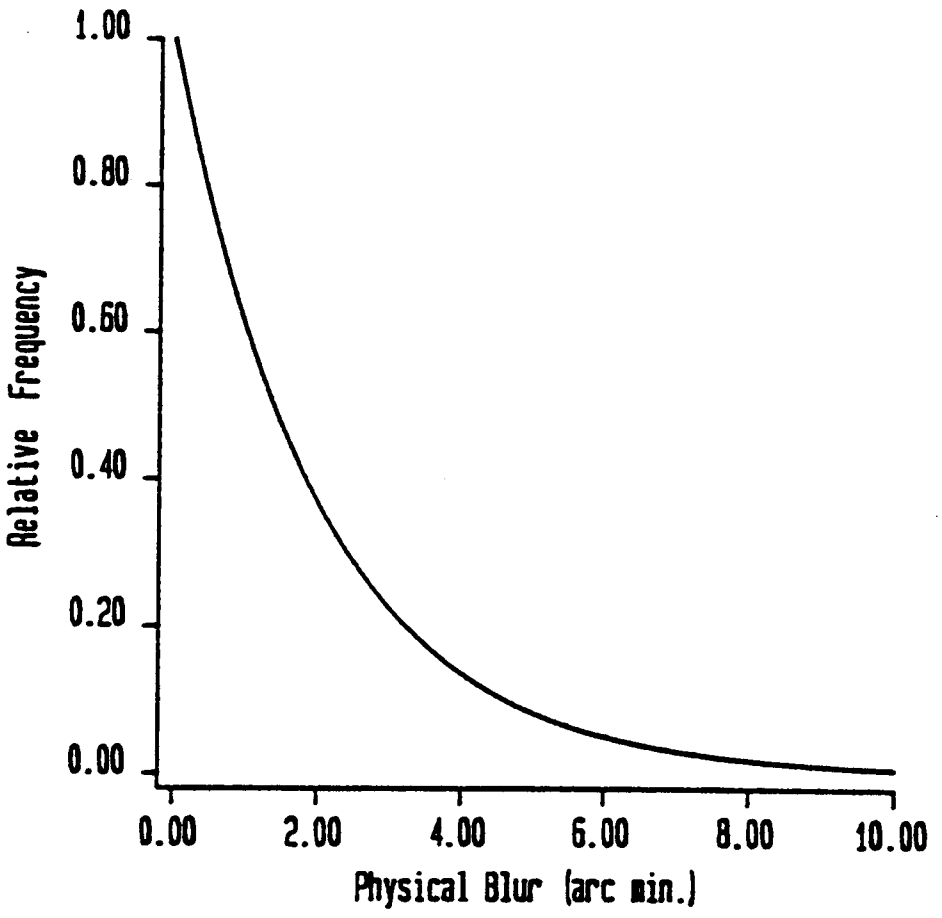
Watt considers the effect of errors in the local polar measurements on the reconstruction process and shows that these propagate through the image: they are additive. This is not the case for other measures such as blur which cause only local errors in the reconstruction process. The problem occurs for position and luminance when reconstructing from their derived measurements, separation and contrast. (See Watt, 1988, p 82-87 for a discussion of the propagation of errors.) It means that, taking any one feature (considered below as a dot), there are many "routes" to another feature in the image and as many measures of their relative positions, because the summed error along each "route" will differ. It is in tackling this problem that the role of spatial scale is introduced:

---

* The distribution is probably similar to that shown in figure 2.1 (b) for blur.



(a)



(b)

Fig. 2.3

(a) A hypothetical distribution of curvatures in images. Physical curvature is plotted as 1/radius (from highly curved in one direction through straight (zero) to highly curved in the other direction). The thin vertical lines show how an ordinal scale might appear (the area under the curve between each line is equal).
(b) A plausible frequency distribution for edge blur in images.
(from Watt, 1988)

> *"What is the solution to this difficulty? In very general terms, the vector joining each dot to each other can be regarded as a constraint on the position of each dot. Any one of these constraints could be adjusted to approach a solution for the geometry of the dot layout. However, this will have implications for the positions of many other dots and may take the system farther from a solution of some other point in the image. It is rather like changing one spring in a bedstead. There is a rather generalised technique, known as* constraint relaxation, *which has been devised to help with situations like this. But relaxation is an iterative process and it imposes a time cost: the length of time taken to reach an acceptable solution will increase with the number of dots or elements involved, even for a parallel machine."*
>
> (Watt, 1988, p 87)

One way of reducing the number of elements involved is to group them together into clusters. The issue of grouping is discussed in the next section.

2.3.2 Grouping by proximity

There are many ways of grouping features in an image, or, in more general terms, elements in a set, some of which are sensible in a given context, some of which are not. Watt shows (p92-94) that any grouping process, even a meaningless one, will speed the constraint relaxation process discussed above. (This applies when the calculation is performed, up to some arbitrary degree of accuracy, for *every* feature in the image: it does not refer to the time which might be saved by analysing only some groups of features in detail.) Watt also considers how many elements there should be within each group for optimal processing and shows that for n features there should be m groups, where $m = n^{1/2}$. The processing of these m groups would, in turn, be best achieved by treating them as elements of larger groups, and so on. So, the optimum organisation of a set of features is into a hierarchy of groups with, at each level, the ratio of number of elements to number of groups equal to the number of groups ($n{:}m = m.$).

This argument applies whatever grouping rule is chosen. Some rules for grouping, however, are more sensible than others. The optimum rule may depend on the task - for example, hunting for strawberries, colour might be a sensible basis for grouping; looking through a window, or at dust on a window, edge blur might be a sensible basis for grouping* - but a rule which has general applicability is grouping by spatial proximity. Neighbouring parts of an image are more likely to be projected from close places in space than are remote parts of an image. This raises the possibility of another time-saving strategy. The calculation of position of

---

* Grouping, including a discussion of Gestalt principles, are discussed by Marr (1976)

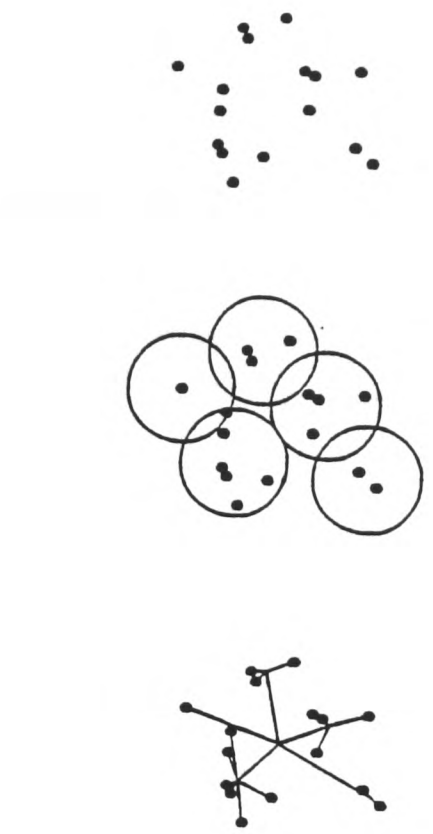
every feature within the nested hierarchy need only be done in the region of the image which is currently of interest.

> *"A grouping that was based on the image proximity of elements would be efficient because the calculated position of the group could serve as an approximate measure of position for each of its member elements. Although there would be no representation of spatial relations between elements of the same group, other than that they were grouped and therefore neighbours, all other spatial relationships would be represented."*
>
> (Watt 1988, p 94)

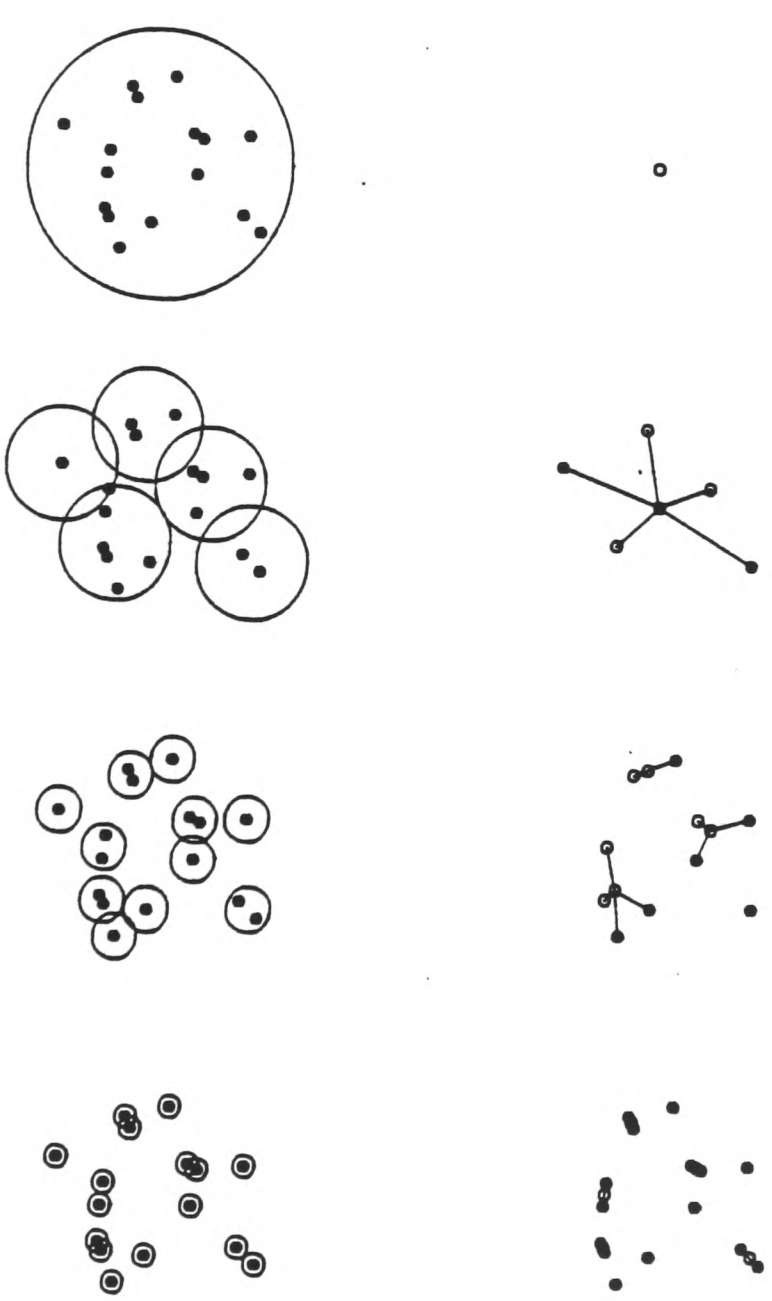
Figure 2.4 shows a set of randomly placed dots and one way in which they might be grouped into clusters. The rule that has been used is simply to group together dots falling within a given distance of one another. The positions of these five groups would be calculated first. Then the position of each dot within a group would be calculated with respect to the other dots in the same group. The number of iterations required to calculate the positions of all the dots is now 10 (for the five groups) plus 15 (for dots within the groups) i.e. a total of 25 compared to 120 iterations for the ungrouped dots. Note that a consequence of this process is that the relative position of two dots in different groups is only known via the positions of their "parent" groups, i.e. position is represented hierarchically. This is illustrated at the bottom of figure 2.4 (a).

Figure 2.4 (b) shows the same dots but now organised into a deeper nested hierarchy. In this example, as before, all the dots which fall within a specific distance of each other are grouped together. Then, within each group (circled) the same process is repeated, this time the grouping distance is half that used before. The process can be repeated at finer scales till no group contains more than one member.

Fig 2.4 (overleaf)
(a) A field of randomly placed dots is shown and beneath it a possible organisation of the dots into five groups. The position of the five groups can be calculated much more rapidly than the position of the 16 dots. At the bottom, a way in which the position of each dot might be represented hierarchically is shown.
(b) Groupings of the dots are shown on the left from the coarsest to the finest scale. On the right the relationship of each group to the coarser scale "parent" is made explicit.



(a)



(b)

Fig 2.4 (legend on previous page)

How might this grouping be carried out? The organisation shown in figure 2.4 (b) is a hierarchical structure in which grouping occurs at a range of spatial scales. But *filtering* at different spatial scales, although it often results in grouping of fine scale features within coarse scale "blobs", does not *always* do so: the grouping can be badly behaved.

An example may help illustrate this point. Figure 2.5 shows an image which has been filtered at a range of spatial scales and one way in which the filtered outputs can be combined into "blobs" which often correspond to significant features in the image. It is an example of a "scale space primal sketch". (This figure is taken from Lindeberg's (1991) thesis: "Discrete scale-space theory and the scale-space primal sketch"). The details of how the blobs are derived is not important. The main point of interest here is that blobs discovered at a fine scale, although they tend to lie within larger scale blobs, do not always do so. Another point which can be seen in the lower panels is that the outlines of the blobs overlap, i.e. one part of the image might be grouped in two coarse scale blobs.

Prazdny (1987) has examined scale-space descriptions of images and emphasises the same point, that coarse scale features do not always provide an orderly grouping of fine scale features. In extreme situations coarse and fine scale channels can carry completely unrelated signals. He says:

> *"[A coarse-to-fine] strategy requires that information in adjacent spatial frequency bands is related in the sense that they are generated by the same physical surface. In other words, the features must persist across spatial frequencies, or equivalently, the spatial averaging process must meaningfully "summarize" the fine scale detail."*
>
> (Prazdny, 1987, p92)

The grouping process described in the next section addresses both these problems: fine scale features are always grouped within coarse scale blobs; and the coarse scale representation provides a much better summary of fine scale detail than the coarse filter output alone.

2.3.3 Grouping as the outcome of MIRAGE

In this section the idea of features being "grouped together" by a filter's response is considered in more detail. The MIRAGE algorithm (Watt and Morgan, 1985), which forms the basis of Watt's hierarchical theory, is described. The grouping behaviour of individual filters and the MIRAGE signals are compared.

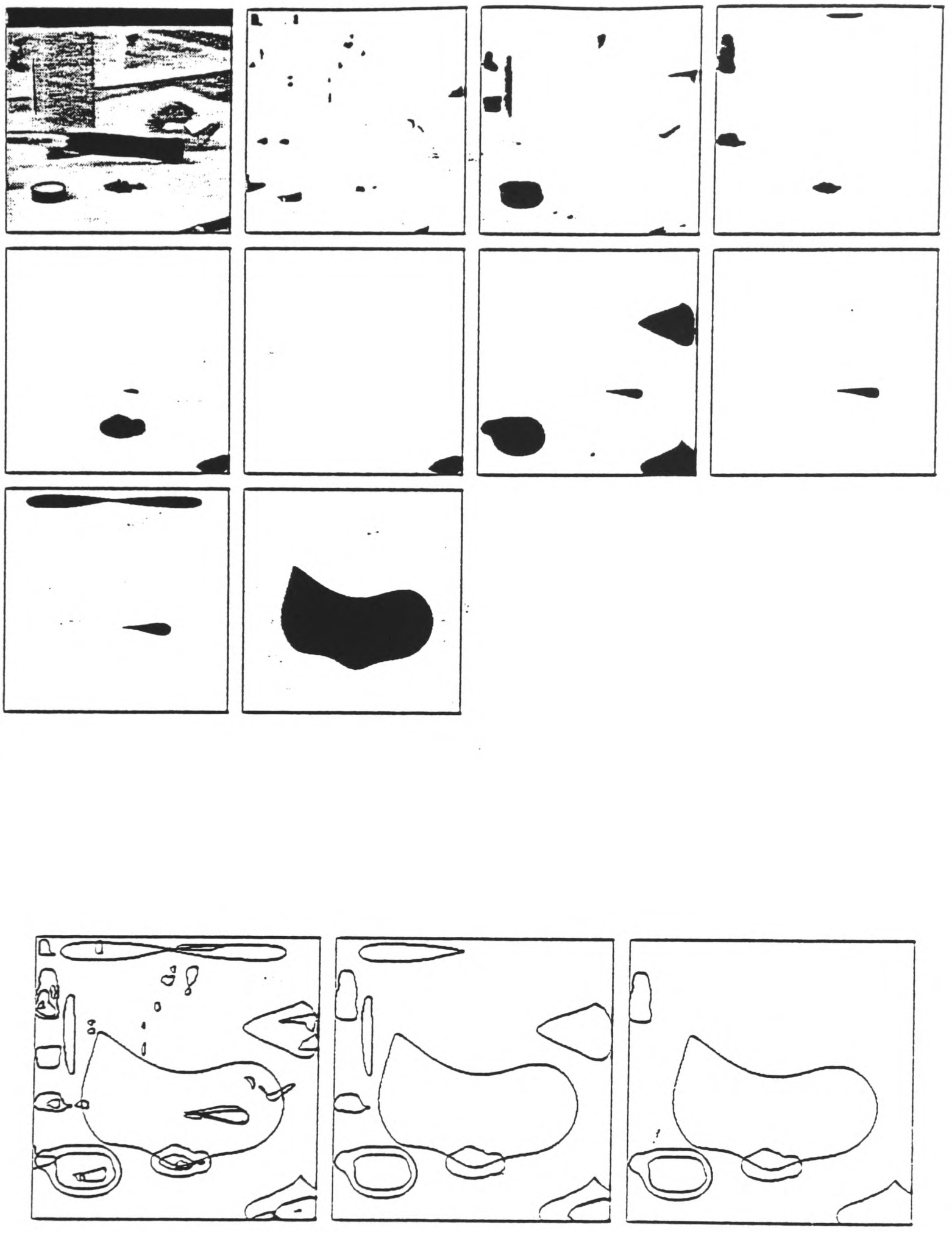


Fig 2.5

An image and its "scale-space primal sketch" as defined by Lindeberg (1991). The original image is shown at the top together with the fifty "most significant" dark blobs in the image. (The significance of a blob is given by its four dimensional volume, the four dimensions being x and y of the image, grey level and scale). The blobs are ordered according to their significance (most significant at the bottom) which roughly corresponds to scale. Below, the outline of these blobs are shown (three different thresholds of significance are shown). In this "primal sketch", fine scale blobs do not necessarily fall within large scale ones - coarse scale features group together some, but not all, fine scale features.
(from Lindeberg, 1991)

Figure 2.6 shows the same 16 dots which were considered in the previous section (shown in figure 2.4), but this time filtered at 4 spatial scales. (The filters used in this example are Laplacian, but in fact any centre-surround filters would do.) This is the same starting point as for Marr and Poggio's algorithm (except that the filters are circularly symmetric rather than elongated). Below the filter responses are shown as a 3-D plot, with positive responses shown as smooth "hills" or "islands" surrounded by a "moat" of negative response. At the top, the outline of the positive responses is shown for each filter. The positive response of the largest filter encompasses all the dots and can be said to have grouped them together. The smallest filter, on the other hand, has resolved almost all the dots, i.e. they form separate islands. At intermediate scales the dots are joined together into smaller groups, groupings which correspond very closely to those in figure 2.4 (b).

In this example, all the islands of positive response lie within a region covered by the positive response at the larger scale. It makes sense, on this basis, to consider each as an element of the group defined by the larger scale response. However, it is not *always* the case that fine scale blobs fall within coarse scale blobs, as discussed in the previous section, and an example of this is shown in figure 2.7. Here half the dots are brighter than the background, half darker. Note that the positive response of the second largest filter, for example, does not lie entirely within the response of the largest filter.

This is never true for MIRAGE signals. Because of the way the filters are combined the response at one scale *always* falls within the coarser scale response. In this sense the coarse scale response can be said to group together fine scale features.

There are four stages to MIRAGE after the filtering stage, i.e. starting from the position illustrated in figure 2.6* . The first is half wave rectification: the positive response of each filter is considered as a separate signal from the negative response. This is equivalent to having a positive or on-centre and a separate negative or off-centre filter and half-wave rectifying the output of each (which is approximately what an on-centre or off-centre retinal ganglion cell does, as it cannot signal negative responses).

---

* Note that the response gain of each filter is not the same in MIRAGE - in fact it varies in proportion to the filter size. The modulation transfer functions of each filter used in the MIRAGE algorithm are shown in Watt, 1988, p44.

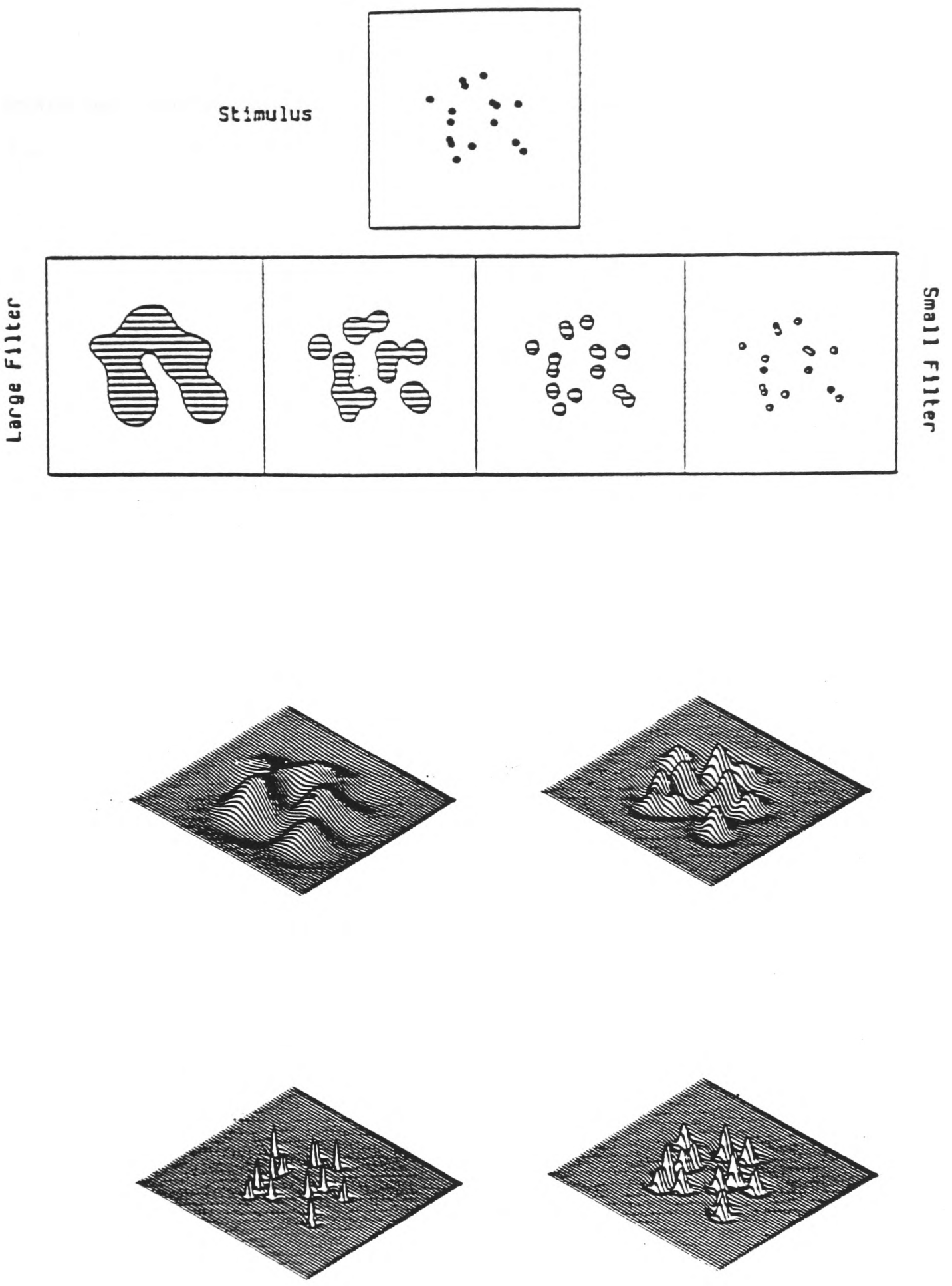


Fig 2.6

The dots of figure 2.4 are shown filtered at four scales. Above, the outline of the positive responses of the filters (i.e. zero-crossings) are shown. Below, the response of each filter is shown as a 3-D contour plot. The filters are Laplacian with space constants in the ratio of 8:4:2:1.
(from Watt, 1988)

The second and most important step is to combine the positive responses of all the filters at each point in the image, giving a so-called S+ signal, and combining the negative responses of all the filters at each point, called the S- signal.

Algebraically this can be written as:

Filtering: $R_i = F_i * I$

i.e the image function, I, is convolved with a set of filters F_i (where i = 1 to n, in figure 2.7 n is 4) giving n different filtered responses;

Rectification:

$$R_i^+ = R_i^+ \quad \text{if } R_i > 0$$
$$R_i^+ = 0 \quad \text{otherwise.}$$
$$R_i^- = -R_i^- \quad \text{if } -R_i > 0$$
$$R_i^- = 0 \quad \text{otherwise.}$$

where R_i^+ and R_i^- are the positive and negative responses of each filter respectively;

and Summation: $S_i^+ = R_1^+ + R_2^+ + R_3^+ ..... + R_n^+$

$$= \sum_{i=1}^{n} R_i^+$$

$$S_i^- = R_1^- + R_2^- + R_3^- ..... + R_n^-$$

$$= \sum_{i=1}^{n} R_i^-$$

The result of this process is shown in figure 2.8 which shows both the outline and the 3-dimensional form of the S+ and S- responses for the 16 black dots of figure 2.6.

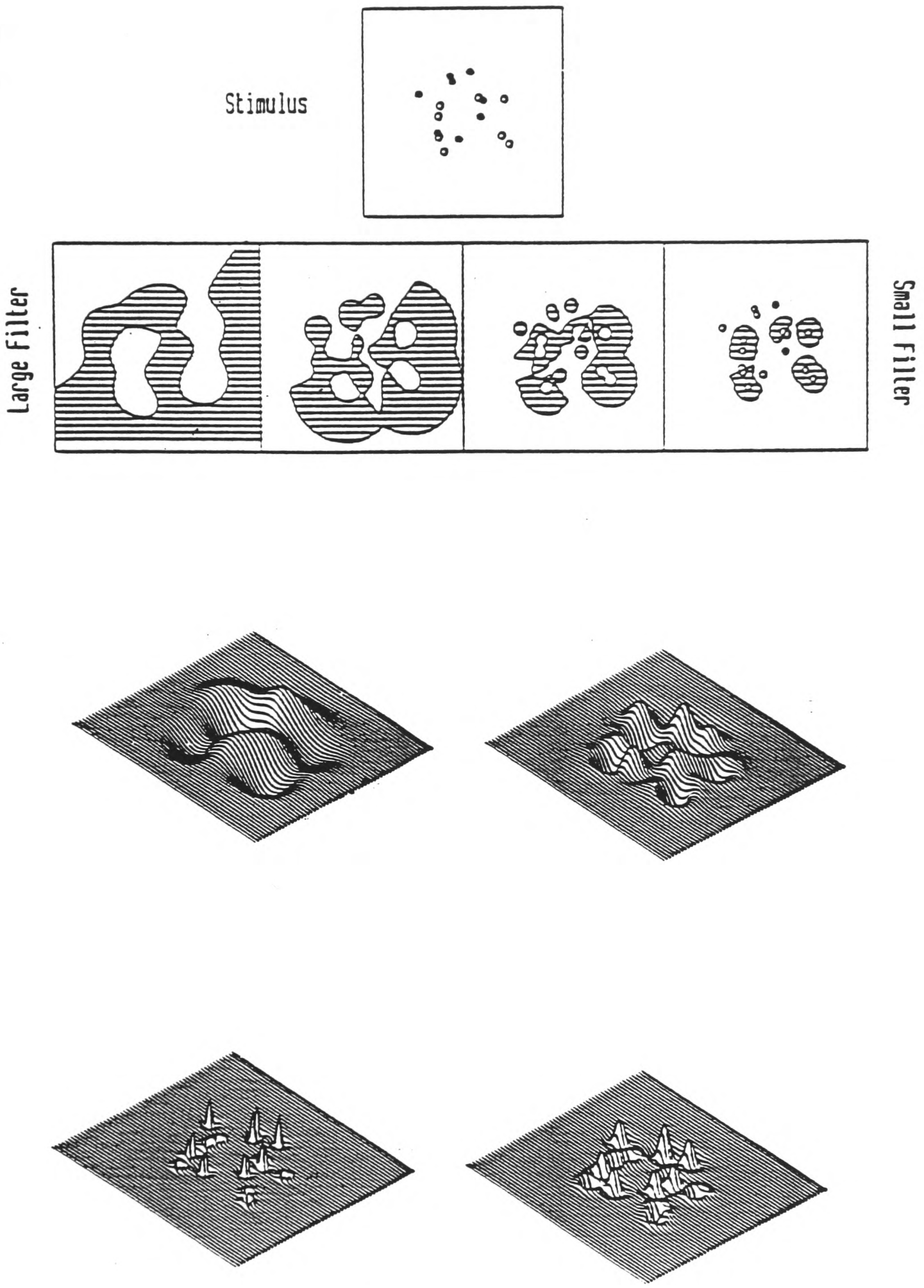


Fig 2.7

In this example half the dots are darker than the background (filled circles), half are lighter (open circles). The responses of the same filters as for figure 2.6 are shown. Unlike the case illustrated in figure 2.6, the smaller filter responses do not necessarily lie within the region of a larger filter's response.
(from Watt, 1988)

(The third and fourth stages of MIRAGE are the extraction and interpretation of primitives from the S+ and S- signals, but the grouping has already been done in the process of forming the S+ and S- signals.)

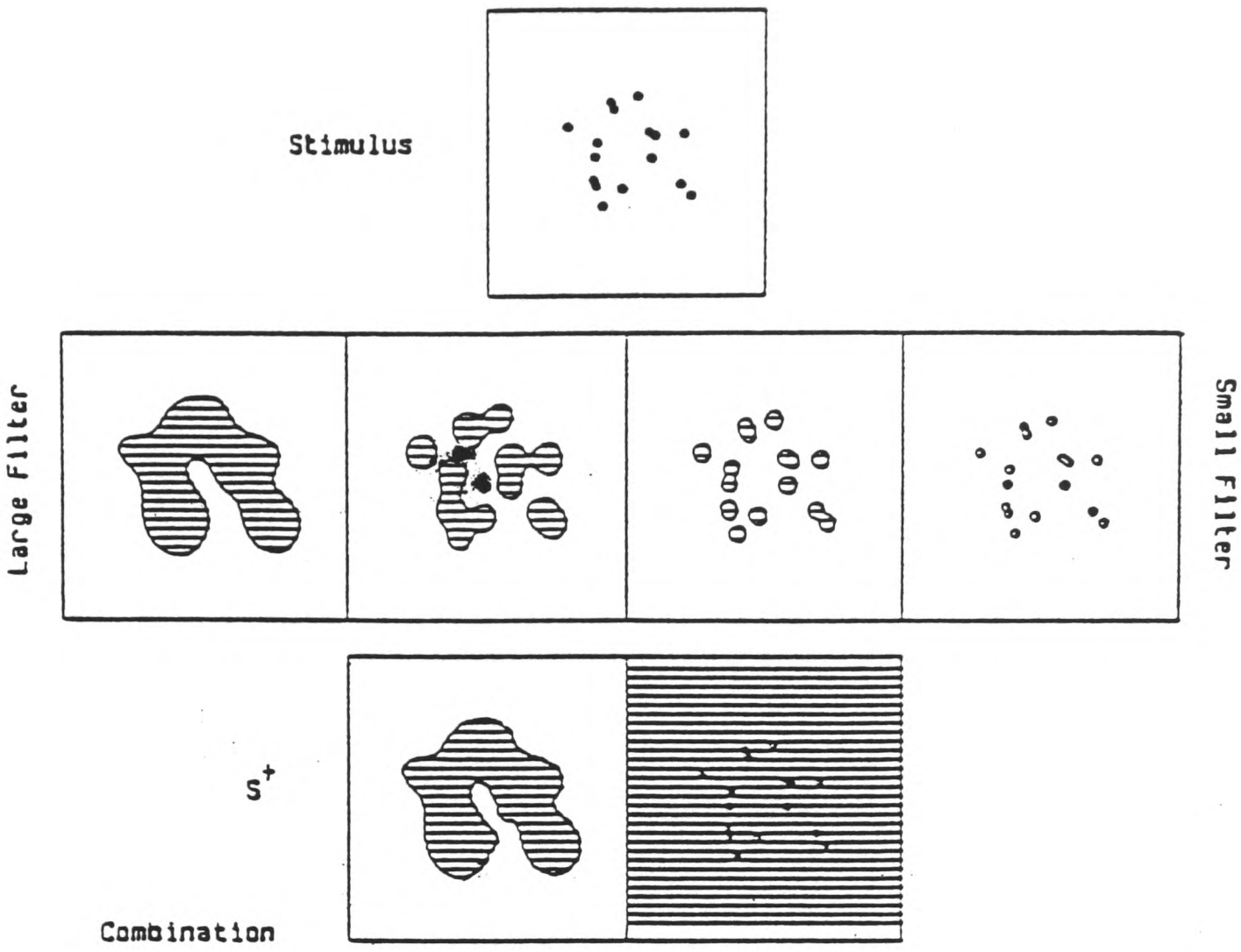
The S+ signal in figure 2.8, or at least its zero-bounded contour, is very similar to that of the largest filter and, as discussed earlier, this can be said to have grouped all the dots together. The finer scale groupings can be "discovered" by switching out the filters one by one, starting with the largest. This is shown in figure 2.9.

The S- signal, as can be seen from figure 2.8 (b), is of much lower amplitude than the S+ signal. Notice that there are small regions of zero response in the signal (in fact there are 16, corresponding to the 16 dots). These remain whatever the number of filters currently "switched in" (figure 2.9, right hand column). The "holes" derive from the smallest filter and are an important part of the MIRAGE signal.

So, MIRAGE can be said to have successfully grouped the 16 black dots at a range of scales. But, as discussed above, it is the black and white dots of figure 2.7 which are a more challenging test. Figure 2.10 shows the S+ and S- signals in response to the black and white dots and the result of switching out the filters progressively. Notice that the S3 response lies entirely within the S4 response, the S2 response within the S3 and S1 within S2 : the grouping is "well-behaved". There is no neat pattern to the groupings as there was for the black dots, but perceptually, for a collection of dots of opposite contrasts, there is no such obvious clustering either.

(Figure 2.11 shows the MIRAGE output for dots of opposite contrast in which the black and white dots are segregated spatially. In this case there *is* an obvious perceptual grouping, and MIRAGE reflects this. These two examples using dots of opposite contrasts emphasise the point that there is no difference in the role of the S+ and S- signals.)

The main point of this section can be stated simply: for the MIRAGE algorithm, fine scale blobs *always* lie within coarse scale blobs. This is so because the fine filter responses (indeed all filter responses) contribute to the coarse scale MIRAGE response.



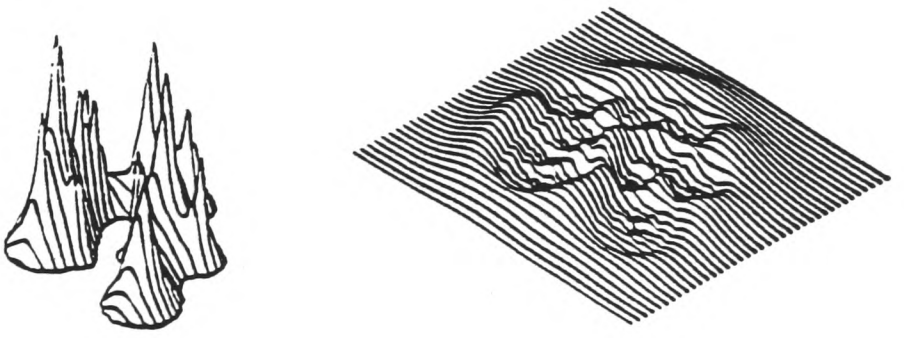


Fig 2.8

The original dots and filter responses are shown as in figure 2.6 together with the S+ and S- signals resulting from the MIRAGE combination of the filter outputs. Below, these are shown as a 3-D plot, S+ on the left, S- on the right. There are "holes" in the S- signal, one corresponding to each dot.
(from Watt, 1988)

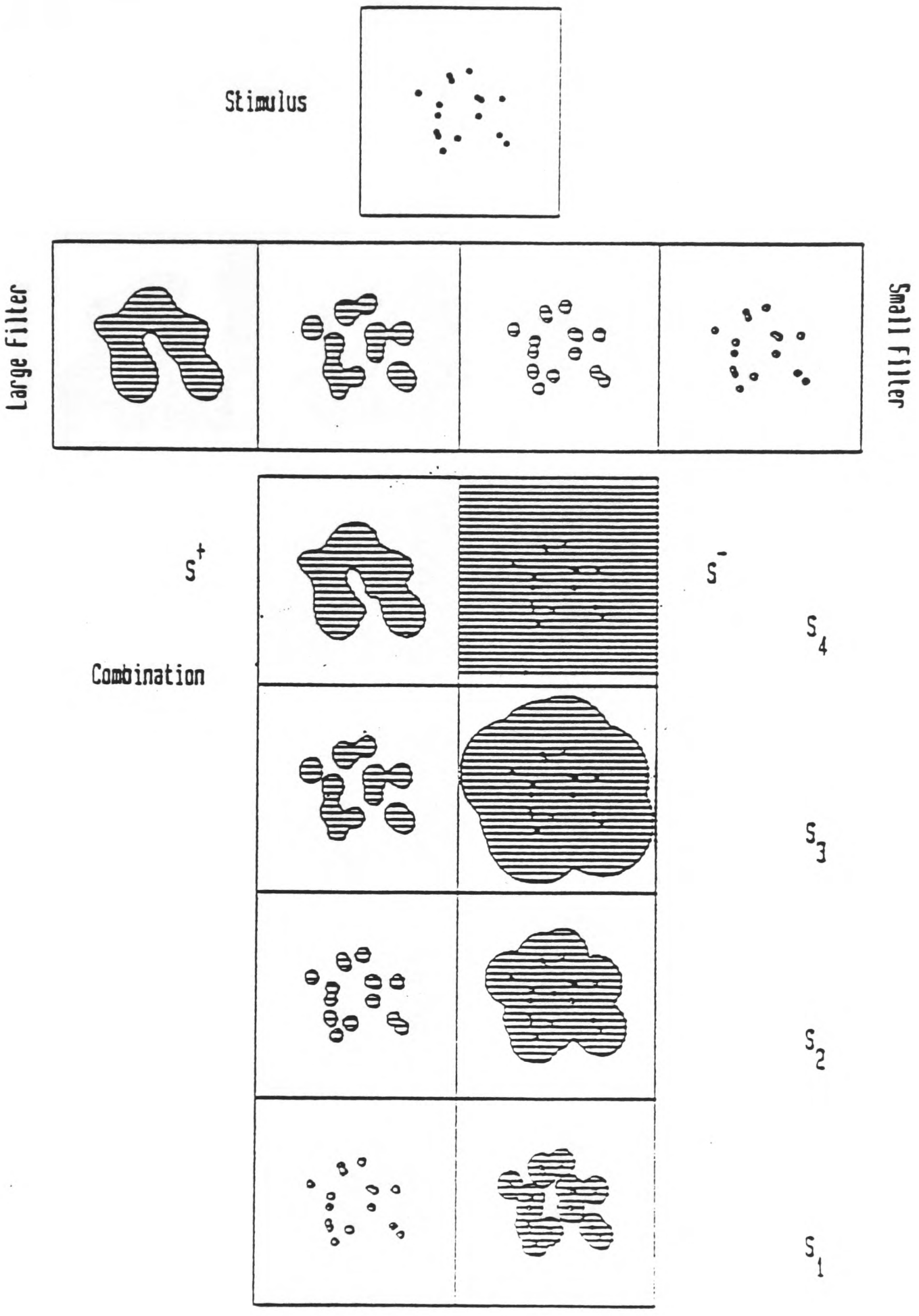


Fig 2.9

The effect of "switching out" coarse filters. The dots and filter responses are shown above, and beneath these the zero-bounded contours of the MIRAGE S signals. S4 refers to the result when all four filters are used: S3 to the result when only the smaller three are used; S2, the two smallest: S1 the smallest alone. The output of the S+ signal provides groupings of the dots similar to those shown in figure 2.4 (b).
(from Watt,1988)

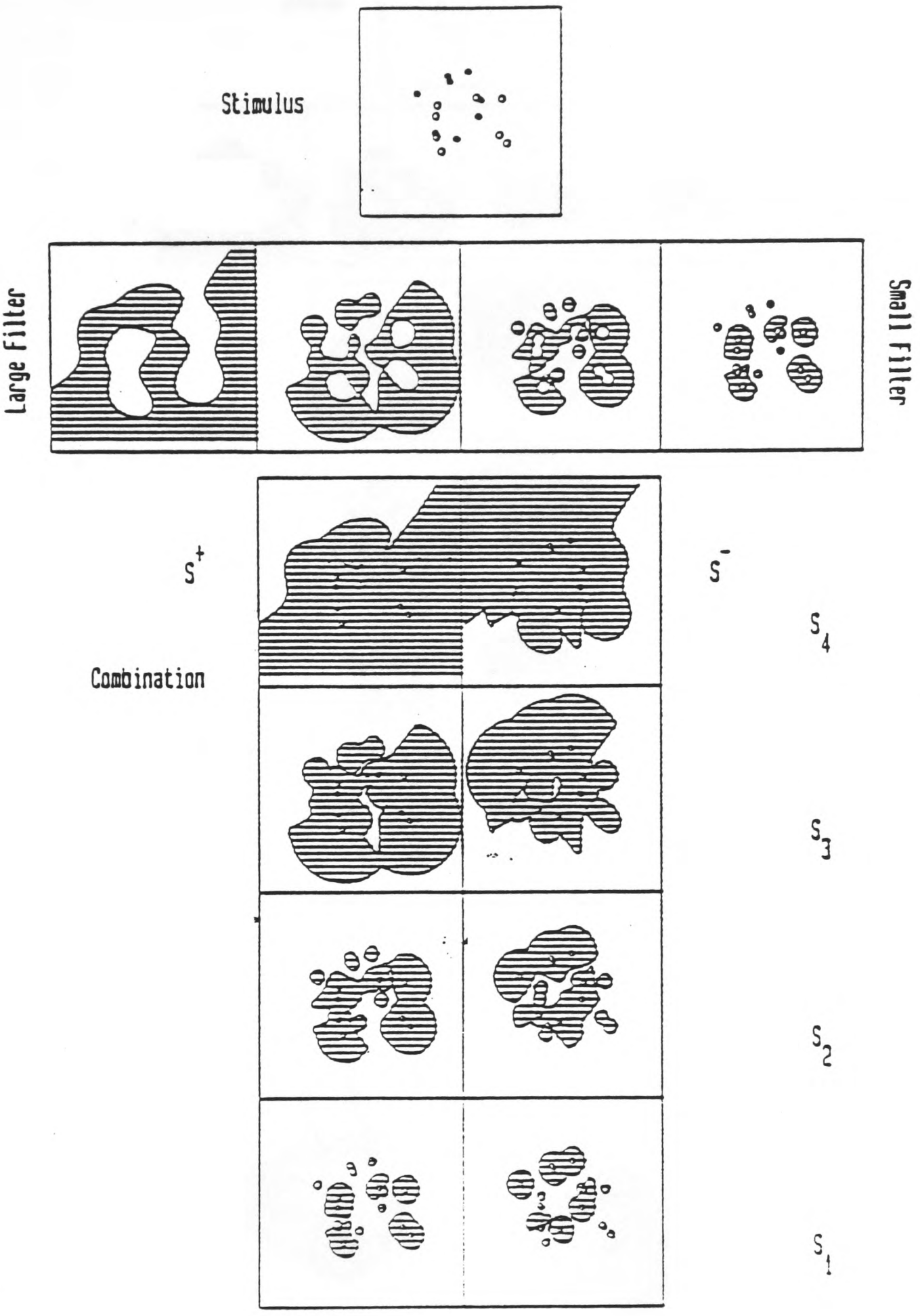


Fig 2.10

The same dots as illustrated in figure 2.7. i.e. half light and half dark, with the individual filter responses and MIRAGE S signals shown below. The grouping of the dots which was seen in figure 2.9 is disrupted but even so the MIRAGE S signals at any scale remain within the bounds of the coarser scale response - switching out a filter can never reveal a "new" area of response. Notice that there is a symmetry between the S+ and S- signals which was not present for figure 2.9.
(from Watt, 1988)

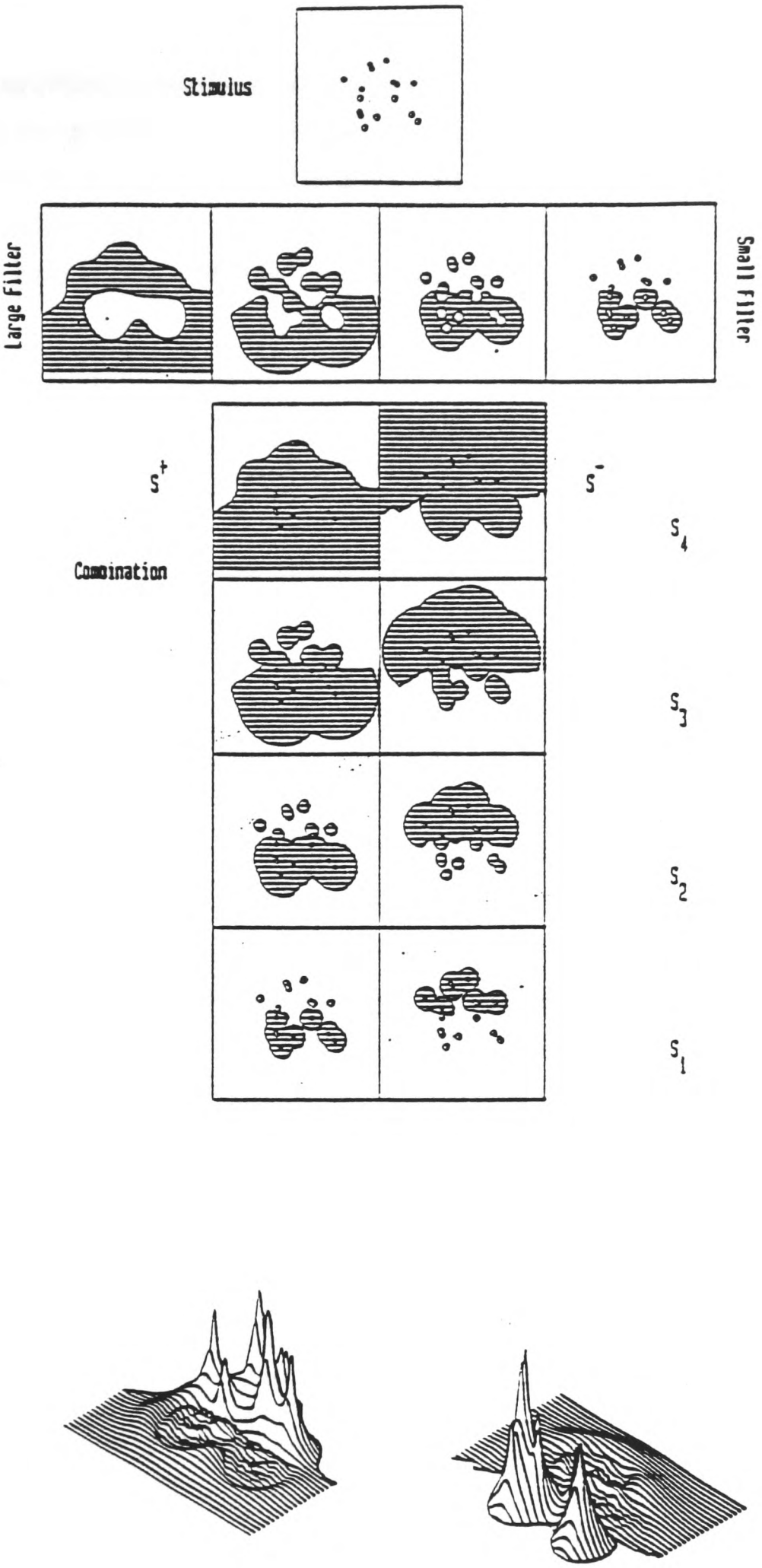


Fig 2.11

In this case the dark dots are spatially segregated from the light dots (dark dots are clustered at the top of the stimulus, light dots at the bottom) and the grouping provided by the MIRAGE responses (best seen in the 3-D plots of the S signals, below) is preserved.

(from Watt,1988)

2.3.4 Hierarchical encoding of position

A way in which the position of the 16 dots in figure 2.4 could be represented hierarchically was discussed in section 2.3.2. That example did not involve filtering (it was assumed that the separation of any pair of dots could be measured) but it can be compared with the MIRAGE representation. The two are, in fact, very similar.

At the finest scale in figure 2.8, the dots are resolved, that is, there is a separate zero-bounded region of response corresponding to all, or almost all, of the dots. These "blobs" all fall within a blob at the next (coarser) scale, which in turn fall within coarser scale blobs and so on. The position of the blobs can be represented in much the same way as the groups of dots were in figure 2.4. This requires that, first, the position of each blob with respect to every other blob at that scale is measured and second, that within a coarse scale blob the position of blobs at the next (finer) scale is known with respect to the position of the coarse scale blob.

One question is important at this stage and it has no definitive answer in MIRAGE: How is the position of a blob defined? In 1-D there *is* a precise answer, it is the position of the "centre of mass" of the response, i.e. its centroid. (The logic is to consider the blob as a distribution of activity, a "zero-bounded distribution", made up of independent samples. Hence the zero-bounded distribution can be described, as any distribution can be described, in terms of its central moments: n (number of samples or "mass"), mean (or centroid) and standard deviation. These are illustrated in figure 2.12). In 2-D the situation is more complex. Watt and Morgan (1985) say of this problem:

> *"In theory MIRAGE could be extended to two dimensions in several different ways, for example by using elongated (orientationally selective) filters and thereby reducing the two dimensions to a set of one-dimensions... Alternatively, circularly symmetric filters could be used, with centroid extraction process being orientationally selective."*
>
> (Watt and Morgan, 1985, p1668)

It is the second of these alternatives which has been chosen for the purpose of modelling in this thesis since for stereo (due to the horizontal separation of the eyes) there is one dimension more appropriate than any other (horizontal) for the extraction of centroids.

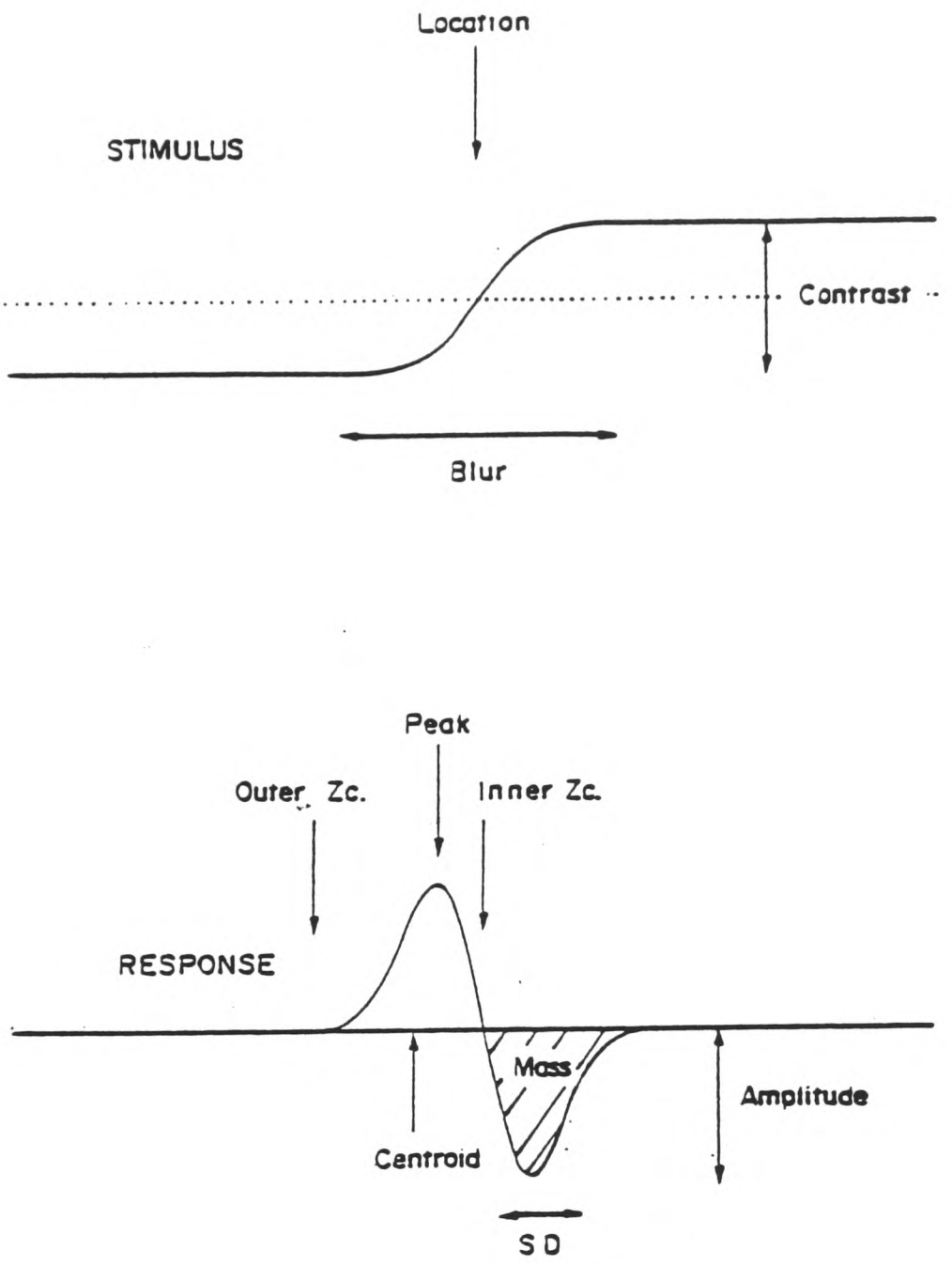


Fig 2.12

The (1-D) luminance profile of an edge. Mean luminance is shown as a dotted line. The edge can be defined by its blur, contrast and location, as illustrated. Below, the response of a triphasic filter (e.g. a Laplacian) is shown, together with potential primitives characterising the response.

(from Watt and Morgan, 1985)

2.3.5 Order and rate of processing.

The logical order in which to calculate the position of the blobs (by the constraint relaxation process discussed in section 2.3.1) is from coarse-to-fine. There are some situations in which prior knowledge about the image and the task is available, such as reading text, in which case analysis might be restricted to a particular scale. This, however, is likely to be the exception and, in general, there are several advantages to a coarse-to-fine order of processing. First, a useful approximation is available straight away, one which can be refined if time allows. Second, the coarse scale representation can be used to guide further processing, indeed this need only be carried out in the region of the image which is of interest. Third, as filter size is reduced new blobs "appear" and their positions must be calculated. If this calculation were the rate limiting step in the process then it would determine the rate at which filter size was reduced. For new blobs to appear at a constant rate, filter size must change with time according to a power law (i.e linear when plotted on log-log axes), a prediction discussed in section 2.4.3.

2.3.6 "Selective attention"

It is important to realise that much more is represented by a coarse scale MIRAGE blob than just the output of the largest filter. The "holes" in the opposite S response convey information about the fine scale texture in the region of the blob (as in figure 2.8, where there are 16 holes corresponding to the 16 dots). Watt is not precise about exactly what information might be available to the visual system about these holes but it would include some idea of the number of holes ("one","two","many", perhaps), their approximate orientation (e.g. "all roughly vertical"), the presence of corners, intersections, etc. This information is all statistical, in contrast to the explicit spatial information about the centroid, mass, orientation and width of a blob. So, before ungrouping there *is* information about fine scale features but it is only a statistical description of the holes bounded by a coarse scale blob. After ungrouping, the information about fine scale features is explicit and spatial.

The initial statistical information might be sufficient to perform many visual tasks and the advantage is that it is rapidly available. In contrast, fine scale spatial analysis, or ungrouping, takes a lot of computation even when it is carried out in restricted areas of the image and this takes time. Watt's suggestion is that this may be what is happening when "subjects experience the need to use selective attention" to perform a task (Watt, 1988, p139). Evidence for this idea and data on the time course of the ungrouping process are considered in the final section of this chapter.

2.4 Psychophysical evidence

In this section some of the psychophysical evidence is considered concerning three crucial aspects of the MIRAGE theory: first, the nature of the primitives used, either for stereopsis or spatial judgements; second, the combination of filter outputs and whether each can be accessed independently and third, the "switching out" of coarse filters and the tasks which can be done before and after this process is complete.

2.4.1 Primitives

Figure 2.12 illustrates the response of a single filter to a luminance step edge. In many ways an edge is the simplest stimulus and any image can be described as made up of edges. A line, for example, can be considered as two edges "back-to-back". Three attributes describe this edge completely. Two are local measures, its contrast and its blur, one is relative, its location. The filter used in figure 2.12 is a Laplacian of a Gaussian but any triphasic, balanced filter such as a DOG would give a similar response. The zero-crossing, peak and trough are marked along with the central moments, i.e. mass, centroid and standard deviation.

The location of the zero-crossing coincides with the position of the edge (in the case illustrated here). The point midway between the peak and trough also signals the edge position, as does the centroid of centroids. The argument that zero-crossings are the simplest signal of edge location does not continue to be the case when the response to a line (impulse) is considered. In this case the zero-crossings are either side of the line by an amount that depends on the scale of the filter and some heuristic is needed to give the line's location (Marr and Hildreth (1980)). The centroid of centroids gives the location of both a line and an edge.

The various candidate spatial primitives behave differently when noise is added to the output of the filter. Watt and Morgan (1984) argue that the central moments are the most robust in the face of low signal-to-noise ratios. Figure 2.13 shows their analysis of the behaviour of each of the features as noise is added. The standard error of estimates of the edge position is plotted against signal-to-noise ratio. Some criterion for the position of the edge must be chosen in the case of each primitive. The highest peak, the centroid of the greatest mass and the zero-crossing with the steepest slope were chosen by Watt and Morgan (1984). The results of this

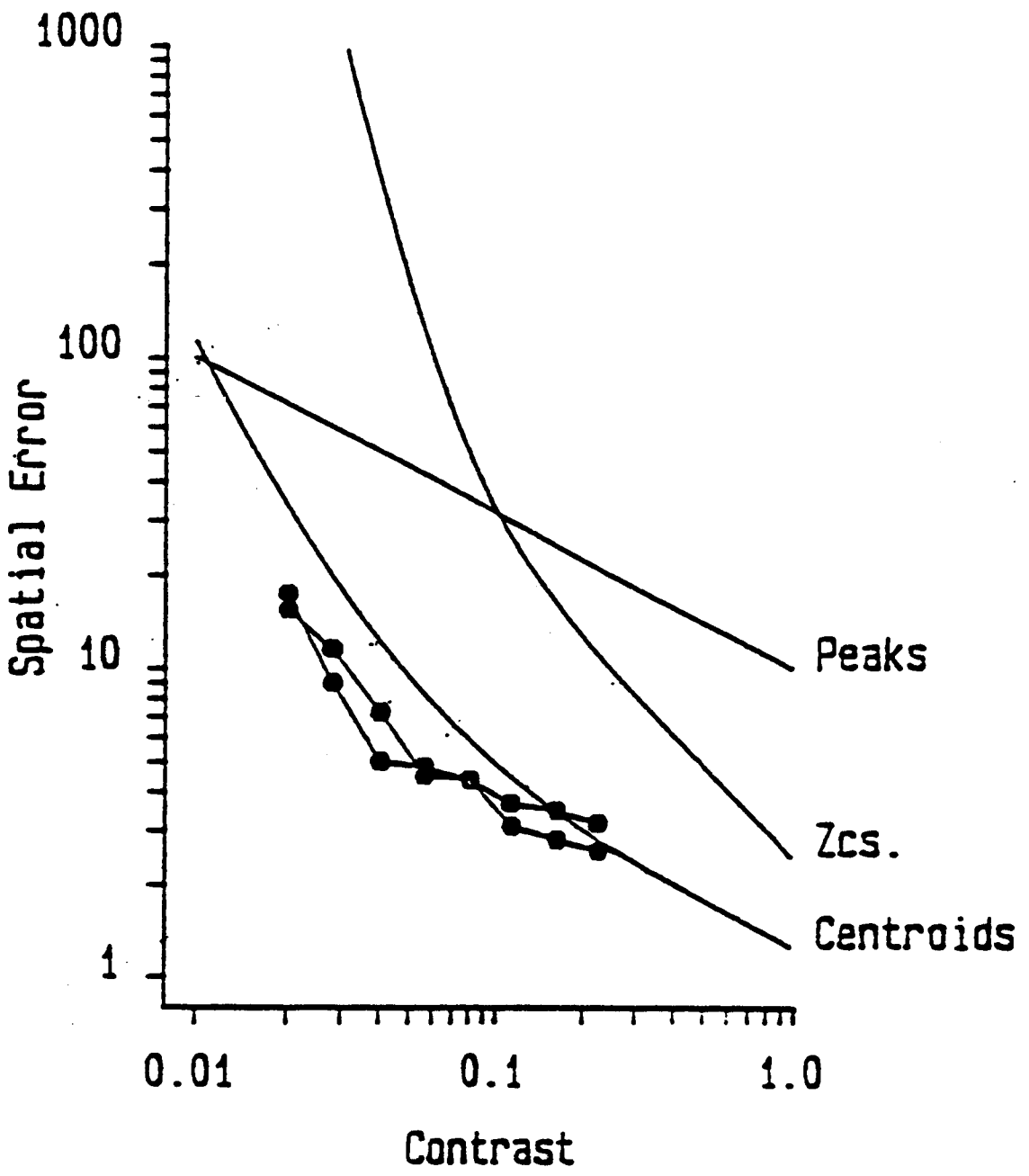


Fig 2.13

The spatial error (standard deviation) of the estimated location of an edge (as defined by one of three primitives) is plotted against the contrast of the edge (which, for a fixed noise level is equivalent to the signal-to-noise amplitude ratio). Data for the vernier alignment of two edges is plotted for two subjects (taken from a study reported by Watt and Morgan, 1984). The horizontal position of the data relative to the model is determined by assuming the threshold for contrast detection is 0.01 (1%). The vertical position of the data is arbitrary. Centroids provide the best fit to the experimental data.

(from Watt, 1988)

simulation were compared to location thresholds for a vernier task using an edge stimulus at a range of contrasts. (The logic of this comparison is that contrast, i.e. peak-to-trough amplitude can be taken as the magnitude of the signal while the magnitude of the intrinsic noise in the visual system is considered to remain constant.) Over a large range of contrasts, location accuracy varied with the square root of stimulus contrast (a slope of -0.5 when threshold is plotted against contrast on log-log axes). At very low contrasts, Watt and Morgan found that performance depended on the type of edge used. For a band-pass edge the square root relationship held right down to the lowest contrasts. For a low-pass edge or, even more so, a high-pass edge, the signal is much more localised and the effect of noise more severe. For these stimuli thresholds rose much more steeply as contrast was reduced. This fits, as shown in figure 2.13, the predictions for centroids.

A very similar analysis was carried out by Legge and Gu (1989). The only difference in their simulation was that they chose as their criterion for the location of an edge signalled by zero-crossings to be the mean of the distribution of zero-crossings. This they show (analytically) to vary inversely with signal amplitude, predicting a slope of -1 on log-log axes. Legge and Gu used sine wave stimuli and their task was a stereoacuity judgement. They found that acuity varied according to the square root of contrast down to the lowest contrasts, i.e. the same result as Watt and Morgan found for a band-pass edge. They chose peaks as the best fitting primitive but, as discussed above, they did not use the most appropriate stimulus for distinguishing between peaks and centroids. In fact, in a later study, Gu and Legge (1991) put forward centroids as the primitive which best fits their data. These are the primitives which are used for modelling in later chapters.

A similar debate exists concerning the relevant primitive for judgements of contrast (mass, peak height, zero-crossing slope: see Watt, 1988 (p47-49)) but the issue is not relevant to the encoding of position and is not considered here.

2.4.2 Combination of filter outputs.

The evidence discussed in this section relates to experiments carried out with unlimited exposures and hence, in terms of the "switching out" of coarse filters discussed above, corresponds to the final state of the filters. Watt (1987) suggests that in this case the four finest filters (space constants 2.83, 1.42, 0.71 and 0.35 arcmin) remain combined. (The experiments were carried out in foveal vision and filter sizes apply to the fovea only.)

Support for the idea that the smallest filters may remain combined comes from two independent experiments which were used by Watt and Morgan to derive the minimum internal blur for judgements of spatial position (Watt and Morgan 84) and relative edge blur (Watt and Morgan 83). Both of these experiments pointed to a value of about 2.8 arcmin (figure 2.14). On the other hand, results of an experiment in which the task was one of resolution (Watt, 1987) were best modelled by a filter of about 0.35 arcmin. This fits with the characteristics of the MIRAGE signal: statistical information derived from the smallest filter (the "holes") would distinguish a dotted from an unbroken line (the resolution task) but information about blur and position are not available at this scale. Thus there is a "gap" in estimates of the minimum blur in the visual system depending on the task being performed. Further evidence for this gap is given in Watt (1987). In that experiment, curvature, line length and orientation discrimination thresholds for long exposures (e.g. 1-2 seconds) are all modelled by a filter of about 3 arcmin, i.e. an order of magnitude larger than the scale for a resolution task. (Details of the experiment are discussed in the next section).

There are other experiments which support the idea that the smallest filters remain combined (whatever the exposure duration), many of which concern the detection of patterns at low contrasts. Because they relate to the measurement of mass rather than position, these are not considered here in detail (see Watt, 1988, p42-50 for a full discussion). One example is an experiment by Henning, Hertz and Broadbent (1975) who showed that an amplitude-modulated high spatial frequency sine wave could affect the detection of a target with a very different spatial frequency. This was not expected on the basis of a model of independent spatial frequency tuned channels and the authors concluded that some "non-linear process" must take place to account for the interaction. MIRAGE is a non-linear process and Watt (1988) shows that it can account for Henning et.al's results as well as the results of a similar experiment carried out by Nachmias and Rogowitz (1983) in which detection thresholds were found to vary with the relative phase of the target and mask.

2.4.3 A dynamic MIRAGE

Quantitative evidence for the change in scale of analysis with exposure duration comes from a paper by Watt (1987). Thresholds for a range of hyperacuity judgements (stereoacuity, curvature, line length and orientation discrimination) were measured over a range of exposure durations (15 to 1000 ms) and, for three

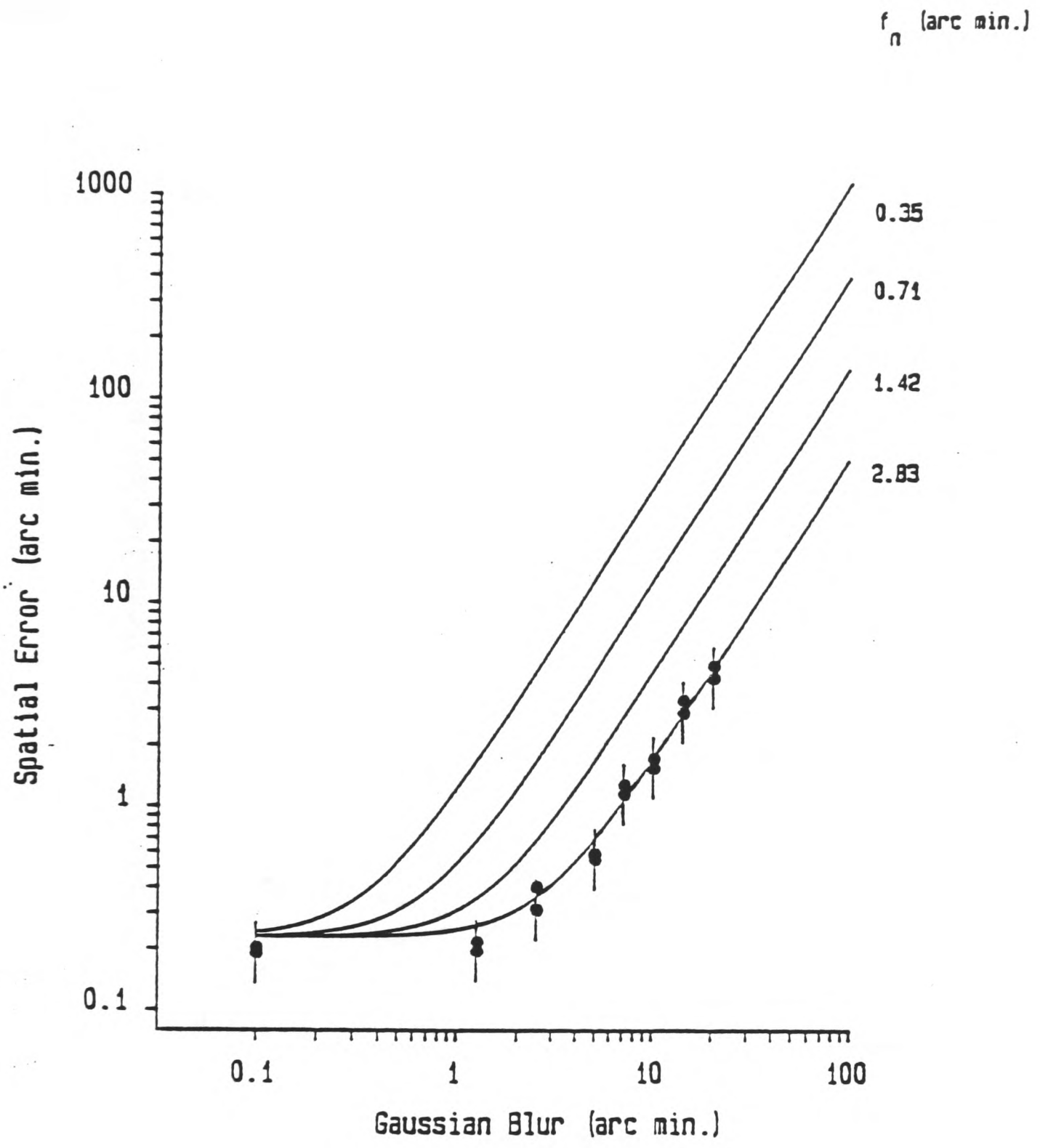


Fig 2.14 (a)

The result of an experiment by Watt and Morgan (1984) in which vernier thresholds for a gaussian blurred edge were measured as a function of blur. Predictions for four different sized filters are shown. (The "heel" of the curve in each case corresponds to the size of the filter.)

(from Watt, 1988)

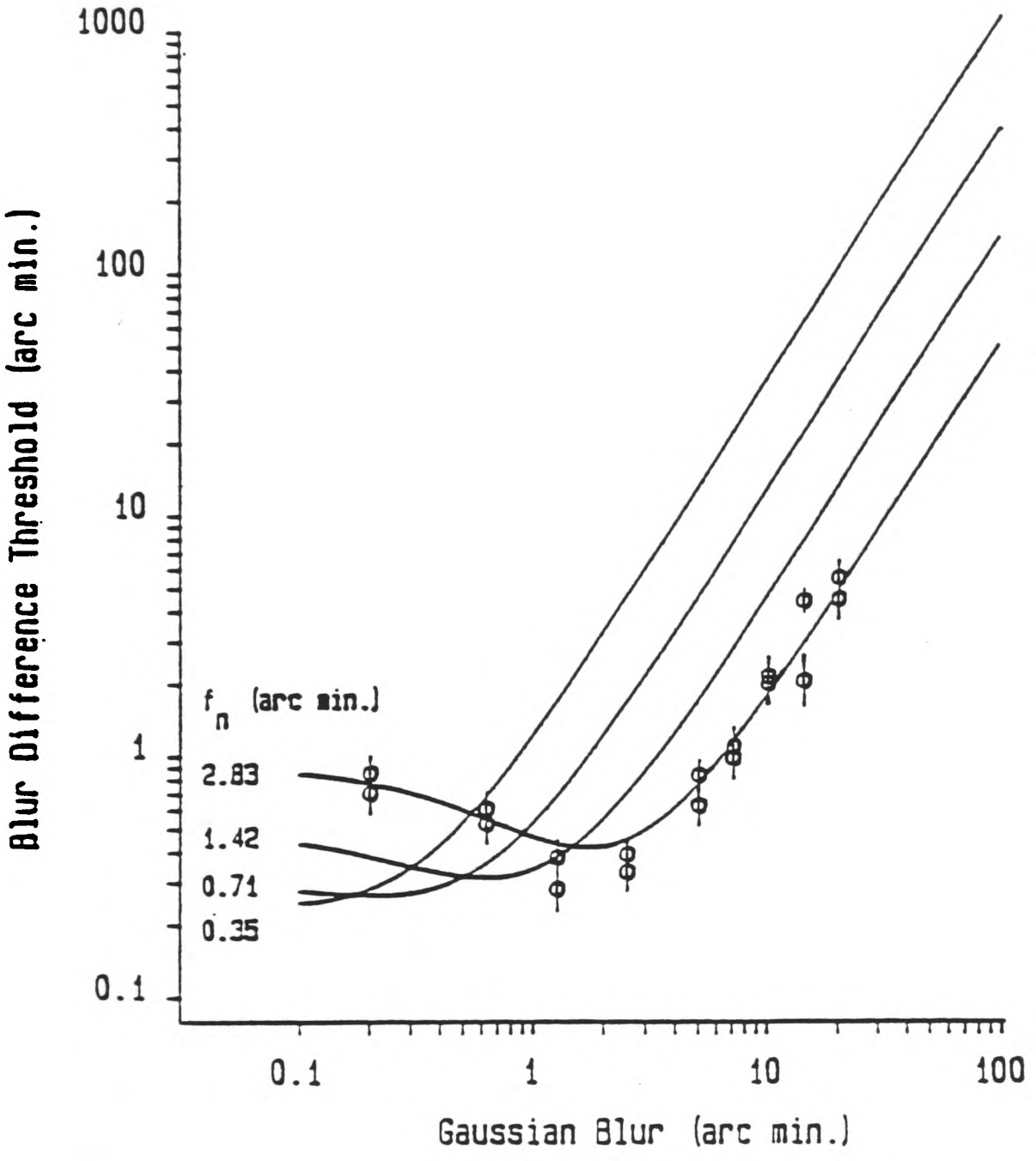


Fig 2.14 (b)

Results of an experiment by Watt and Morgan (1983) in which thresholds for discriminating which of two edges was the more blurred were measured as a function of blur. Predictions for four filter sizes are shown.

(from Watt, 1988)

of these tasks, models were derived from which the size of the largest filter could be deduced. One of the models is discussed here as an example of the method used.

The task was to judge the orientation of a line tilted to the left or right of vertical. The minimum orientation for which subjects could perform this task reliably was measured as a function of exposure duration and line length. The model considered the effect that internal blur might have on the stimulus, effectively making it into an ellipse. The reliability of correctly identifying the orientation of an ellipse in the presence of noise depends on the ratio of the lengths of its two principal axes (in the limit, for a ratio of 1 (a circle), the threshold is infinite). This ratio depends both on the length of the line and the size of the filter. Data for this experiment are shown in figure 2.15 along with the model for three filter sizes. Results for exposure durations of 35, 100 and 500 ms fit well with the model for filter sizes of 64, 16 and 4 arcmin respectively. In this way, the scale of analysis at different exposure durations could be deduced. Figure 2.16 shows the result. The agreement between models of filter size derived from curvature, line length and orientation discrimination thresholds is very close. In each case filter size reduces inversely with exposure duration (the slope is about -1). The filter used for the resolution task (discussed above) shows a very different pattern, varying very little over the range of exposure durations tested. Note that this result is not inconsistent with the findings of Robson (1966) that contrast sensitivity for high spatial frequencies is poor at high temporal frequencies. The detection of a low contrast target is likely to be improved by a process which averages over time or space or both. The resolution of dots in a high contrast stimulus, on the other hand, is not necessarily affected by short exposure durations (or high noise level) in the same way. Certainly the two tasks, resolution and the detection of a grating, would be modelled differently in MIRAGE.

In the previous section (2.4.2), evidence was discussed which indicated a difference between the minimum internal blur or scale of analysis for some tasks than for others. This experiment extends that observation and suggests that the difference is more marked at short exposures. In terms of the MIRAGE algorithm this can be interpreted as a wider *range* of filters contributing to the S signals: the size of the smallest filter stays roughly constant (as the resolution data shows) but the coarsest filter is much larger at short exposures.

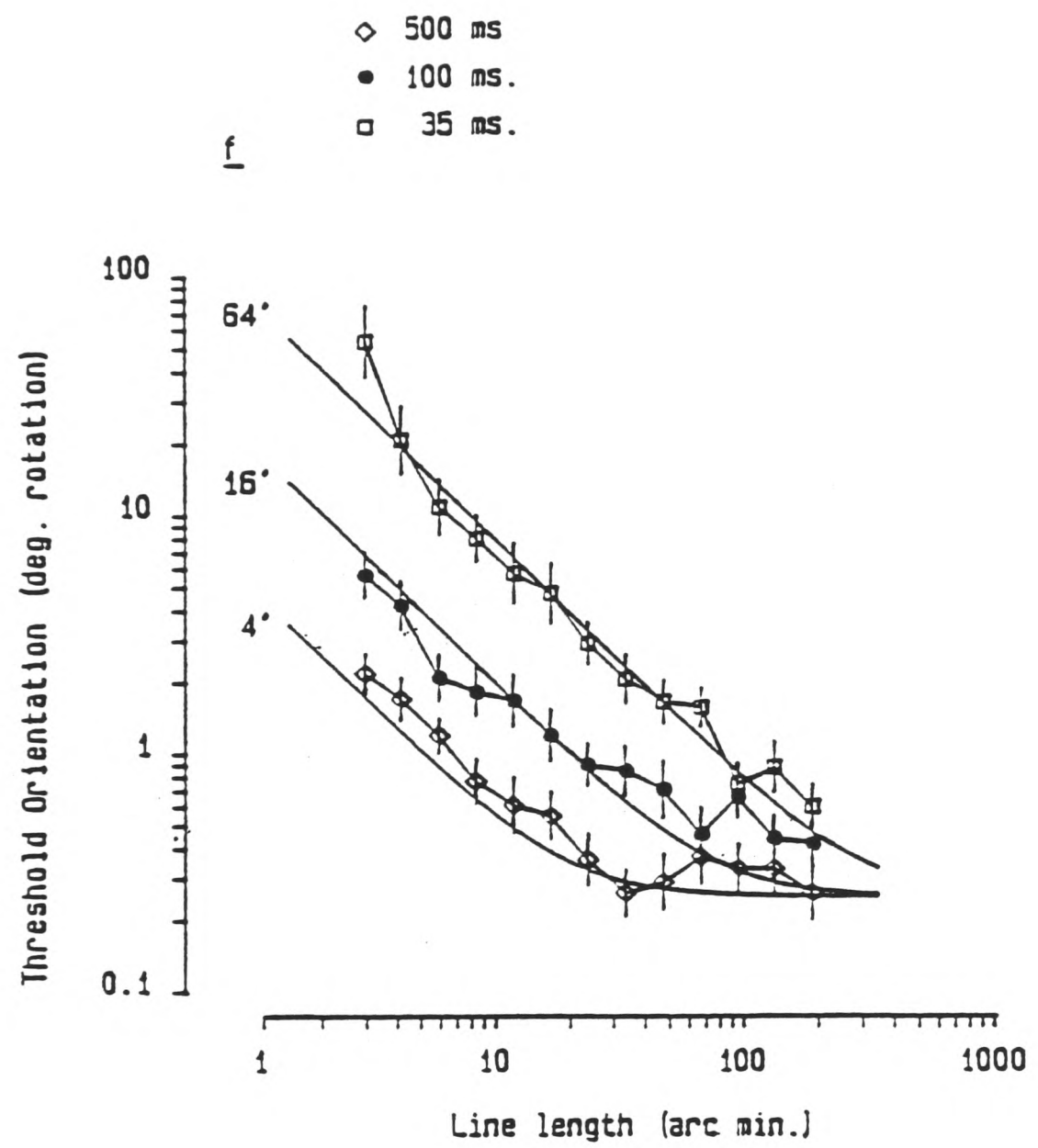


Fig 2.15

Results from an experiment by Watt (1987). Orientation thresholds are plotted as a function of line length for three exposure durations. The continuous lines are models of orientation thresholds for three different filter sizes (see text).

(from Watt, 1988)

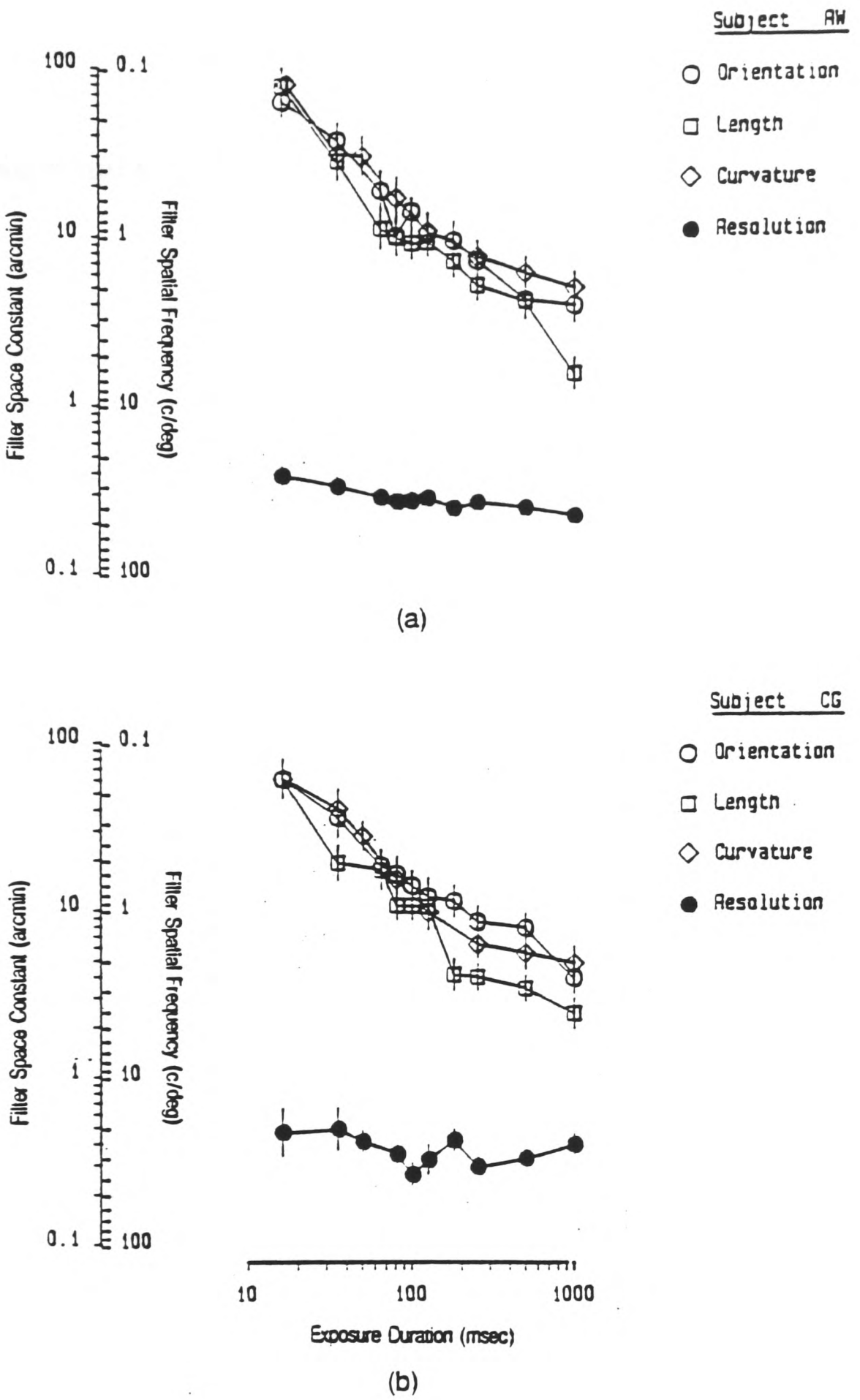


Fig 2.16

A model of the effective filter size used by the visual system to perform different tasks at a range of exposure durations (up to one second) after the onset of a stimulus. Data are taken from an experiment by Watt (1987). A set of experiments was carried out for three hyperacuity tasks: a curvature discrimination task, an length discrimination task and an orientation discrimination task (see results shown in figure 2.15). A model of filter size was derived for each and used to determine the values shown in this plot. The three models co-incide closely. A resolution task (distinguishing a dotted line from a continuous one) gave very different results: the modelled filter size for this task changes very little over the range of exposure durations tested. Data for two subjects are shown in (a) and (b).

(from Watt, 1987)

2.5 Summary

This chapter began with a discussion of "neural shift" models of stereoscopic matching as an alternative to Marr and Poggio's (1979) coarse-to-fine algorithm. Although these models avoid the need for eye movements, they do so by "brute force", i.e. by proposing that a cell exists to signal every possible disparity at every possible retinal location. Marr and Poggio's model is more economical because one group of cells measuring fine scale disparities is "moved around" the scene using eye movements. In a sense this model is hierarchical because, if the $2^1/_2$-D sketch is considered as a record of coarse scale disparities, fine scale disparities are all measured relative to coarse scale features. However, fine scale disparities are always measured as absolute disparities (i.e. with respect to the fixation plane) and disparities are recorded in the $2^1/_2$-D sketch at one (fine) scale, so it cannot be said to be truly hierarchical. Certainly it was not described by Marr and Nishihara (1978) or Marr and Poggio (1979) in those terms.

For the disparities of features to be measured relative to coarser scale features rather than relative to the fixation plane requires that their *position* be defined in a relative way as well. Such a proposal has already been made (Watt, 1987 and 1988) and the rest of the chapter was devoted to a discussion of this theory.

The most important aspect of a hierarchical model of position is that a new map of the image is "built up" from information about the orientation and separation of features. The "local sign" information about a feature in the image (its retinal location) is completely lost (although theoretically it could be re-computed). It was emphasised that not all coarse-to-fine systems are necessarily hierarchical. For instance the output of a coarse filter does not necessarily provide an obvious grouping for all fine scale features, whereas in the MIRAGE algorithm the relationship of each fine scale blob to its coarse scale "parent" is made explicit.

In the next chapter, some of the consequences of a hierarchical system of encoding disparity are examined.

CHAPTER 3

3.1 A synthesis: Marr, Poggio and Watt

In the previous two chapters the coarse-to-fine algorithm for stereo matching put forward by Marr and Poggio (1979) and the coarse-to-fine method of encoding position put forward by Watt (1988) have been discussed. In this chapter, the two theories are brought together.

The hypothesis is simple to state: it is that the disparity of a feature is based on its "hierarchical" position in the left and right eyes' images rather than its location on the two retinae. The implications of this proposal are examined in the next section in the context of three examples: the disparity of features on a fronto-parallel surface; on a slanted surface; and on a curved surface.

3.2 "Hierarchical" disparity

3.2.1 A fronto-parallel surface

The simplest example to consider which illustrates the effect of encoding disparity hierarchically is a fronto-parallel surface off the fixation plane. This is situation shown in figure 3.1 . The representation of coarse scale features is very similar to that in Marr and Poggio's scheme. It is the representation of fine scale features which displays the different characteristics of a hierarchical model.

Figure 3.1 depicts an observer looking at two fronto-parallel surfaces, one nearer to the observer than the other. One "blob" on each surface is illustrated. The coarse outline of these blobs, as seen by the two eyes, is shown below and, within each, several fine scale blobs. The observer is fixating the left hand blob, so that its disparity is zero. The right hand blob has a large uncrossed disparity (it is further from the observer).

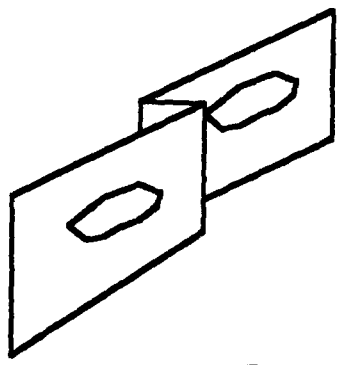
In Marr and Poggio's (1979) model, or any non-hierarchical model, the position of fine scale features is measured using the same co-ordinate framework as for coarse scale blobs. Hence the fine scale blobs within the right hand coarse blob have very different x-co-ordinates in the left and right eyes' images. This means that, in Marr and Poggio's scheme, they could not be matched without a vergence eye movement (e.g. fixating on the right hand blob) to bring the fine scale features into closer correspondence.

This is not necessary in a hierarchical model. There is, in a sense, a separate co-ordinate framework for each blob. The position of fine scale features is recorded only with respect to the coarse scale centroid. This means that the "hierarchical" position of a fine scale feature in the left and right eyes' images is always the same whatever the present angle of vergence happens to be. It follows that the *difference* in position of the fine scale feature in the left and right eyes' images - its disparity - also remains invariant with changes in vergence angle. When the observer is fixating on the coarse scale blob Marr and Poggio's model and a hierarchical one are equivalent.

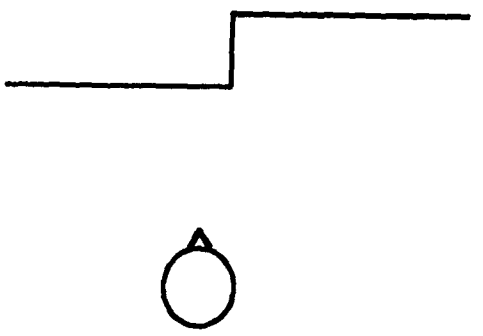
There are other theories which make explicit the relative position of features (e.g. horizontal separation or orientation) and use these measurements to calculate the

---

Fig 3.1 (overleaf)
This figure depicts the left and right eyes' views of two blobs which lie on two fronto-parallel surfaces at different depths from the observer, as illustrated above. The coarse scale outline of each blob is shown as an ellipse. Fine scale features within each blob are drawn in dotted lines. The left eye's view of the blobs is shown above, the right eye's view below.
Immediately beneath each blob the x-co-ordinate of the coarse scale centroid is given and, below that, the x-co-ordinates of the fine scale features. These values are retinal co-ordinates (in arbitrary units). Below these the hierarchical co-ordinates of the same blobs are shown for comparison. In this case, fine scale positions are all given relative to the coarse scale centroid.

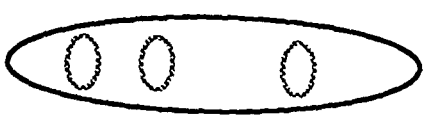


Side view



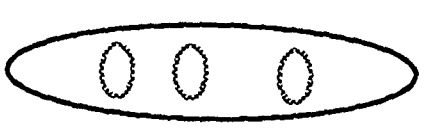
Plan

L



0

-3 -1 +3



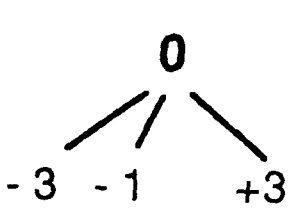
20

19 20 22

coarse blob

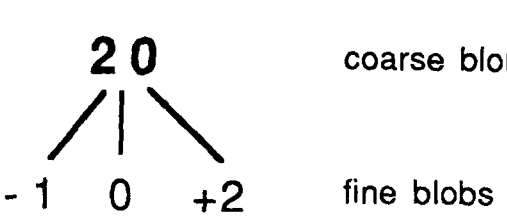
fine blobs

Retinal co-ordinates



0

-3 -1 +3



20

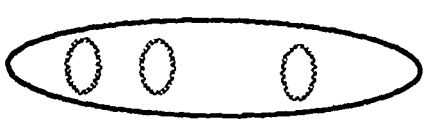
-1 0 +2

coarse blob

fine blobs

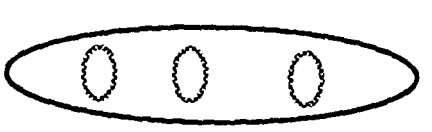
Hierarchical co-ordinates

R



0

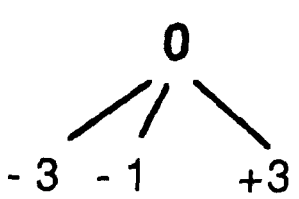
-3 -1 +3



30

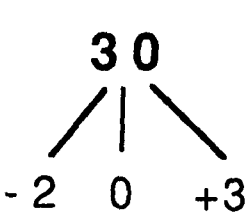
28 30 33

Retinal co-ordinates



0

-3 -1 +3



30

-2 0 +3

Hierarchical co-ordinates

Fig 3.1 (legend on previous page)

relative disparity* of the features. (e.g. "spatial frequency disparity" (Blakemore, 1970a); "diffrequency" (Tyler and Sutter, 1979); "orientation disparity" (Cagenello and Rogers, 1988); "curvature disparity" (Rogers and Cagenello, 1989)). The output of these models are local measures and, like hierarchical measurements of disparity, do not vary with the vergence angle of the eyes. In fact all these types of disparity are easily incorporated into a hierarchical model (as discussed in the next sections). A hierarchical model provides a logical framework for these local measurements and, by carrying out the process at a range of scales, enables disparities across the whole image to be compared. Thus a hierarchical model goes much further than previous models of relative disparity, since the outcome is a global description of the image, in many ways similar to the $2^1/_2$ - D sketch proposed by Marr and Poggio (1979).

It is an empirical question as to how far the coarse-to-fine analysis of a blob might proceed when the blob is not in the fixation plane or is outside the fovea. In either case, a likely scenario is that as "attention" is switched to a coarse scale blob (see section 2.3.6) finer scale analysis might begin to be carried out. But the analysis might only be completed down to the finest scale once an eye movement had taken place, thus making available the much finer resolution of the fovea.

Marr and Poggio (1979) noted that subjects could make accurate eye movements to fine scale features over relatively large changes in depth. They assumed that this was only possible after the subject had built up a $2^1/_2$ - D sketch of the object, i.e. a stored representation of the 3-D distance of features discovered when the observer first viewed the object. In a hierarchical scheme this is not required: computation of fine scale disparities can precede a vergence eye movement (a prediction borne out, for example, by experiments on the perception of slanted surfaces, discussed in the next section). It is also possible that the positions (and hence disparities) of fine scale features in several different blobs can be "remembered" without the need for time consuming calculations to be repeated, in an analogous way to the that used in the $2^1/_2$-D sketch, although 3-D (z) distances would not be represented.

---

* "Absolute disparity" is the disparity of a point with respect to the fixation plane (or, strictly, the Vieth-Müller circle which includes the point at which the line of sight from the two foveas intersects). "Relative disparity" is the difference in disparity of two points. It could be calculated from the absolute disparity of each point or from the relative position of the points in each eye's image. (See figure 2.1 and discussion in section 2.2)

3.2.2 A surface slanted about a vertical axis.

Mitchison and McKee's (1987a) study of matching in slanted surfaces (discussed in section 1.3) is one of the most important challenges to Marr and Poggio's theory. No vergence eye movement could account for their results and their model, of matching being guided by a tilted interpolation plane, is radically different from Marr and Poggio's scheme.

Figure 3.2 illustrates the hierarchical representation of features on a slanted surface. The left hand blob is on a fronto-parallel surface as before, the right hand blob is on a surface slanted about a vertical axis so that it faces towards the observer. As a result, its image in the left eye's view is wider than the in the right. The differential widths of the blobs in the left and right eye signal the coarse scale slant of the object explicitly.

The position of the fine scale features within the blob have been written in two ways. The first is as for figure 3.1. The second gives their positions as a proportion of the width of the "parent" blob. This means not only that, as well as position being relative, the *metric* for describing position is relative, too. Thus a fine scale feature with an x co-ordinate of -3 with respect to the coarse scale centroid is instead described as having in x co-ordinate of $-1/4w$ where w is the width of the coarse scale blob. Described in this way, the positions of fine scale features remains invariant not only to the vergence angle of the eyes, as discussed above, but also to the slant of the object. In the example illustrated in figure 3.2, the fine scale features have zero disparity when described in this metric, that is, they are flat in the plane of the slanted surface discovered at a coarse scale.

Blakemore and Campbell (1969) suggested that, if the visual system analyses the distribution of spatial frequencies in an object, then the ratios of these frequencies

---

Fig 3.2 (overleaf)
This figure illustrates a refinement of the hierarchical model shown in figure 3.1. Again, the left and right eyes' views of two blobs are shown. In this case the right hand blob lies on a slanted surface (slanted about a vertical axis) as illustrated above. Beneath each blob the coarse scale centroid is shown and the width (w) of the blob. The position of fine scale features is given relative to the coarse scale scale centroid in two ways. First, as in figure 3.1, the positions are given in the same units as those used at the coarse scale. Below this, the positions are given as a proportion of the width of the coarse scale blobs.

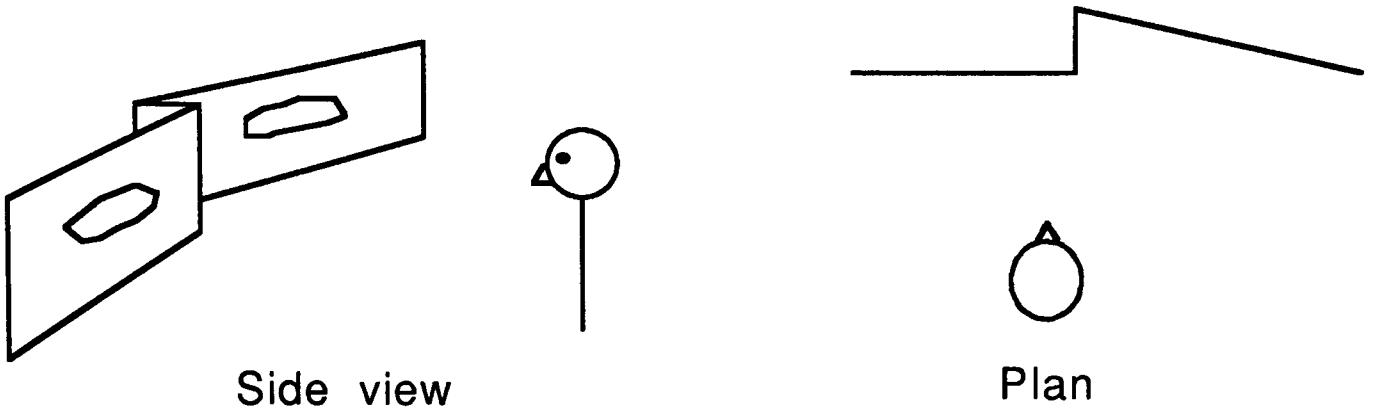


Fig 3.2 (legend on previous page)

could be used to identify the object at different distances. This idea is related to the one described here in which the metric for fine scale blobs is dependent on the coarse scale width of a blob. However, it is less clear in a Fourier-based model how the image would be segmented so that different transformations could be applied to different parts of the image, or how such a model could predict the matching results of Mitchison and McKee (1987a) for grids of dots.

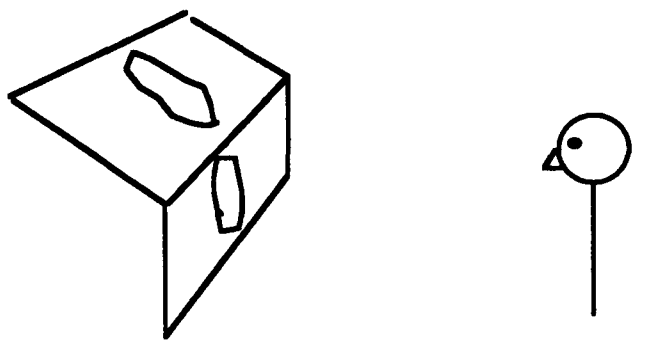
3.2.3 A surface slanted about a horizontal axis

Features which lie on a surface which is slanted about a horizontal axis have an orientation disparity rather than a width disparity, as illustrated in figure 3.3. Cagenello and Rogers (1988) suggest ed that orientation disparity may be used directly to signal surface slant (about a horizontal axis). The orientation of the principal axis of a blob is one of the measurements recorded explicitly in the hierarchical scheme proposed here, so it is natural to assume that orientation disparity is as well.

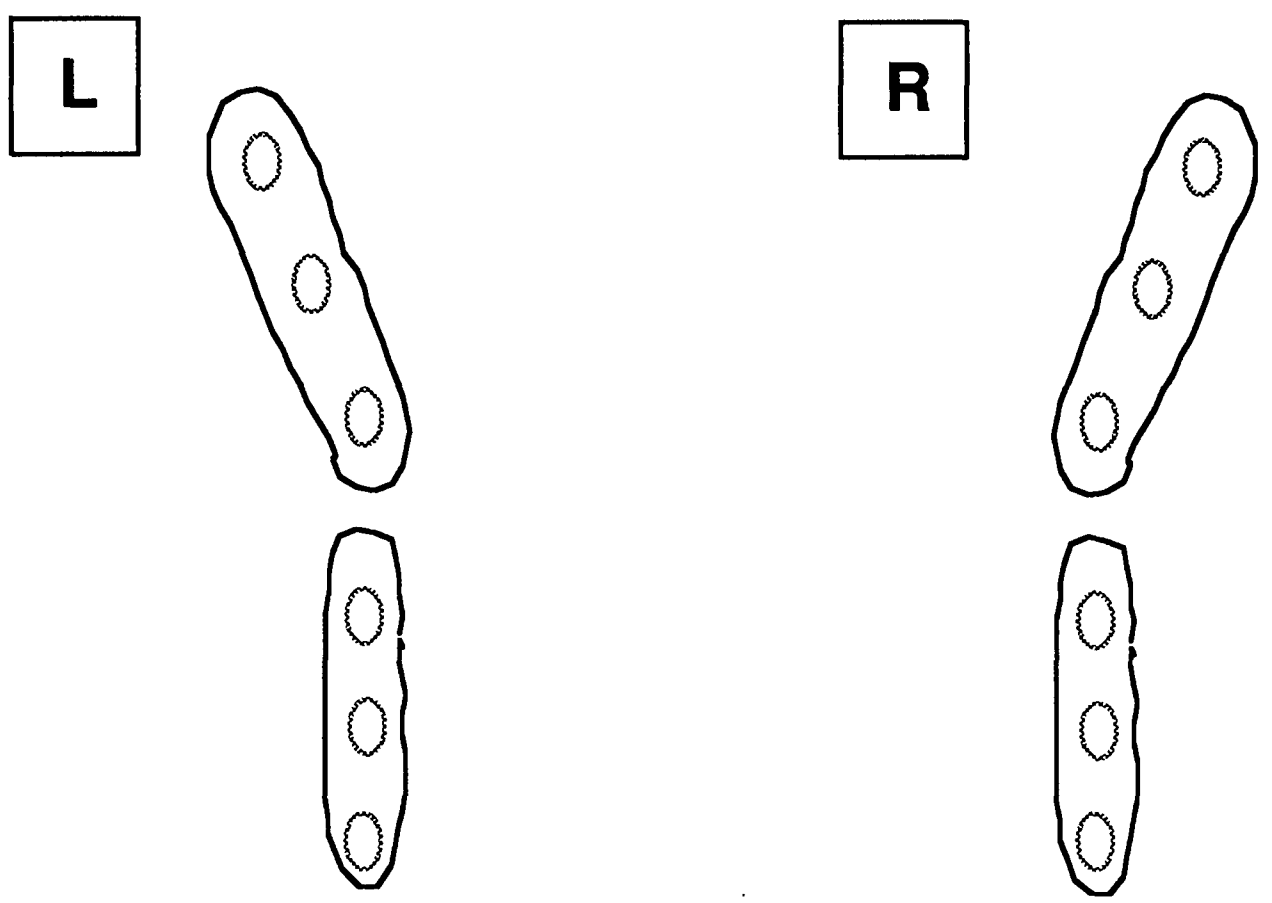
It is possible that orientation, like position, is also defined hierarchically. That is, the orientation of a fine scale blob might be defined relative to the principal axis of the coarse scale "parent" (as illustrated in figure 3.3 (below)). The consequences of this type of representation are discussed in more detail in section 7.2 (in relation to the perception of simultaneous depth contrast in slanted surfaces).

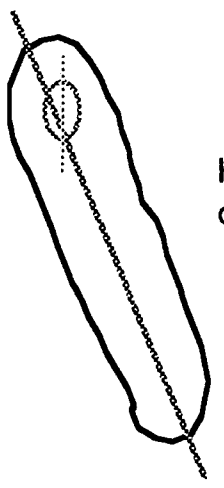
3.2.4 A curved surface

Finally, figure 3.4 illustrates the representation of features on a surface which is curved in a horizontal direction. The interesting aspect of this example is that the coarse scale representation is exactly the same as for the slanted surface shown in figure 3.2. Two consequences follow. One is that fine scale features cannot be represented in relation to the the coarse scale curvature (as, for a slanted surface, they were represented in relation to the coarse scale slant) because there is no such thing as the "coarse scale curvature". The position of fine scale features is instead likely to be recorded in terms of the width of the parent blob, as described above, and so each fine scale blob will have a non-zero disparity (which will be the disparity with respect to the plane joining the two ends). The model is very similar to the matching scheme proposed by Mitchison and McKee (1987a) who describe matching in stimuli which are curved in a horizontal direction. A detailed



Side view





Hierarchical representation of blob tilt.

Fig 3.3

At the top, an observer is shown viewing two surfaces surfaces, the top one of which is slanted about a horizontal axis, the bottom one is fronto-parallel. Beneath this, left and right eyes' views of two blobs on the surfaces are shown. The difference in orientation of the top blob in the two eyes' images (its orientation disparity) can be detected at a coarse scale. The (2-D) tilt of fine scale features could be defined relative to the angle of the coarse scale principle axis. If this were the case then orientation disparity, like width disparity, could be defined hierarchically.

comparison of Mitchison and McKee's theory with a hierarchical model is made in section 5.1. The point of interest is that they found their results could be predicted by considering matching relative to an interpolation *plane* , including a tilted plane, but not relative to an interpolation *curved* surface.

The second consequence is that the representation of curvature in a horizontal and a vertical direction will be different. For curvature in a horizontal direction, as has been said, there is no such thing as the "coarse scale curvature" of a blob. For curvature in a vertical direction, as illustrated in figure 3.5, the same is not true. The upper blob shown in figure 3.5 curves in opposite directions in the left and right eyes' images. The (2-D) curvature of a coarse scale blob is likely to be recorded explicitly as well as the *difference* in curvature of a blob in the two eyes' images, i.e. the "curvature disparity" (Rogers and Cagenello, 1989). This means that at a coarse scale the fact that a surface is curved rather than slanted can be recorded explicitly, as can the direction of curvature and, to some extent, the magnitude of the curvature. (A simple metric for curvature disparity would be difficult to construct as it would depend on the monocular curvature of the feature.) None of this information can be gained from the coarse scale outline of a single blob on a horizontally curved surface. The difference in the representation of curvature in horizontal and vertical directions may play a part in the "anisotropy" of perception of sinusoidally corrugated surfaces (e.g Rogers and Graham, 1983), a possibility discussed in section 7.2.

3.2.5 Explicit versus implicit information

An important theme emerges from these examples. It is that, in a hierarchical model, some pieces of information are explicit and others are implicit, a difference

---

Fig 3.4 (overleaf)
In this figure the right hand blob lies on a surface which curves in a horizontal direction, as illustrated above. Below, as in figure 3.2, the left and right eyes' views of the blobs are shown. Beneath each is shown the coarse scale centroid and blob width and the hierarchical co-ordinates of the fine scale features within the blob. The coarse scale representation of the blobs is exactly the same as for figure 3.2. This reflects the fact that there is no information about the curvature of a single blob. Only the positions of the fine scale features indicate that the surface is curved.

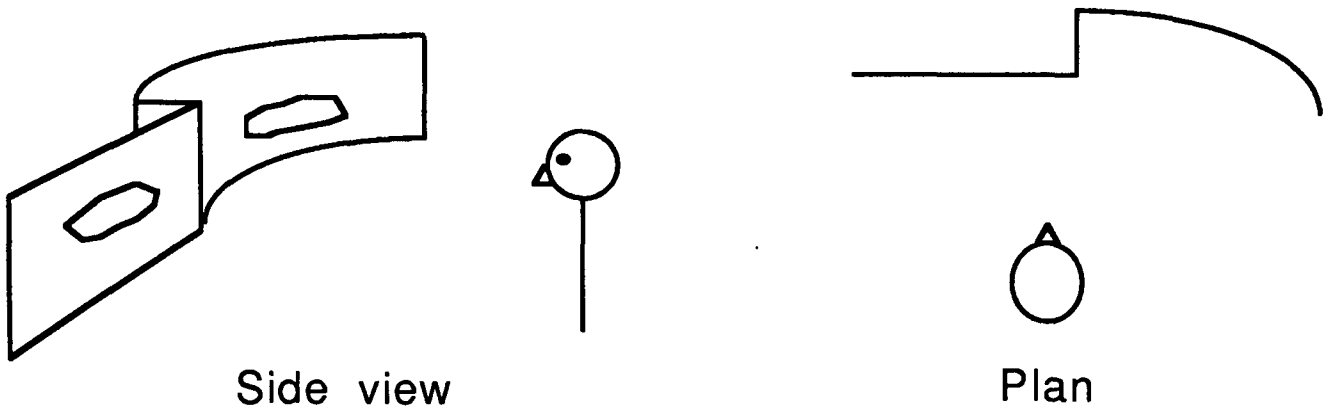


Fig 3.4 (legend on previous page)

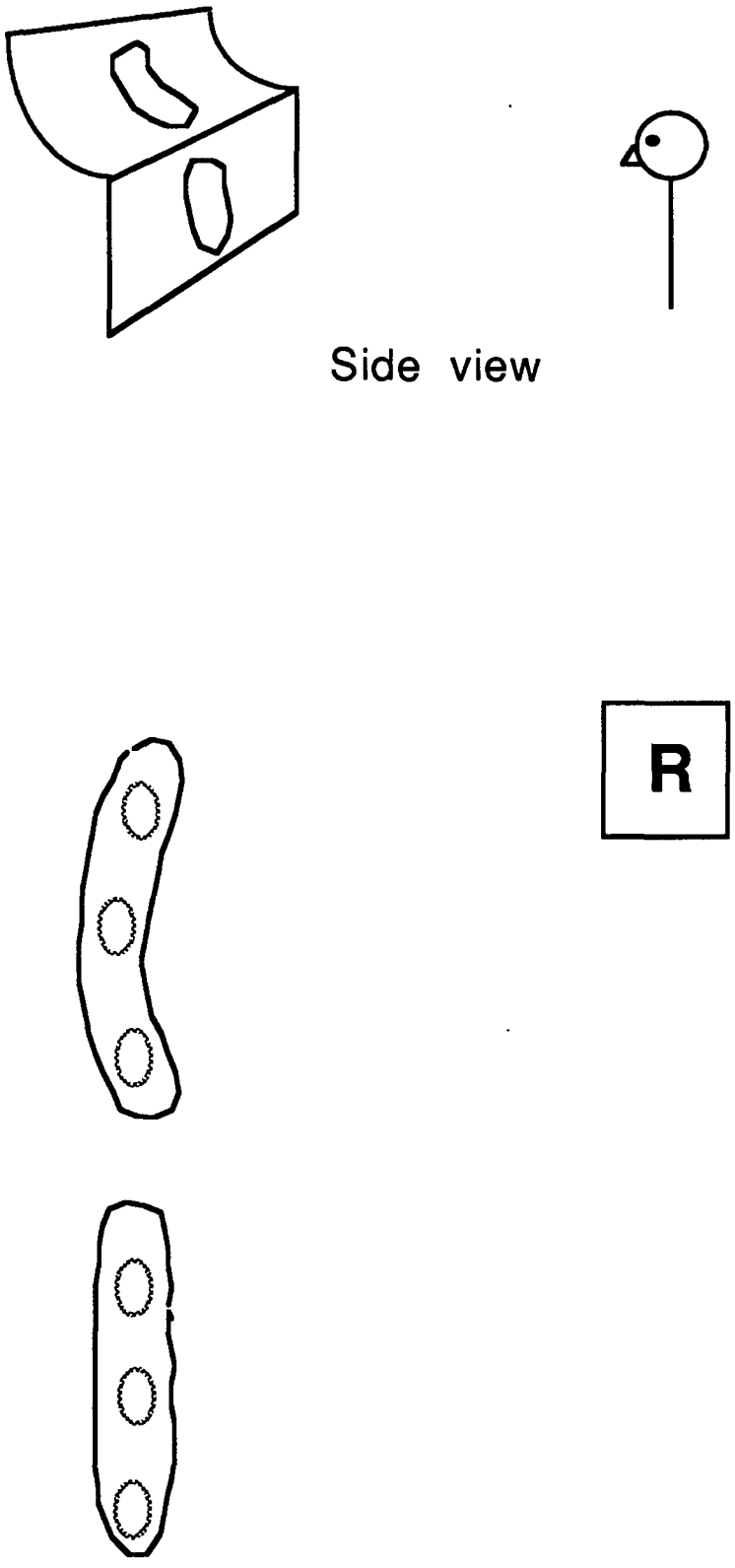


Fig 3.5

At the top, an observer is shown viewing two surfaces surfaces, the top one of which is curved in a vertical direction, the bottom one is fronto-parallel. Beneath this, left and right eyes' views of two blobs on the surfaces are shown. The difference in curvature of the top blob in the two eyes' images (its curvature disparity) can be detected at a coarse scale.

which may be reflected in our perceptions and in psychophysical performance. For instance, fine scale features in a blob away from the fixation plane have a precise disparity with respect to the fixation plane but this is represented only implicitly (as the combination of a coarse scale and a fine scale disparity). Similarly, fine scale features on a slanted surface (slanted about a vertical axis) have a disparity with respect to the fronto-parallel but this is not represented explicitly (only the disparity with respect to the coarse scale blob). And, as a slightly different example, the horizontal curvature of a surface is represented implicitly, because the disparity of all the fine scale features is known and could be used to "build up" the curvature, but nowhere is the curvature recorded explicitly (i.e. as information about one blob (or one pair of blobs) at one scale).

This issue, i.e. how information from a hierarchical database is used in performing visual tasks, relates to two of the experiments described in the following chapters and is the main question examined in chapter 5.

3.3 Experimental approach

3.3.1 Rationale for avoiding filtered stimuli

Many of the experiments which have addressed the role of spatial scale in binocular stereopsis have used stimuli which are band-passed in luminance spatial frequency. Some of these have been discussed in chapter 1. Filtered stimuli have not been used in the experiments described in this thesis for two reasons. First, if results are found to vary with the spatial frequency of the stimulus, it is difficult to know whether this is due to different "mechanisms" in the visual system being stimulated or whether it is just that the information in the stimulus is changing (examples from recent experiments are discussed in the next section). In other words, it is easy to end up studying properties of the stimulus rather than properties of the visual system. Second, for high spatial frequency filtered stimuli, the coarse spatial structure of the stimulus often remains (particularly in contrast modulated or sparse patterns) even when all *luminance* low spatial frequency energy has been removed. This has led to a debate about whether mechanisms in the brain can "reconstruct" the low spatial frequencies in the image through various non-linear stages. In section 3.3.3 the MIRAGE definition of "coarse scale structure" is illustrated using a variety of different images. It is argued that, in general, the luminance low spatial frequency and coarse spatial structure co-incide

closely. (Note that the same debate does not arise for low-pass filtered stimuli. Fine scale information once lost cannot be retrieved.)

To clarify these points, several experiments are reviewed in the next two sections.

3.3.2 Do results reflect properties of the stimulus or of the visual system?

This question was raised in chapter 1 in relation to studies on stereoacuity (section 1.5). Legge and Gu (1989), for example, showed that stereoacuity thresholds, measured using sine wave stimuli, are inversely proportional to the spatial frequency of the stimulus (below 3 c/deg.). One possible explanation is that different spatial frequency tuned mechanisms are being stimulated, each with a different acuity. Alternatively, the results may reflect an inevitable limitation of the stimulus. In other words, the quality of the information in the stimulus may change with spatial frequency and account for the change in performance. There is some evidence in favour of the second hypothesis. Blake and Zisserman (1987) show that the precision with which a step edge can be located in noise deteriorates as scale is increased (their figure 4.16). A similar result has been derived analytically by Canny (1986). He considers the theoretical precision with which an edge can be localised in noise in a linear system. He shows (p681) that the standard deviation of the estimate of the location of an edge increases with edge blur. (Specifically, he shows that the standard deviation of the estimate of edge location (δx_0) is inversely related to the second derivative of the convolution of the edge with the filter at the real location of the edge ($H''_G(0)$).) A similar result has been derived by Watt (1988, p52). He shows that the error in locating a centroid in the convolution of an edge with a Laplacian of Gaussian filter increases with the blur of the edge (If some level of internal blur is assumed then thresholds only increase for edge blurs above some critical value, which is consistent with Legge and Gu's results).

The same problem of interpretation arises in studies of two-frame apparent motion using periodic stimuli (e.g. Cleary and Braddick, 1990a; Chang and Julesz, 1983; Bischof and Di Lollo, 1990). In these experiments, the task was to discriminate the direction of motion in a 2-alternative forced choice (2AFC) procedure, and the displacement between frames at which performance breaks down (d_{max}) was measured. The stimuli used were 50% random dot patterns filtered with narrow band, isotropic band-pass filters. Over a wide range of spatial frequencies, these authors found that d_{max} increased in inverse proportion to the centre frequency of the stimulus, i.e. the coarser the filter the bigger d_{max} they measured. As in the

previous example, it is unclear whether this result indicates the presence of a range of spatial frequency tuned motion mechanisms, each with a different d_{max}, or whether the results reflect a property of the stimulus (for instance, the spacing of false targets in the stimulus) which would limit performance in any case. (This particular issue is discussed in more detail in chapter 6).

So, the demonstration of a "spatial frequency effect" does not necessarily imply the existence of spatial frequency tuned mechanisms. Even if such mechanisms do exist, filtered stimuli may not be the best way to study them. The original aim of using band-pass filtered patterns was to stimulate one "mechanism" while leaving all the others silent. Cleary and Braddick (1990a) admit that this is not necessarily the case:

> *"..it must be noted that Fc* [the centre frequency of the stimulus], *need not reflect the preferred frequency of the mechanisms mediating direction discrimination...If the direction discrimination performance was mediated by sensors tuned to frequencies 1 octave below Fc, d_{max} could be expressed as 1/2 a cycle of the detectors' preferred frequency."*
>
> (Cleary and Braddick, 1990a, p313-314)

In other words, even assuming such "mechanisms" are present in the visual system, it is not possible to decide which one is responsible for the observed performance.

3.3.3 Can coarse scale mechanisms be "silenced"?

Another group of experiments, also using filtered stimuli, gives rise to a different problem of interpretation. The objective of these experiments is to study the properties of fine scale mechanisms in isolation by using high frequency filtered patterns. The assumption is that coarse scale mechanisms are "silenced" by the filtering process.

A typical example of this type of experiment is one by Burbeck (1986). She measured thresholds for discriminating the separation between two gabor patches* and found they were independent of the "carrier" spatial frequency within the gabor patch. (Similar results have been obtained, for a bisection task and a three patch alignment task, by Toet and Koenderink, 1988). Unfortunately, these results do not distinguish between the hypothesis that high spatial frequency "mechanisms" can accurately measure large distances and one claiming that low

---

* A gabor patch is a luminance sine wave grating (the "carrier") modulated in amplitude by a gaussian envelope.

spatial frequency mechanisms do the task, after a non-linear process such as rectification has taken place. Various hypotheses about non-linear processes which might take place in the visual system have been put forward (e.g. Chubb and Sperling, 1988; Morgan et al., 1990).

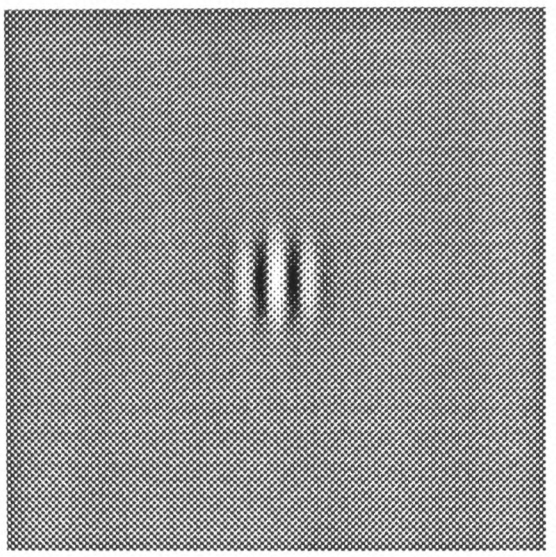
A different argument may be put forward in relation to the MIRAGE algorithm described in chapter 2. In that model, a coarse scale representation is made up from the output of a range of filters (summation after half-wave rectification). A gabor patch containing only relatively high frequencies would selectively stimulate finer filters. Figure 3.6a illustrates such a case. The output of three Laplacian of Gaussian filters in response to the gabor patch are shown on the left. The space constants of the filters are 16, 8 and 4 pixels. The formula for a Laplacian and the method by which filter outputs were combined to give the MIRAGE S+ response is given in Appendix B. On the right are shown the S+ MIRAGE responses if the output of all three filters were summed (top), only the smallest two summed (middle) and the finest filter alone (bottom). The "coarse scale" MIRAGE response (top right) carries information about the gabor patch, and would do so even if the coarse scale filter output were zero (in this example the coarse filter output is of much lower amplitude than the other filters). Burbeck wished to investigate the possibility that coarse scale mechanisms might exist and have different properties from fine scale mechanisms, for instance in relation to length judgments. The assumption made in her paper is that, if such mechanisms existed, it would be possible to selectively stimulate each with stimuli filtered at different spatial scales. A MIRAGE analysis of these and other stimuli prompts a different approach. As discussed in the previous chapter, the emphasis is not on coarse or fine scale mechanisms but on the properties of coarse representations (a MIRAGE S+ signal with all the filters "switched in") and fine scale representations (which are derived from only the finest filter output, but the information about each fine scale blob depends on the hierarchical structure of the image).

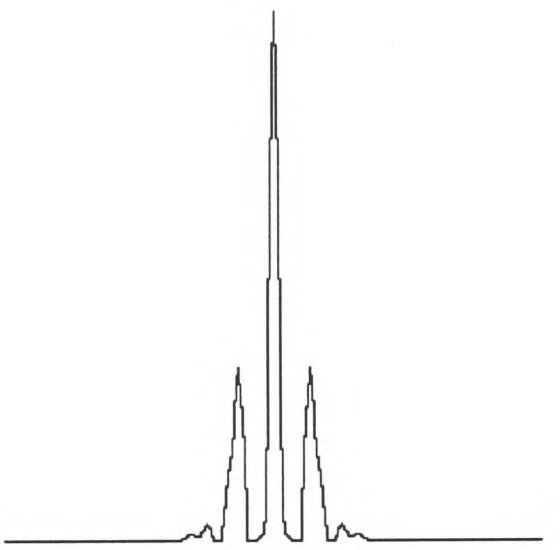
It is untypical for the coarse scale MIRAGE representation to be dominated by the output of fine filters as it is for a gabor patch. The MIRAGE responses to several other images are shown in figure 3.6b-e. As for figure 3.6a, the responses of three Laplacian filters is shown on the left and the MIRAGE S+ responses, from coarse to fine scales, on the right. The MIRAGE response at the finer scales has been shown with the outline of the coarser scale MIRAGE response shown as a dotted line in order to emphasise the grouping structure. As discussed in chapter 2, each new "blob" which appears as a filter is switched out lies within the boundaries of

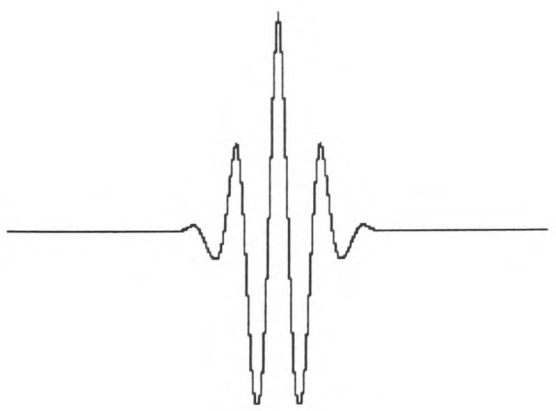
the blob at a coarser scale. For the low density random dot pattern (figure 3.6b) the positive responses of all three filters occur in the same place and so the spatial layout of the coarse scale MIRAGE blobs is very similar to the coarse filter output. All the fine filter responses lie within, or are grouped together by, the coarse filter output. The same is not true of the 50% random dot pattern (figure 3.6c). Here the coarse scale MIRAGE response is more crowded than the coarse filter response. There are many small blobs which are derived from the fine filter outputs and these fill in the spaces between the larger blobs. Thus, although the amplitude spectrum of a low density and a high density random dot pattern are both flat (see section 6.8) they have very different coarse scale representations in MIRAGE.

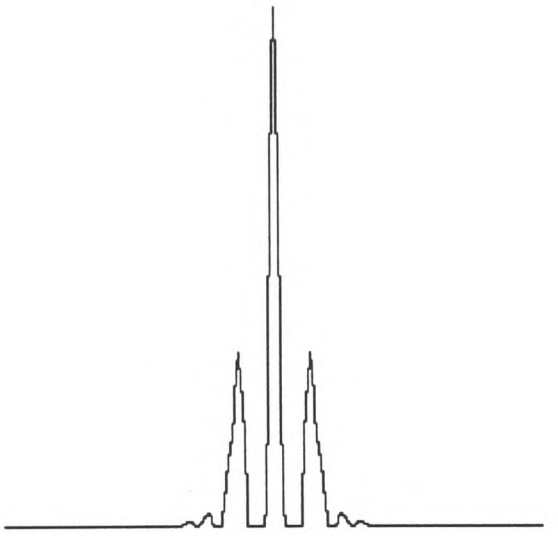
Figures 3.6d and e show the MIRAGE responses to a 1/f spectrum noise pattern and a natural image, which has similar spectrum (i.e. the amplitude of the fourier energy at any spatial frequency is proportional to the frequency). Most natural images have a spectrum which is approximately 1/f (e.g. Field, 1987). The coarse scale MIRAGE response for both patterns is heavily dominated by the output of the coarsest filter since the amplitude of other filter outputs is relatively small.

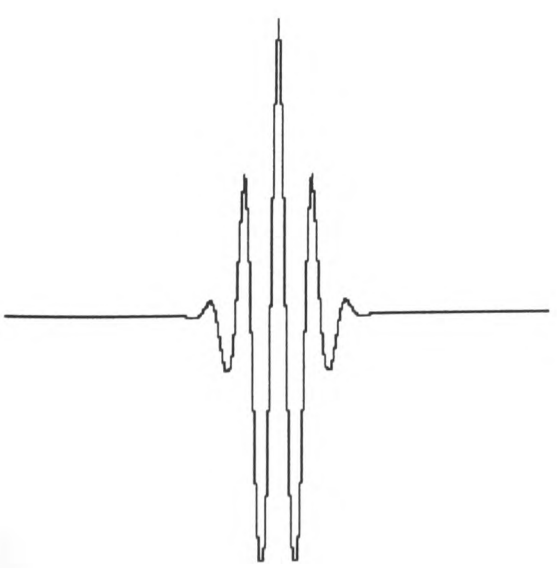
Fig 3.6 (overleaf)
This figure illustrates the MIRAGE response to five different images (a-e). The images are (a) a gabor patch, (b) a 0.05% density random dot pattern, (c) a 50% density pattern, (d) a noise pattern with a "1/f" spectrum and (e) a natural image. On the left below each image are shown the responses of three Laplacian of Gaussian filters to the image at the top. The image is in each case 256 by 256 pixels and the space constants of the filters are 16, 8 and 4 pixels (from top to bottom). The output shown is a one-dimensional slice through the filtered image (at line j = 85). The impulse response of all the filters has an equal peak-to-trough amplitude (details are given in appendix B). The amplitudes of the filter outputs can be compared directly (for any one image, they are all plotted on the same y-axis). On the right are shown the MIRAGE S+ responses when all three filter outputs (16, 8 and 4 pixels) are added together (top), when only the smaller two are added together (centre) and for the smallest filter alone (bottom). For images b - e, a dotted line shows the S+ response when a larger filter is included (e.g. in the centre, 16+8+4 is shown as dotted, 8+4 shown as a solid line). This helps to illustrate the hierarchical structure of the MIRAGE representation: blobs which appear at a fine scale always lie within the boundaries of blobs present at a coarse scale. for any one image, the same y-axis has been used to display all three MIRAGE responses (but it is different from that used for the filter responses). The standard deviation of the Gaussian for the gabor patch (a) is 12 pixels, the period of the sine wave is 16 pixels. The "1/f" noise pattern (d) was created by multiplying the fourier amplitude spectrum of the 50% pattern (c) by an appropriate mask (amplitude = frequency) and then carrying out an inverse fourier transform. Thus the filter responses to the 50% and "1/f" patterns differ only in relative amplitude.











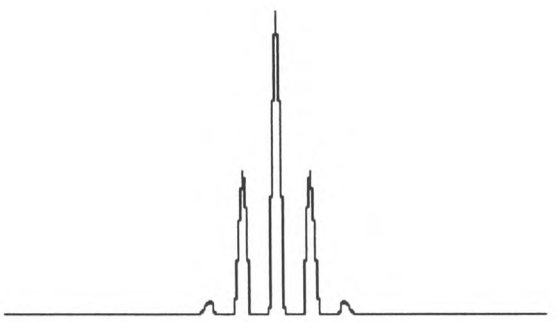
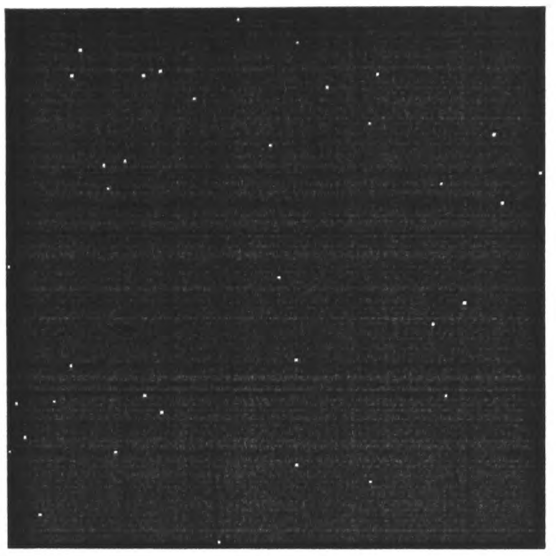


Fig 3.6a (legend on previous page)



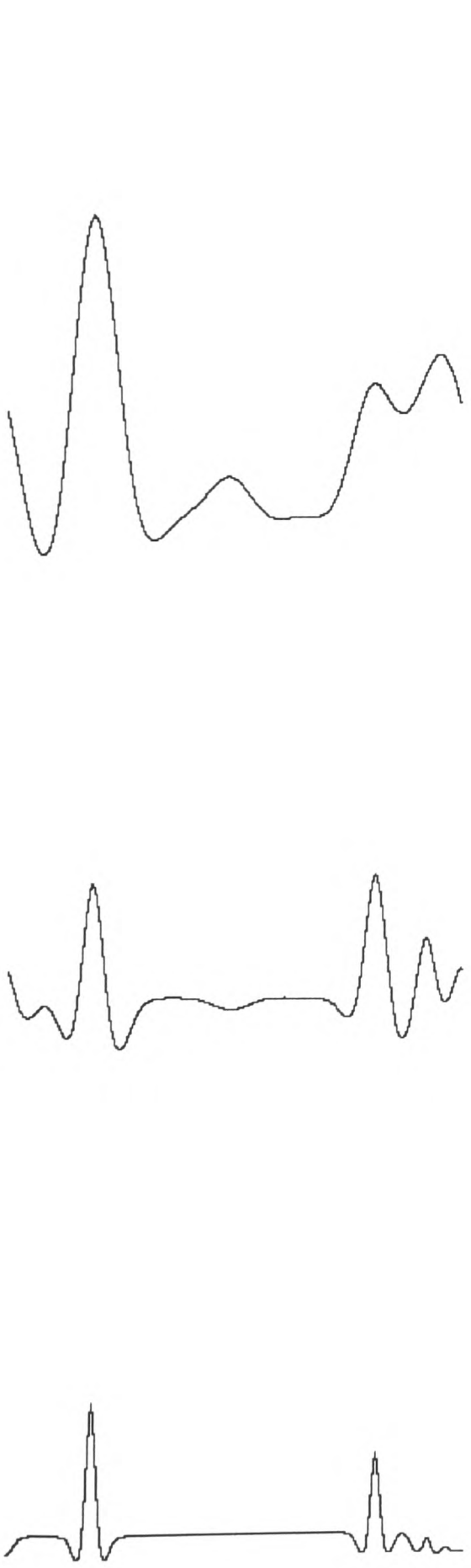
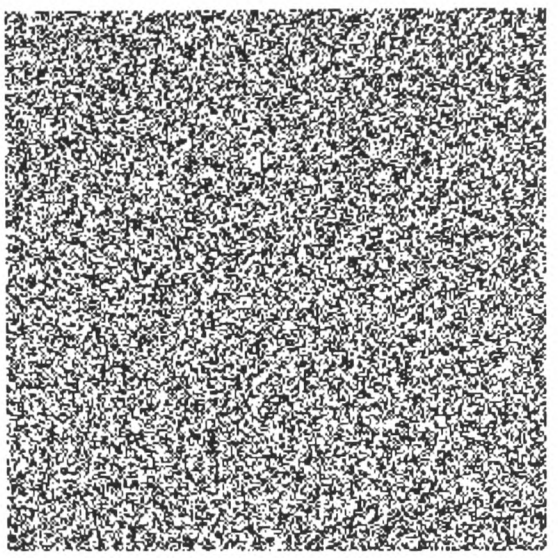
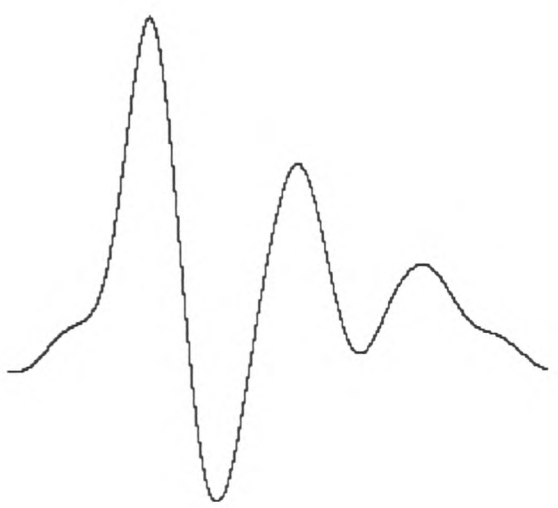
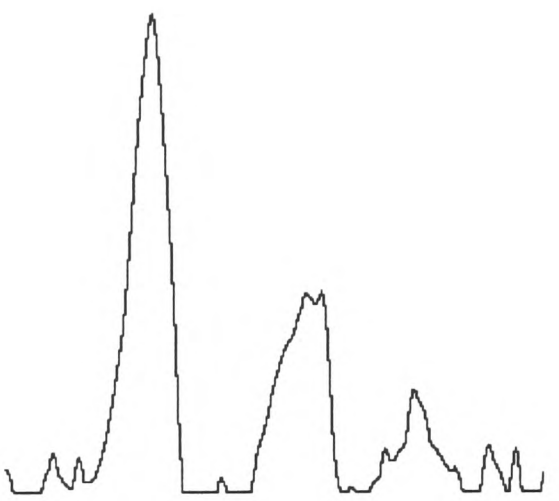
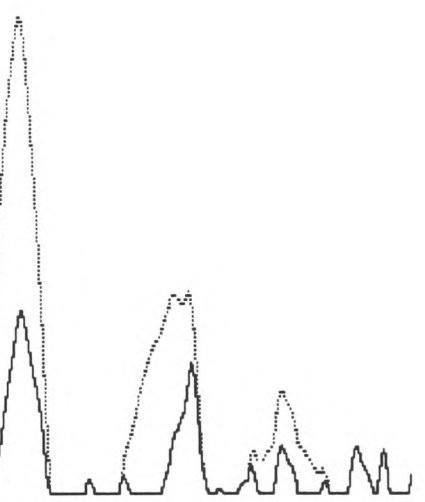


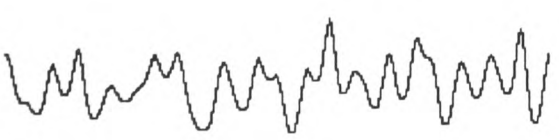
Fig 3.6b (legend on previous page of text)











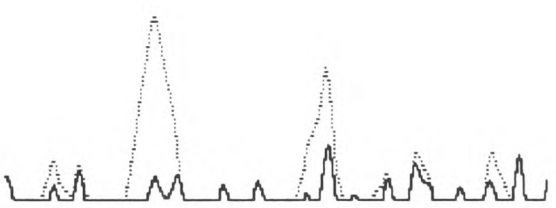
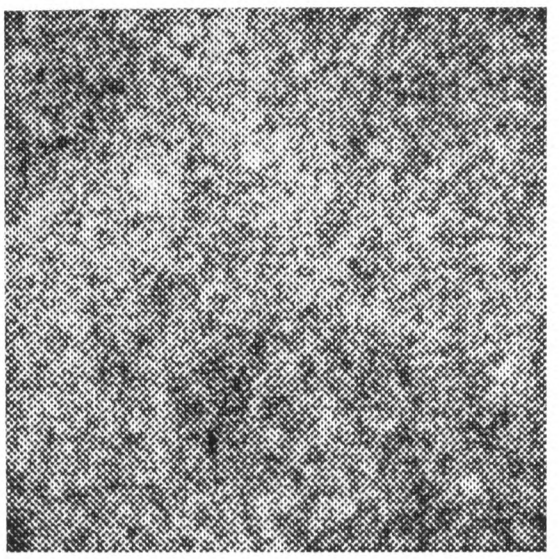
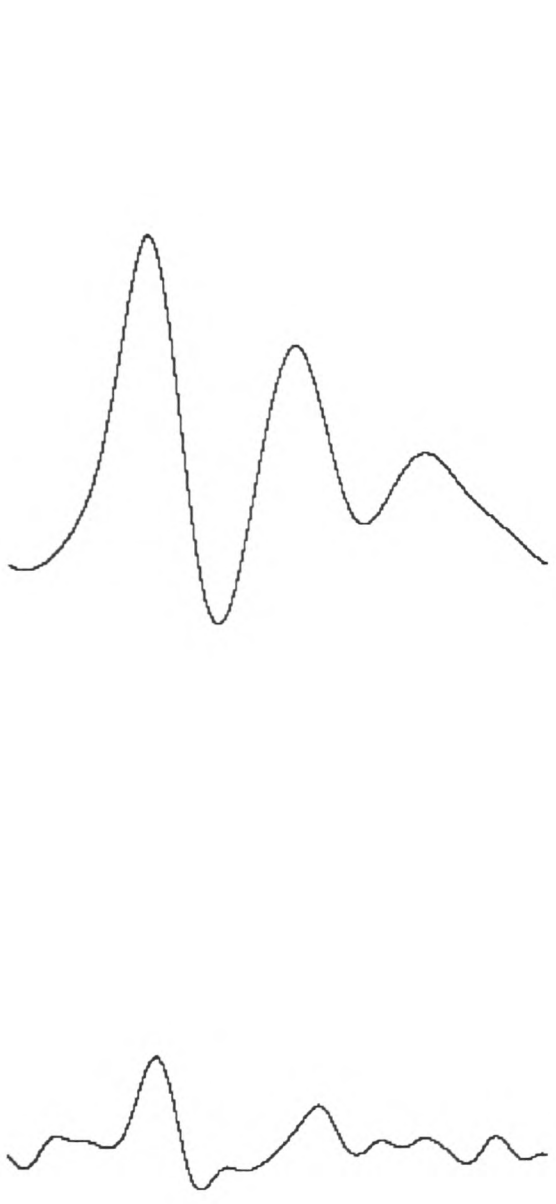
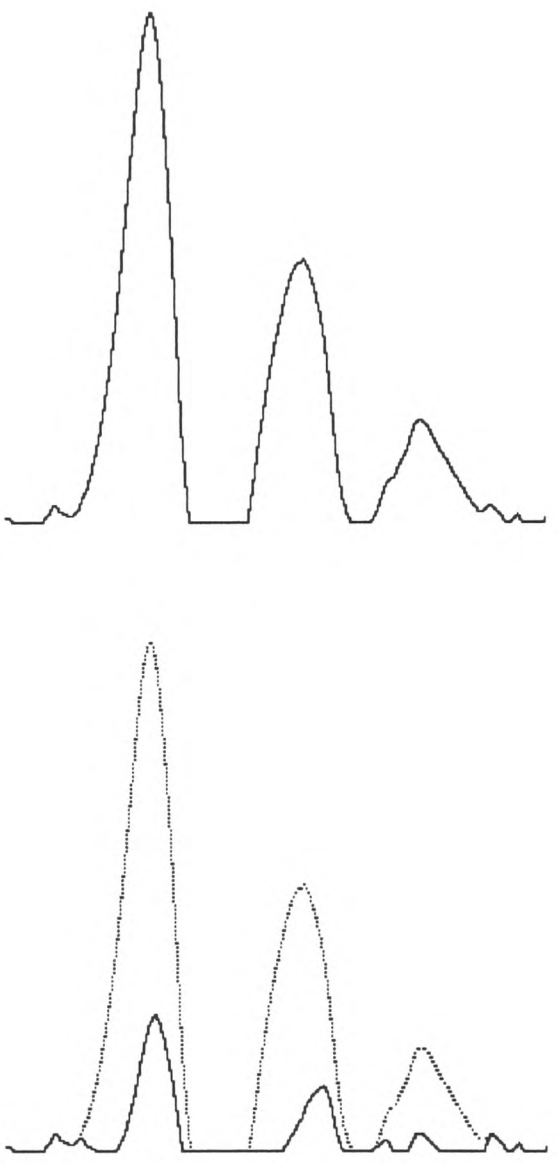
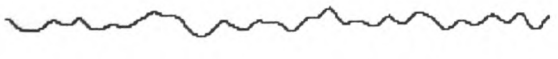


Fig 3.6c (legend on previous page of text)









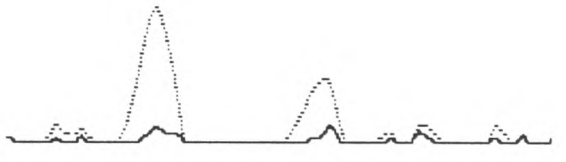
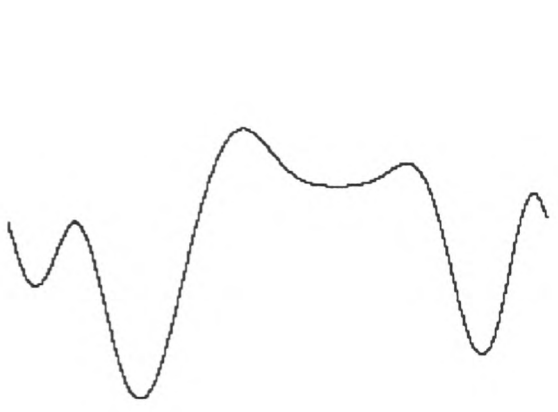


Fig 3.6d (legend on previous page of text)



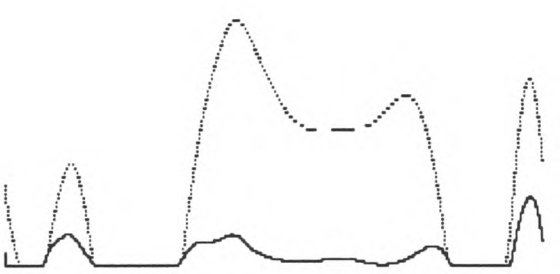


Fig 3.6e (legend on previous page of text)

The spacing of zero-bounded distributions (blobs) in the 1/f pattern must be the same as in the 50% pattern since it differs only in the relative amplitude of filter outputs. However, if centroids below some threshold mass were ignored (for instance because they were not significantly different from the background noise level) then the mean centroid spacing might differ in the two cases. This question is of relevance to the experiment and modelling described in chapter 6.

The examples shown in figure 3.6 illustrate some of the differences between a "coarse scale" MIRAGE representation and the output of a coarse filter, or low spatial frequency channel. However, for many types of images there is a strong similarity between the form of the coarse scale MIRAGE response and the output of the coarsest filter and an analysis of the coarse filter output alone serves as a reasonable approximation to the coarse scale MIRAGE response. Figure 3.6b illustrates this point for a low density pattern. In chapter 6 it is shown that for low density patterns such as this the spacing of centroids is the same for either the MIRAGE response or a single coarse filter output. For stimuli made up of thin bright lines on a dark background, as are used in chapter 5, the coarse scale MIRAGE S+ signal is essentially the same as the coarse filter output and this approximation has been used in modelling the results. (Note that the S- signal for this stimulus is very different from the negative response of the coarsest filter, a fact which is relevant to the discussion of how stimuli containing high spatial frequency information can be distinguished from blurred stimuli at short exposure durations (see chapter 2).)

The characteristics of the MIRAGE representations illustrated in figure 3.6 have been discussed in detail because they have influenced the design and modelling of experiments described in the following chapters. For instance, the grouping of the blobs derived from fine filter outputs in a MIRAGE representation of some images leads to predictions about the performance of subjects when asked to compare the separation of fine scale features which are grouped differently. This issue is explored further in chapter 5. The MIRAGE representations of a low and high density random dot patterns relate to the experiment described in chapter 6. Marr and Poggio (1979) assumed that matching was limited by the density of false targets in the image. The maximum disparity which can be detected in random dot stereograms of different densities is measured and the results compared to the density of MIRAGE blobs in the images. The experiment described in chapter 4 is related to the idea that first the coarse and then progressively finer filters might be "switched out" as exposure duration is increased.

3.4 Summary

In this chapter a hierarchical system for coding disparity has been examined. It is derived from the hierarchical system for encoding position (Watt, 1988) which was introduced in the last chapter. An addition to the theory was suggested, i.e. that the position of fine scale features might be encoded in terms of the *width* of the parent coarse scale blob. This would make a nearest neighbour matching scheme in a hierarchical model very similar to the matching scheme proposed by Mitchison and McKee (1987a).

The design of the experiments described in the following chapters have been influenced to a large extent by characteristics of the MIRAGE model and examples of MIRAGE responses to a range of stimuli were discussed.

CHAPTER 4

4.1 Local-to-global or coarse-to-fine?

The aim of this chapter is to provide some direct evidence for a coarse-to-fine process in human stereopsis. The strategy is to try to "catch" the stereoscopic process "in the act" of moving from coarse to fine spatial scales. Stimuli in this experiment are presented for very brief exposure durations (from 60 ms up to one second), after which a mask is presented to prevent retinal persistence and limit the subsequent processing carried out on the image. The stimuli were chosen so that, if stereoscopic matching is in fact a local-to-global process, i.e. starting with fine scale information (including many ambiguous matches) and "building up" a representation of the surface then, at least for a simple local-to-global model, all the

stimuli should be seen equally easily. But if stereoscopic matching is a coarse-to-fine process, then one of the stimulus types should be especially difficult to see at short exposures, when the scale of analysis is still large.

It is important to note that the demonstration of a temporal sequence of analysis from coarse to fine spatial scales would not, by itself, provide evidence for a hierarchical model of stereopsis. The experiment described in chapter 5 is more directly concerned with how the information about an image might be stored, i.e. it examines the *purpose* of a coarse-to-fine analysis.

The stimuli used in this experiment were first used, and more important, their spatial frequency characteristics analysed, by Parker and Yang (1989). Their paper is discussed in the next section.

4.2 Parker and Yang

A paper by Parker and Yang (1989) investigated the perception of a type of stimulus that carries different information at different spatial scales. A version of their stimulus is used in the experiment described in this chapter. The aims of the two experiments are rather different, but it is worth discussing their paper in some detail, partly because it provides an introduction to the properties of this type of stimulus at a range of spatial scales and partly because the issue that Parker and Yang addressed is relevant to Marr and Poggio's (1979) theory.

Parker and Yang investigated "depth averaging" in random dot stereograms in which two surfaces with very similar disparities were present in one area. The patterns they used were 50% density random dot patterns. One surface was described by the disparity of odd rows within the region, the other by the disparity of the even rows. When the two surfaces could be distinguished one appeared "transparent" with the other visible behind it. When their disparities were similar they were perceived as a single surface lying at a depth approximately equal to the average depth of the two surfaces. Parker and Yang found that the range over which this "disparity averaging" occurred, i.e. the maximum difference between the disparity of the two surfaces before they were seen as two*, varied according

---

* In fact the criterion for the breakdown of "disparity averaging" they used was not that the subject saw two planes but that the perceived depth of the single surface seen (as judged by the bias of the psychometric function (50% correct point)) was shifted from the average disparity of the two planes by an amount greater than the subject's stereoacuity. (The subject's stereoacuity

to their average (or pedestal) disparity . The further away from the fixation plane they were the greater the disparity difference between the surfaces before disparity averaging "broke down".

4.2.1 Two interpretations of disparity averaging

Parker and Yang discuss two possible mechanisms for "disparity averaging". One is that the disparity for each individual dot is calculated first and followed by a process of "interpolation" (i.e. averaging in the disparity domain). If the disparity difference between odd and even rows is sufficiently small then a single plane would be "interpolated" through the dots at a disparity mid-way between the disparity of the two sets of dots. If the disparity difference exceeded some critical value then dots would be interpreted as lying on two separate planes. The critical value would depend, according to this model, on the distance from the fixation plane.

The other explanation of disparity averaging they put forward is that a stage of monocular blurring precedes the matching process. It is not immediately obvious that blurring the left and right eyes' images should result in disparity averaging. One way to think of it is in the spatial domain. When two features in the image are blurred together the resultant "blob" lies between the two, in fact, on average, exactly half way between the two. So, if the *positions* of the features in odd and even rows have been averaged, inevitably their disparity will also be averaged.

The effect of monocular blurring can also be thought of in terms of the cross correlation of left and right images, i.e. in the disparity domain. Parker and Yang describe the search for two disparity planes as a task of resolving two peaks in an area-based correlation of the left and right eyes' images (within the target area). Examples of this type of analysis are shown later in the chapter (figure 4.11). For an unfiltered pattern the two planes form two spikes in the cross-correlation function. For low-pass images the two planes form Gaussians in the cross-correlation function whose width is proportional to the size of the filter. When the disparity difference of the two planes is small compared to filter size, the Gaussians coalesce to form one peak. The larger the filter, the further apart in disparity the two planes must be for them to be resolved. Parker and Yang's results, according to this scheme, imply that surfaces further from the fixation

---

was measured in a separate experiment as the threshold for distinguishing the depth of two spatially separated patches within the random dot stereogram.) This criterion for the limit of disparity averaging correlated with, but was always smaller than, the disparity difference at which two planes were seen.

plane are processed by larger monocular filters. (How this might occur is explained in the next section).

Although they do not consider that there is sufficient evidence to rule out either hypothesis, Parker and Yang favour the second ("monocular filtering") model. They conclude their paper by saying,

> *"The fact that larger disparity differences can be tolerated in mixed disparity targets when the average disparity is offset from the fixation plane implies that the spatial analysis for non-zero disparity targets takes place through filters of a larger spatial scale. In other words, larger disparities are processed through coarser filters than small disparities. This inference suggests that the neural apparatus is available to implement a* coarse-to-fine *strategy in stereo matching (Marr and Poggio, 1979)."*
>
> (Parker and Yang, 1989, p 1537)

4.2.2 Relationship of pedestal disparity and filter size

It is worth reiterating the logic that links Parker and Yang's result to Marr and Poggio's model. The parts of the image with a large average, or "pedestal", disparity are filtered at every scale, including the finest scale, but the argument is that the output of the finest filters cannot be matched (for any disparity greater than ± w of the filter). Therefore, the minimum filter size giving a *coherent* disparity signal is greater for larger pedestal disparities. In other words, the extent of disparity averaging gives an indication of the smallest filter that can solve the correspondence problem at that disparity.

Marr and Poggio (1979) assumed that the same reasoning would apply to stereoacuity, which they argued also reflected the smallest filter giving a coherent signal. Data on stereoacuity at different pedestal disparities is given in Parker and Yang's paper, but the results do not show quite the same pattern. (Exactly the same paradigm was used to measure stereoacuity as for the disparity averaging experiment except that the two spatially separated patches within the stereogram, the test and comparison, both had a uniform disparity.) The rise in stereoacuity thresholds with pedestal disparity is much shallower than the extent of disparity averaging (in fact, Parker and Yang take the mean stereoacuity over the whole range for use in the criterion of disparity averaging). This result might be argued to indicate that coarse filters play more of a role in acuity at small disparities than Marr and Poggio supposed*.

---

* As discussed in section 1.5, other studies have found an "exponential" rise in thresholds with increasing pedestal disparity (Ogle, 1953; Blakemore, 1970b; Schumer and Julesz, 1984) but all

In general, Parker and Yang's results agree very well with Marr and Poggio's theory. However, they do not fit so well with the predictions of a hierarchical model. Because the position of fine scale features is recorded relative to coarse scale features in a hierarchical scheme, it cannot be argued that the fine scale features give rise to an "incoherent disparity signal" at a large pedestal disparity. It is possible that, because in most cases (i.e. apart from when viewing stereograms) features with a large disparity are blurred (accommodation is inappropriate), this might lead to a strategy in which blobs at a large disparity were not analysed down to the finest scale. The simpler heuristic "continue to switch out filters until the representation no longer changes" (at least within the focus of attention) is more attractive but would not explain the results.

Exposure duration was unlimited in Parker and Yang's experiment. In view of evidence presented later in this chapter for a coarse-to-fine analysis over the first second of viewing, it would be interesting to discover the effect of exposure duration on disparity averaging for the type of stimuli they used.

4.2.3 Related experiments

Other authors have used similar stimuli. Schumer (1979) investigated disparity averaging close to the fixation plane using random dot patterns in which odd rows were given a disparity while even rows remained in the background plane. (This is just one case in the space covered by Parker and Yang). Stevenson, Cormack and Schor (1989) investigated different types of resolution tasks using random dot stereograms in which, as in Parker and Yang's experiment, separate planes with different disparities were presented within one area of the pattern. Instead of using alternate rows to define the planes, Stevenson et al.. displayed the different planes on alternate frames (at a rate of 60 Hz). The task in their experiment that is most closely related to that used by Parker and Yang is one they called "superresolution". The subject was required to discriminate a single disparity plane from a stimulus containing two planes with slightly different disparities but the same average disparity (zero). Stevenson et al. obtained superresolution thresholds of between approximately 25 and 45 arcsec which is comparable with the limits of disparity averaging reported by Parker and Yang for targets with an average disparity of zero (between approximately 55 to 85 arcsec). The method used by Stevenson et al.. could be used to investigate the temporal frequency response of the mechanisms involved, although they did not pursue this. The method used in

---

these studies have used much larger pedestal disparities and their results are consistent with a slow change in acuity for pedestals up to 4 or 5 arcmin as Parker and Yang found.

both the experiment of Parker and Yang and the one described in this chapter are better suited to investigating the spatial properties of the putative mechanisms.

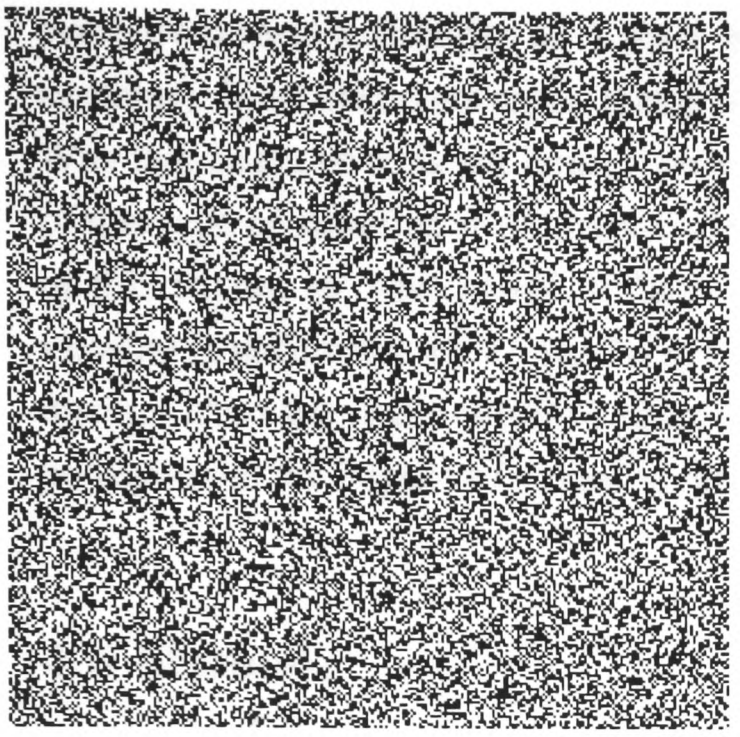
Nakayama and Silverman (1984) have studied very similar stimuli but in the motion domain (measuring the upper displacement limit for motion, d_{max}). They also point out that large cells, whose receptive fields cross several strips, "will not generate a directionally selective signal" (p 296). They do not specify whether they think this is because the cells receive equal and opposite motion signals (i.e. they represent the sum of coherent motion signals) or that they fail to find any consistent correlation over time (i.e. they signal incoherent motion). In the following experiment the distinction is an important one: a signal of zero disparity is quite different from an uncorrelated signal.

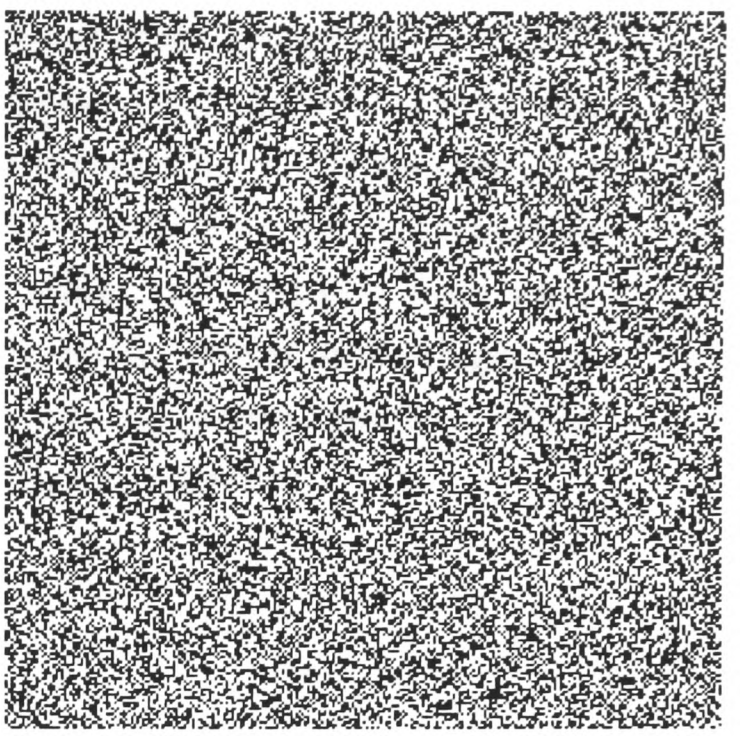
4.3 The rationale for this experiment

Parker and Yang's experiment was concerned with the finest filter that signalled a coherent match between left and right eye's image (at least, if the "monocular filter" model of their results is accepted). The experiment described in this chapter is concerned with the signal from the coarsest filter. A coarse monocular filter will blur together odd and even lines to a greater extent than fine filters and, as Parker and Yang argue, will signal the average disparity of the patch. The most interesting case is when this average value is zero, i.e. the coarse filters do not "see" a disparate square at all.

If stereo matching is a coarse-to-fine process as Marr and Poggio suggest then, at short exposure durations, a 2-plane stimulus with zero average disparity should be difficult if not impossible to see. As finer scale information is analysed the differences between left and right eye's images will be "discovered" and the stimulus should become apparent. Thus it should be possible to measure the time course of the coarse-to-fine process. That is what this experiment sets out to do.

Figure 4.1 illustrates the types of stimuli used in the experiment. The three squares in the stereogram are defined in three different ways: (a) a uniform disparity of 2 pixels, (b) odd rows have a disparity of 3 pixels, even rows a disparity of 1 pixel (called, for short a "+3,+1" stimulus) and (c) odd rows have a disparity of +1





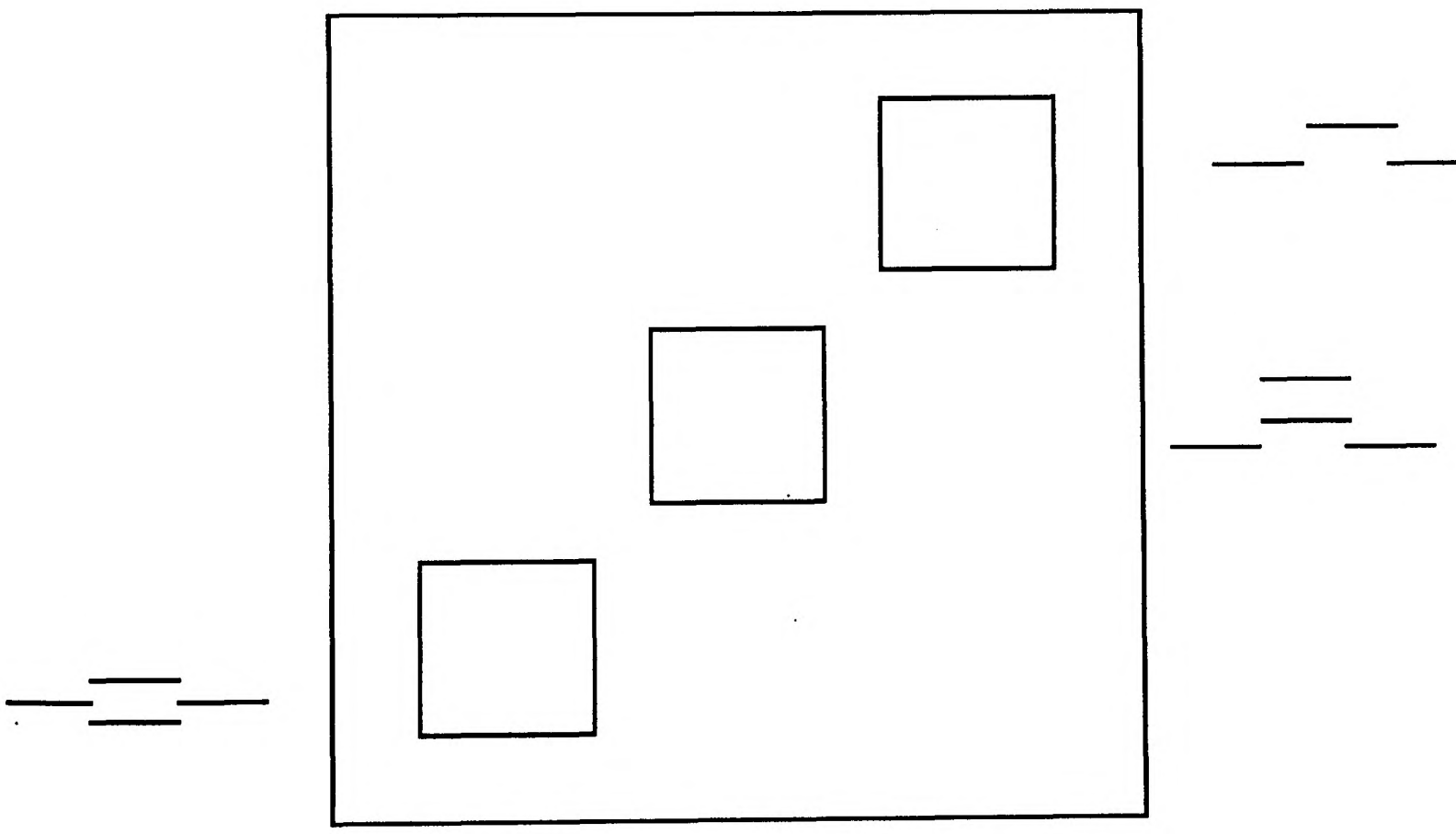


Fig 4.1

Readers who can fuse the random dot patterns at the top should see three squares, as illustrated below. The top right square is defined by a normal disparity shift of 2 pixels and should appear as a flat square floating above the background (for cross-eyed fusion). The middle square contains mixed disparities: odd rows have a disparity of 3 pixels, even lines a disparity of 1 pixel with respect to the background. It should appear to be at the same average disparity as the first square but with a "ruffled" or corrugated surface. The square at the bottom left is more difficult to make out. It is made up of odd rows with a disparity of +1 pixel (in front), even rows with a disparity of -1 pixel (behind), i.e. it has the same average disparity as the background. In the text this is referred to as the "+1,-1" disparity stimulus.

pixel, even rows a disparity of -1 pixel with respect to the background (a "+1,-1" stimulus). Most observers report that this last square is more difficult to make out. Why this might be so is not immediately obvious, even considering the properties of this type of stimulus at different scales, as discussed above, for the perception of this square as "less solid" or "less easily picked out" remains, even when viewing is unlimited. This issue (which is discussed at the end of the chapter) is distinct from the main question addressed by this experiment, that is, how the perception of the 3 types of stimuli varies with exposure duration.

4.4 Methods

4.4.1 Subjects

Subjects were all young adults with normal (6/6) or corrected-to-normal vision.

4.4.2 Apparatus

Stimuli were generated on a Macintosh II computer and displayed on two continuous grey-scale monitors set up in a Wheatstone configuration at a distance of 57 cm (illustrated in figure 4.2). At this distance the pixel size was 2 arcmin.

Subjects sat with their head in a chin rest in front of two front-silvered mirrors mounted on holders screwed to the bench. The angle between the mirrors was 90°, i.e. 45° either side of the median plane. The monitors were placed within tightly fitting wooden surrounds fixed to the bench, at right angles to, and equidistant from, the median plane, as shown in figure 4.2. The centre of the screen was slightly in front of the position of the observer's eye which is consistent with the viewing geometry of a virtual screen at 57 cm (see figure 4.2). As a result, the line of sight is slightly oblique with respect to the monitor screen, and hence the pattern of vertical disparities is identical to that created by a real screen at this distance. The fine adjustment of the monitor and mirror positions was achieved by comparing the apparent depth of a rod placed at 57 cm from the observer in the median plane with the stereoscopic image of a vertical line drawn in the centre of the two screens.

A separate 8-bit video card was used to display stimuli on each monitor. Subjects responded to trials in the experiment by closing one of two switches. The signal was relayed to the computer via a National Instruments digital input/output board.

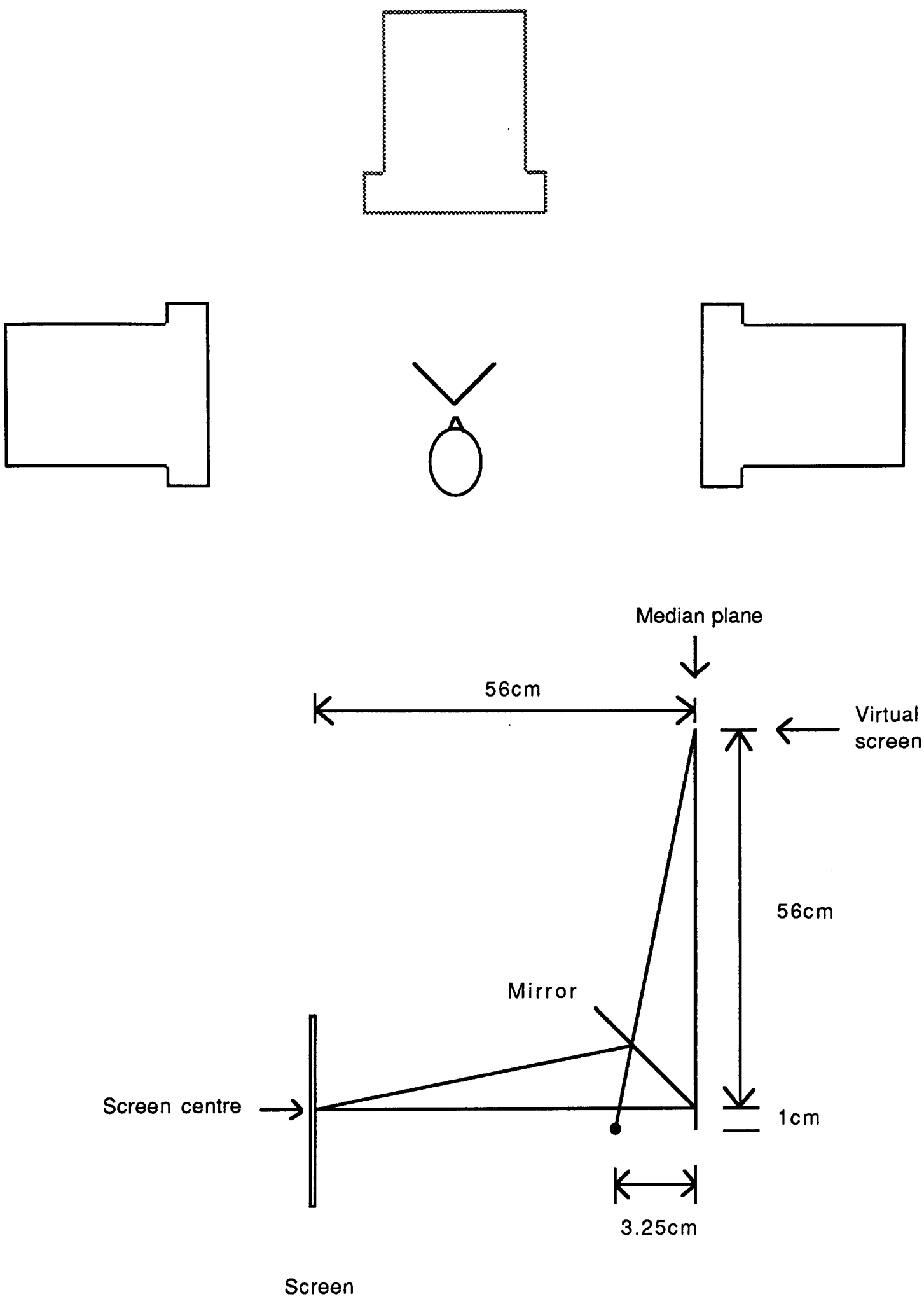


Fig 4.2

A plan view of the Wheatstone set up used to display stimuli in all the experiments described in this thesis. Subjects sat with their head in a chin rest in front of two front-silvered mirrors which were set at an angle of 45° either side of the median plane. The monitors were fixed at right angles to, and equidistant from, the median plane as shown so that a virtual screen is seen directly ahead of the observer at a distance of 57cm (dotted outline). The geometry of the set up is shown below (left half only) is shown below. The left eye is shown as a black dot 3.25 cm from the median plane. The line of sight to the centre of the virtual screen and, reflected in the mirror, to the centre of the real screen are shown.

4.4.3 Stimuli for experiment I and II

On each trial a 50% density, 1-bit random dot stereogram (256 by 256 pixels, i.e. 8.5 by 8.5°) was displayed containing a disparity-defined rectangular "target" region (as illustrated in figure 4.3). (The disparity used to define the target varied between experiments). The luminance of the bright pixels was 32 cd/m^2 and the dark pixels 0.12 cd/m^2 . The luminance of the screen surrounding the stereogram was also 0.12 cd/m^2 .The room in which the experiment was carried out was dimly lit.

The stimulus was displayed for a given exposure duration (which was constant within one experimental run) after which a mask, consisting of a different random dot stereogram (correlated, zero disparity) was presented. The mask remained on the screen until the subject gave their response to the trial, and this triggered the next display.

The stimuli used for each trial and the mask were 1-bit images. Alternation between the stimulus and mask was achieved by changing the colour look-up table, which occurs between frames. The minimum exposure duration using this technique was 4 frames (i.e. 60 ms, frame rate was 66.7 Hz). Exposure duration was controlled from the program using the computer's internal clock which runs at a rate of 50 Hz (a "tick" is 16.7 ms). This led to an occasional extra frame being displayed (as measured using an oscilloscope attached to the video output). The exposure duration shown in the results are multiples of the frame duration (15 ms), they do not take account of the occasional extra frame (which occurred on less than 1 in 10 of presentations).

4.4.4 Task

Subjects had to identify, in a 2-alternative (single interval) forced choice procedure, whether the target rectangle was horizontally or vertically elongated. The independent variable (or "cue") was the height:width ratio of the rectangle. To prevent subjects from using either the height or the width alone to make their judgements, the position and size of the target was jittered randomly between trials. (The centre of the target rectangle was on average at the centre of the random dot pattern but could vary in its horizontal or vertical position by up to +/-16 pixels. The size of the target was on average 64 by 64 pixels but varied between trials by up to +/- 16 pixels. That is, the notional square target to which a cue was added might be as small as 48 by 48 pixels or as large as 80 by 80 pixels.)

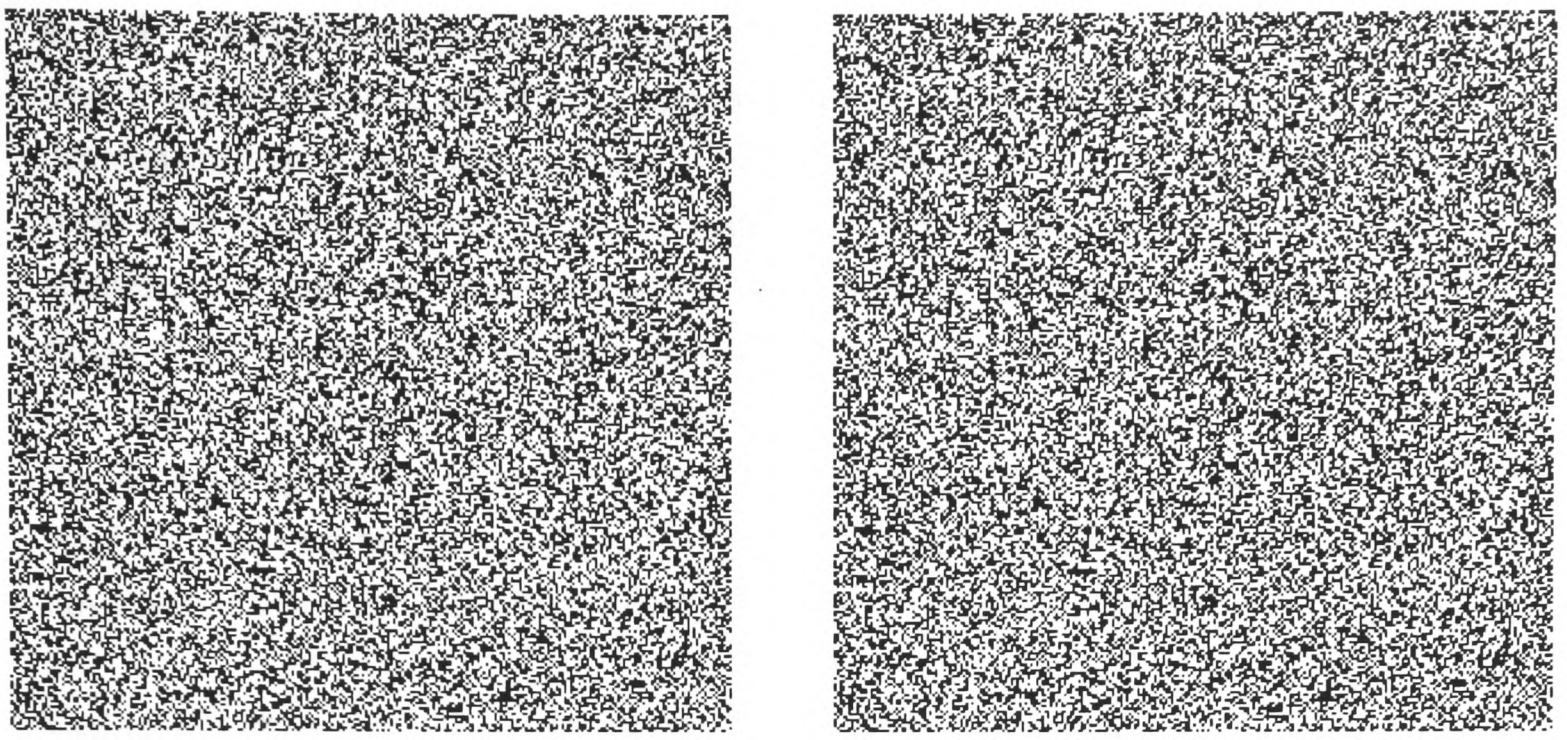


Fig 4.3

The stereogram above shows a single disparity-defined target rectangle, as illustrated below. The subject's task was to identify the target as horizontally or vertically elongated. The more square the target the more difficult the task. The threshold height:width ratio was obtained for a range of exposure durations and stimulus types.

4.4.5 Psychometric procedure

The threshold height:width ratio required to accurately perform the task was determined over an experimental run of 147 trials. In each run 7 height:width ratios were used (each presented 21 times, in random order). These ratios were equally spaced (on a log scale) and centred around a ratio of 1 (square). An appropriate range of ratios for each condition was determined in a pilot run of 50 trials. An example set of results is shown in figure 4.4. A cumulative Gaussian was fitted to the data using probit (according to the method described by Finney (1971), implemented in a program written by Dr. M. Treisman). The anti-log of standard deviation of the Gaussian was defined as the threshold height:width ratio.

Estimates of the standard error were obtained by dividing the run into three blocks of 49 trials (in one block each cue was presented seven times) and using probit to determine the standard deviation for each block. The mean of these values usually corresponded very closely to the standard deviation obtained using all 147 trials. The standard error of the mean is plotted for each data point in figures showing the results.

4.5 Experiment I: Time course

In the first experiment, the target rectangle was defined in one of the three ways described above and illustrated in figure 4.1, i.e either having a uniform disparity of 2 pixels or a mixed disparity of 3 and 1 pixels or a mixed disparity of +1 and -1 pixels.

4.5.1 Results

Figure 4.5 shows results for 2 observers. Threshold height:width ratios are plotted against exposure durations for the three target types. Shape discrimination thresholds for an ordinary 2 pixel disparity target vary very little with exposure duration, down to 60 ms. The results for the 3 and 1 pixel disparity stimulus are broadly similar. Thresholds are always slightly worse and for one subject are significantly worse at the shortest exposure. The results for the target defined by +1 and -1 pixel disparity are very different. Thresholds are higher for all but the longest exposure. At exposures of about 100 ms there is a steep rise in thresholds. Subjects report that at these short exposures it is very difficult to see the "+1,-1" disparity target at all.

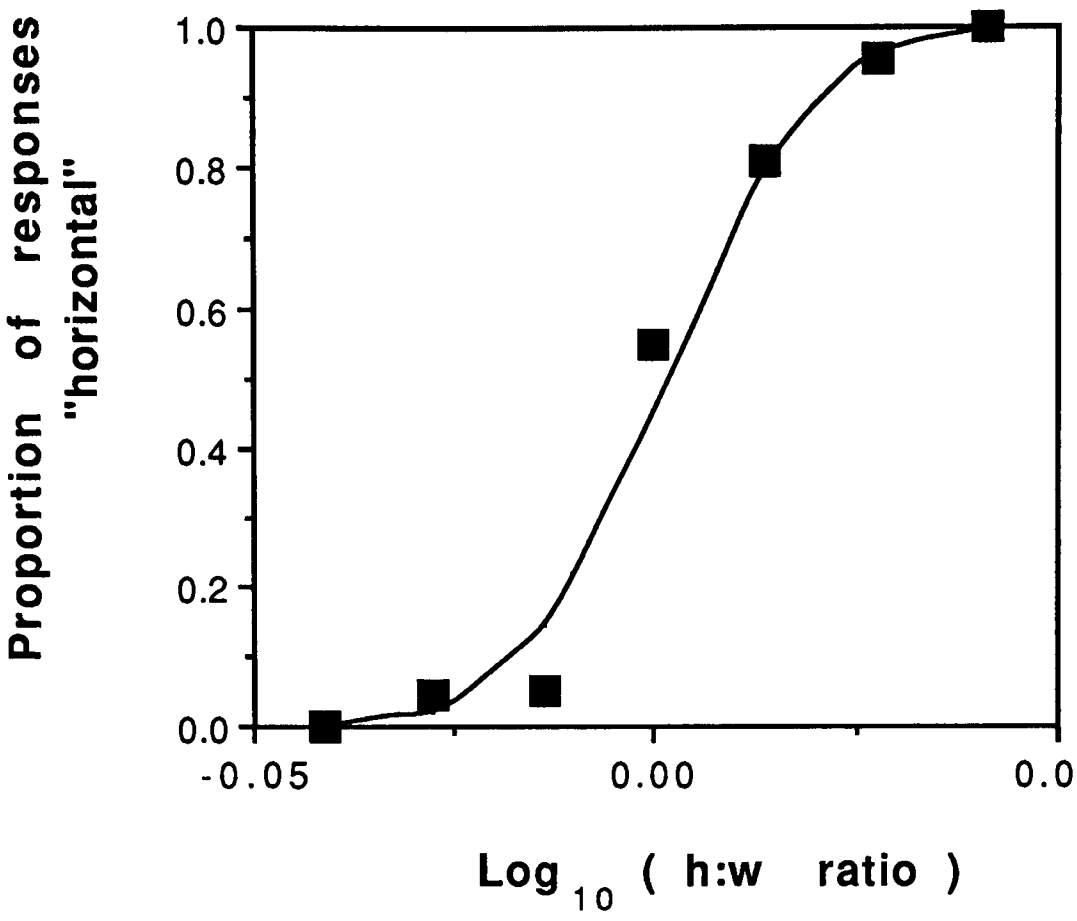


Fig 4.4

Data from one experimental run as an example. The proportion of trials in which the subject saw the target as horizontally elongated is plotted against the log of the target's actual height:width ratio. A cumulative Gaussian has been fitted to the data by probit. The inverse log of the standard deviation of this function is defined as the threshold height:width ratio

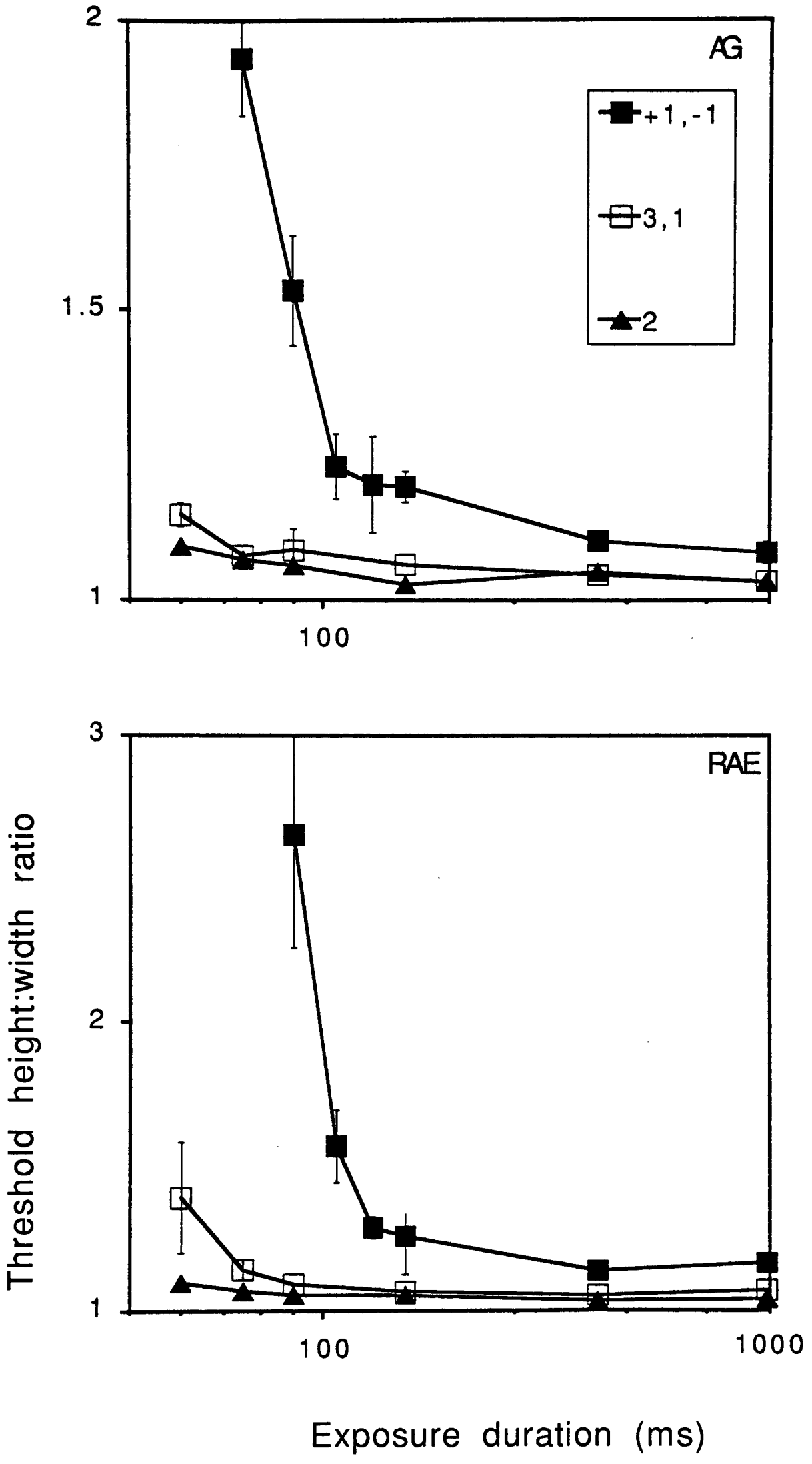


Fig 4.5

Results for experiment I are shown for two observers (top and bottom). Threshold height to width ratios are plotted as a function of exposure duration for the three stimulus types, i.e. the target defined by odd and even rows having a disparity of +1 pixel and -1 pixel with respect to the background (filled squares); odd and even rows having a disparity of 3 and 1 (crossed) disparity with respect to the background (open squares) and the target defined by a uniform 2 pixel disparity (triangles). Error bars indicate ± one standard error

4.6 Experiment II: Varying strip height

In this experiment the height of the strips in the "+1,-1" mixed disparity stimulus was altered. (A strip height of 2 pixels (4 arcmin), for instance, means that two consecutive rows have a crossed disparity of 1 pixel, the next two rows an uncrossed disparity of 1 pixel and so on.)

4.6.1 Results

Results for 2 observers are shown in figure 4.6. Results from the previous experiment (for the "+1,-1" condition only) are shown as a dotted line (i.e. a strip width of 1 pixel). Increasing strip height to two pixels moves the curve to the left, i.e. the task is possible at slightly briefer exposures. Increasing the strip width to *eight* moves the curve even further to the left. (In the latter condition there were only about 4 complete corrugations within the target region.)

4.7 Experiment III: Control stimuli

The results for the "+1,-1" target in experiment I are very different from those obtained for the uniform disparity target (2 pixels crossed disparity) and the "+3,+1" target. It is important to determine whether the difference is simply a consequence of the smaller disparities used in the "+1,-1" target. The results of experiment II, which show that increasing the strip height in a "+1,-1" target improves thresholds at short exposures, suggests that when strip height is equal to target height (i.e. a uniform +1 disparity target) performance should be considerably better than when strip height is 1 pixel as in experiment I. In this experiment, thresholds were measured for a uniform +1 disparity target at a range of exposure durations (60 ms to 500 ms) for one subject and for a second subject at 75 ms (i.e. a short exposure where predictions about performance are most critical).

In addition, thresholds were measured for a target in which odd rows had a crossed disparity of 3 pixels and even rows an uncrossed disparity of 3 pixels ("+3,-3"). The target in this stimulus is similar to the "+1,-1" stimulus in having zero average disparity and, it could be argued, in producing equal excitation in crossed and uncrossed disparity detectors. Thresholds were also measured for a target in which odd lines had a crossed disparity of two pixels and even rows zero

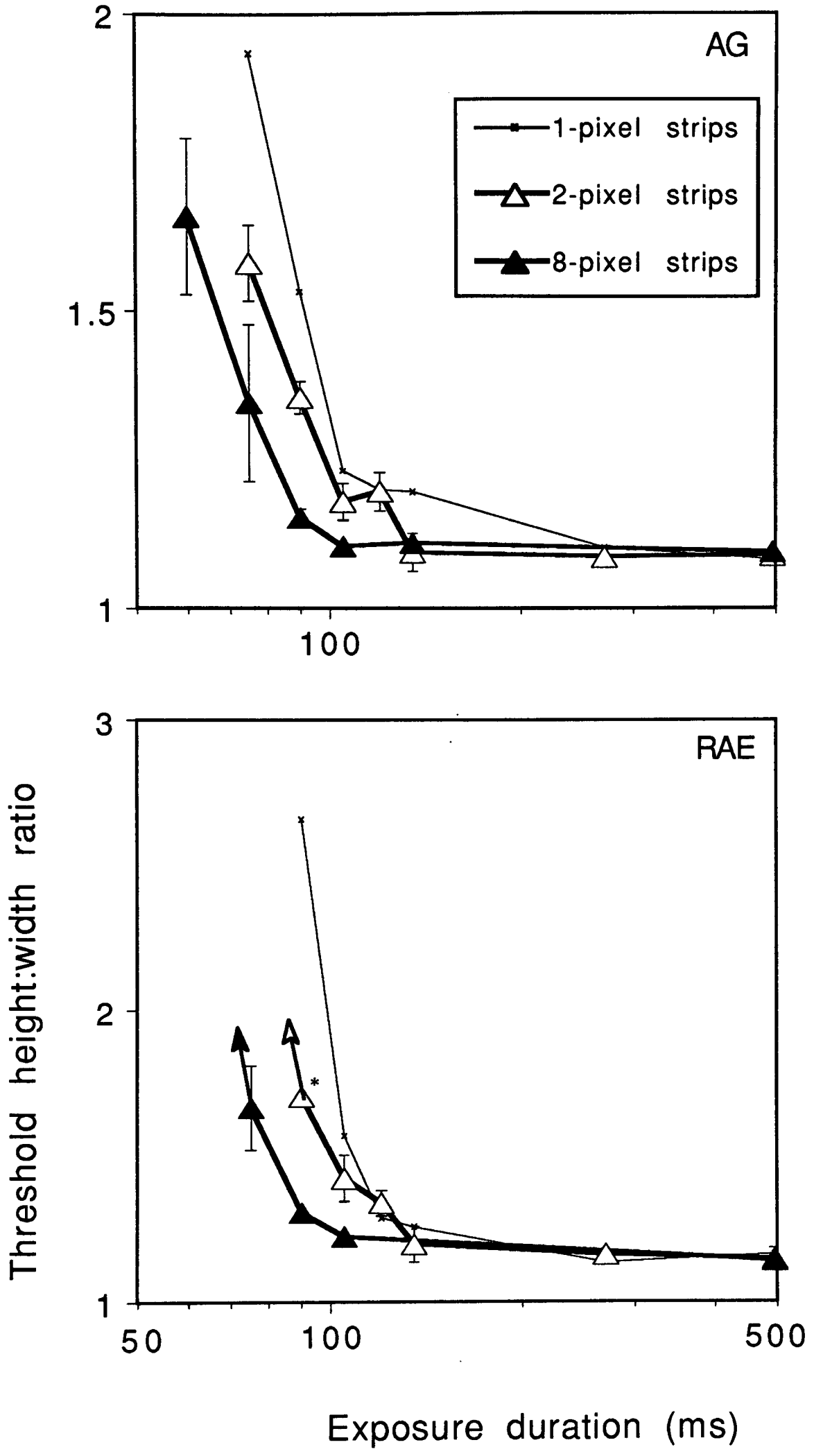


Fig 4.6

As in figure 4.5, threshold height to width ratios are plotted as a function of exposure duration for two observers. Results are all for "+1,-1" disparity targets for different strip heights, i.e. a strip height of 8 pixels, 2 pixels and, taken from experiment I, shown as a thin line, 1 pixel. Results for subject AG are shown above, for RAE below. * The error bar for this point was too large to be shown on this graph. Variation in performance at or near the steep rise in thresholds is very large.

disparity with respect to the background ("+2,0"). This stimulus has mixed disparities *and* a small average disparity, thus the effects of a small disparity and a mixed disparity can be examined both together or in isolation. Finally, thresholds were measured for a stimulus in which the dots in the left and right eyes images were uncorrelated within the target region. The background dots, as in all the other conditions, were correlated (zero disparity).

4.7.1 Results

The results are shown in figure 4.7a for subject AG (showing data for exposures between 60 and 1000 ms) and in figure 4.7b for subject RAE (exposure duration 75 ms). Thresholds for the "+1,-1" and +2 uniform disparity targets are re-plotted on the graphs for comparison.

Thresholds for the uniform +1 pixel disparity target for exposure durations down to 60 ms do not show the same pattern as the +1,-1 disparity target indicating that it is not the small disparity of the +1,-1 target *per se* that causes the rise in thresholds at brief exposures.

On the other hand, the threshold for the "+2,0" target rises at the briefest exposure (figure 4.7a) indicating that targets that have *both* a small average disparity *and* contain mixed disparities are difficult to see at brief exposures. The most extreme example of this is the "+1,-1" target.

Thresholds for both the "+3,-3" and uncorrelated target are relatively low even at the briefest exposure for both subjects. It is suggested in the discussion section that these two types of stimulus may have similar properties at a coarse scale.

4.8 Experiment IV: Low-pass stimuli

In the final experiment low-pass filtered stimuli were used to address the question of whether the removal of high frequency information affects performance for the "+1,-1" target more than for a uniform disparity target. Earlier in this thesis the use of filtered stimuli has been criticised because of the problem of distinguishing whether a result is due to the properties of the stimulus or of the visual system. In this experiment it is the properties of the stimulus that are of interest, so it is

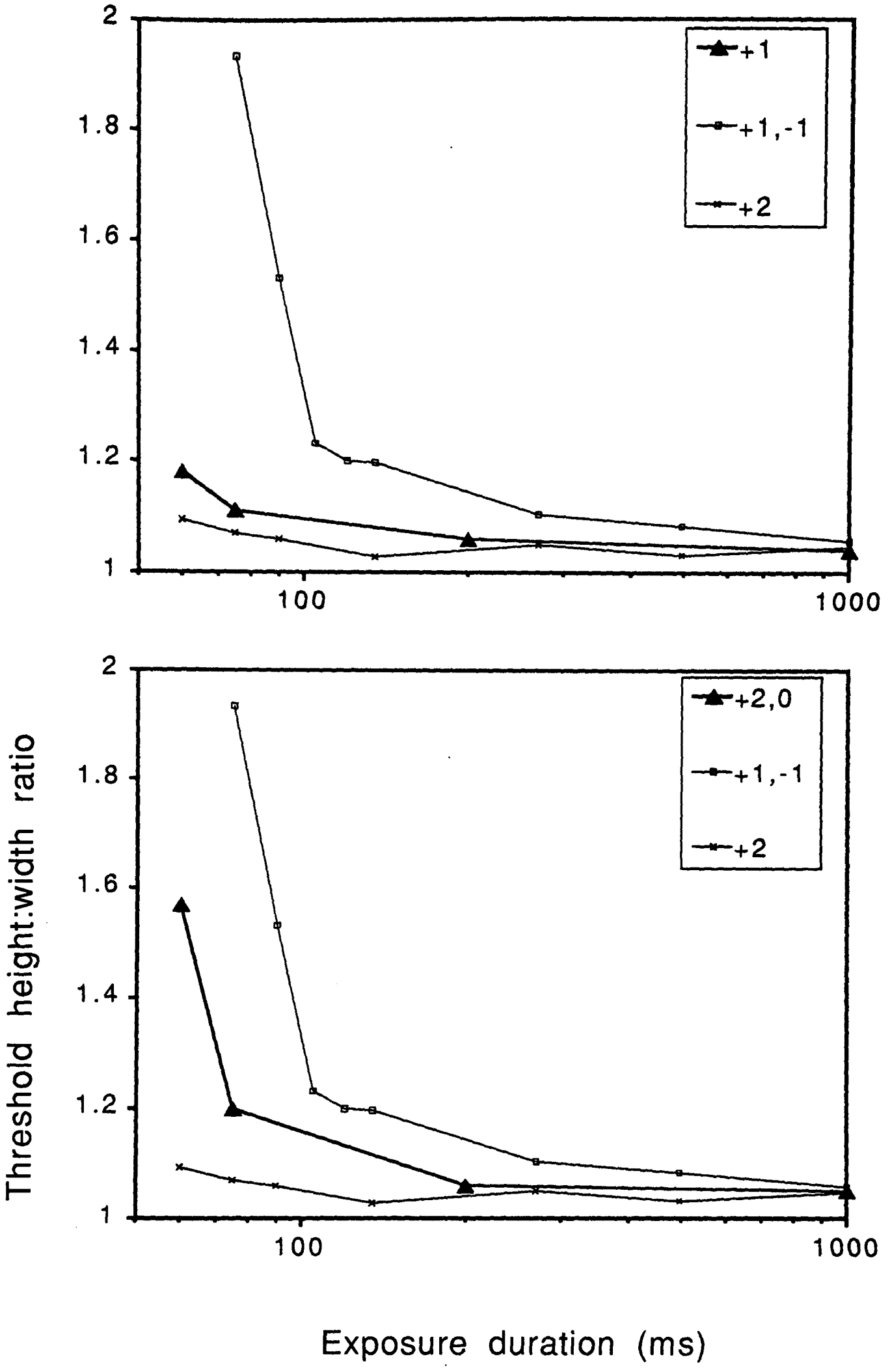
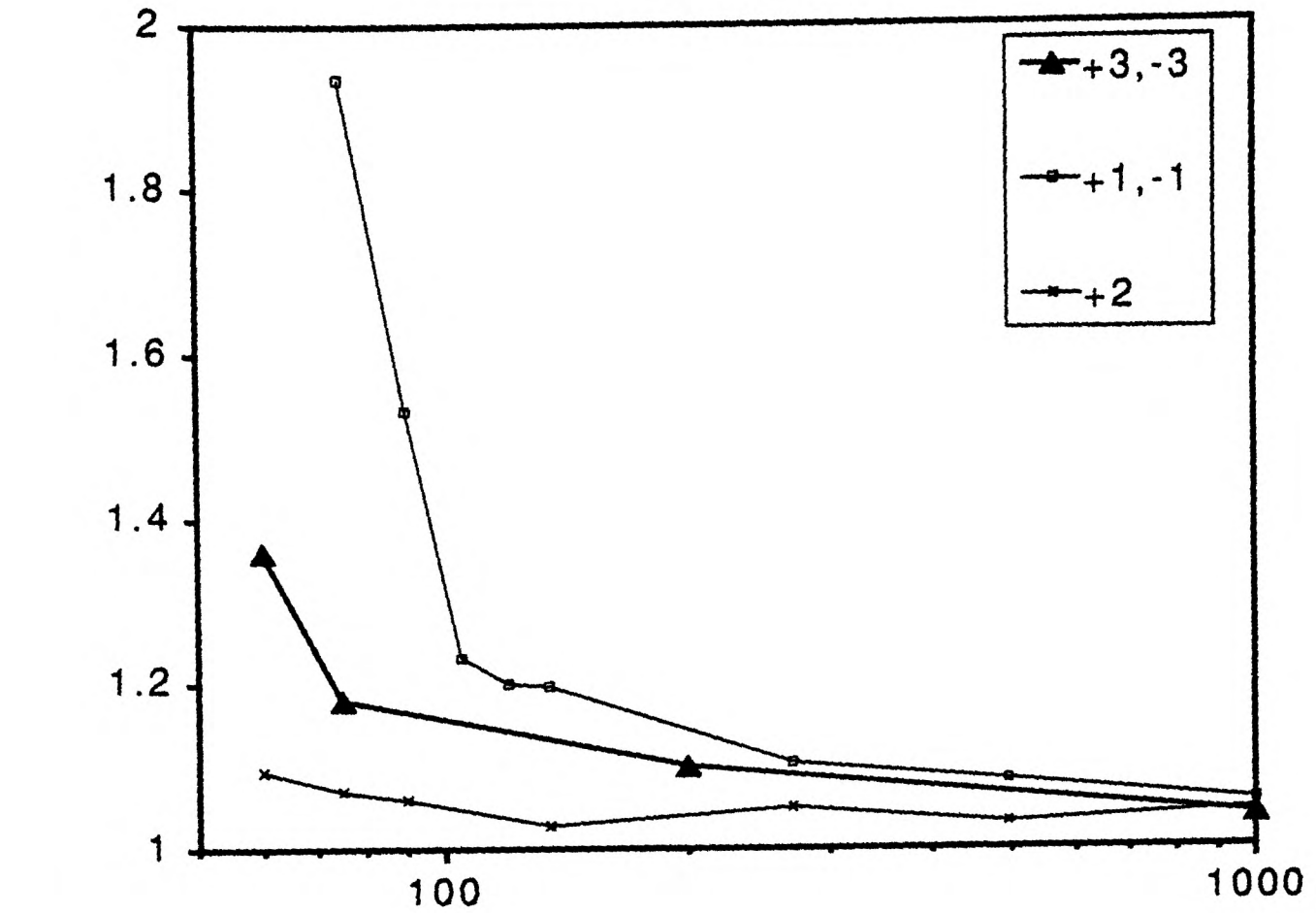


Fig 4.7a

Results for experiment III, subject AG. As in figures 4.5 and 4.6, threshold height:width ratios are plotted against exposure duration (ms). Data are shown for a target with a uniform crossed disparity of 1 pixel (top), and a target in which odd rows had a disparity of two pixels, even rows a disparity of zero with respect to the background (bottom). Data from figure 4.5 ("+1,-1" and "+2" targets) are shown as thin lines for comparison.



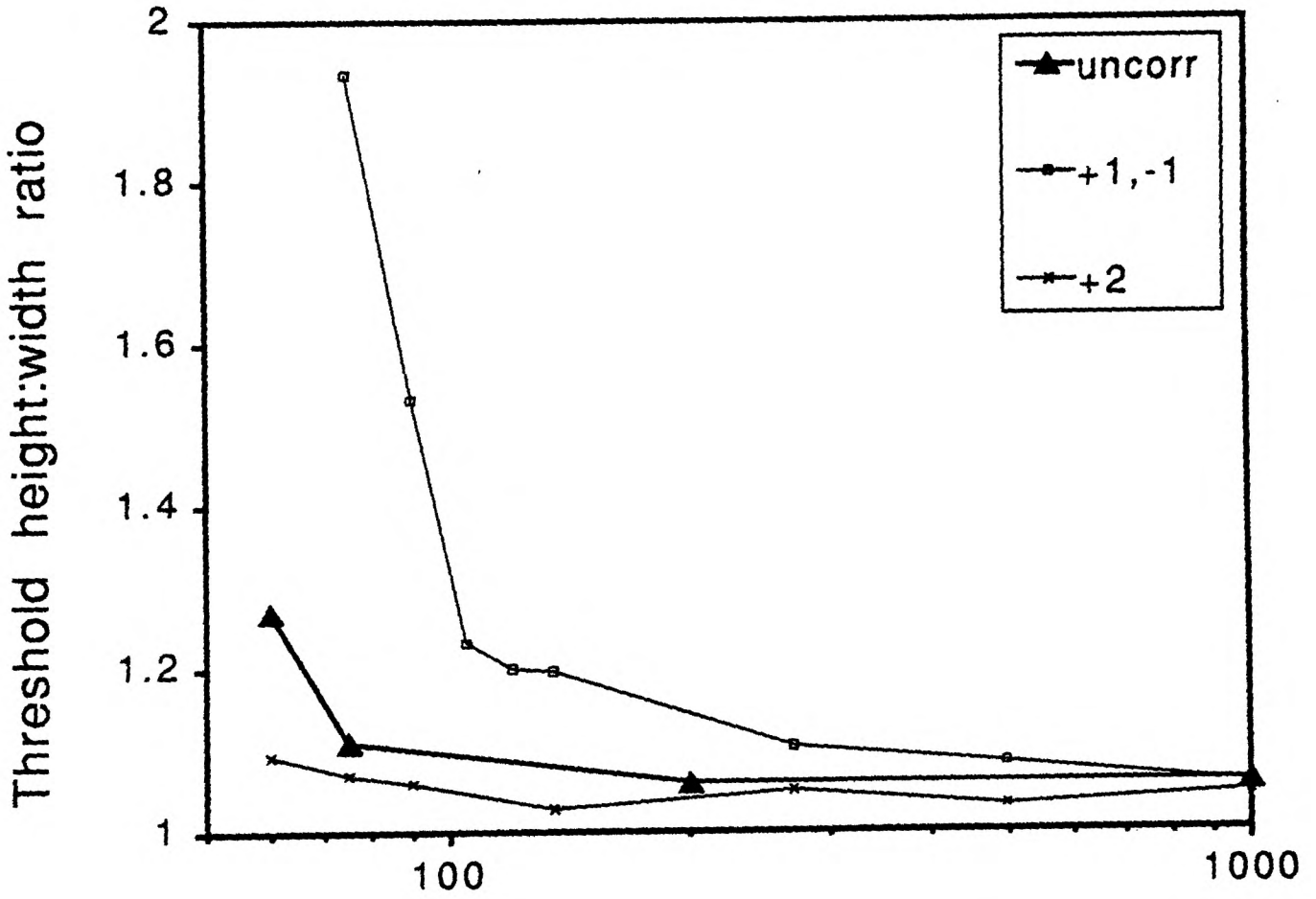


Fig 4.7a (cont)

Data are shown for a target in which odd rows had a crossed disparity of 3 pixels, even rows had an uncrossed disparity of 3 pixels (top), and for a target in which the pixels in the left and right eyes images were uncorrelated (bottom).

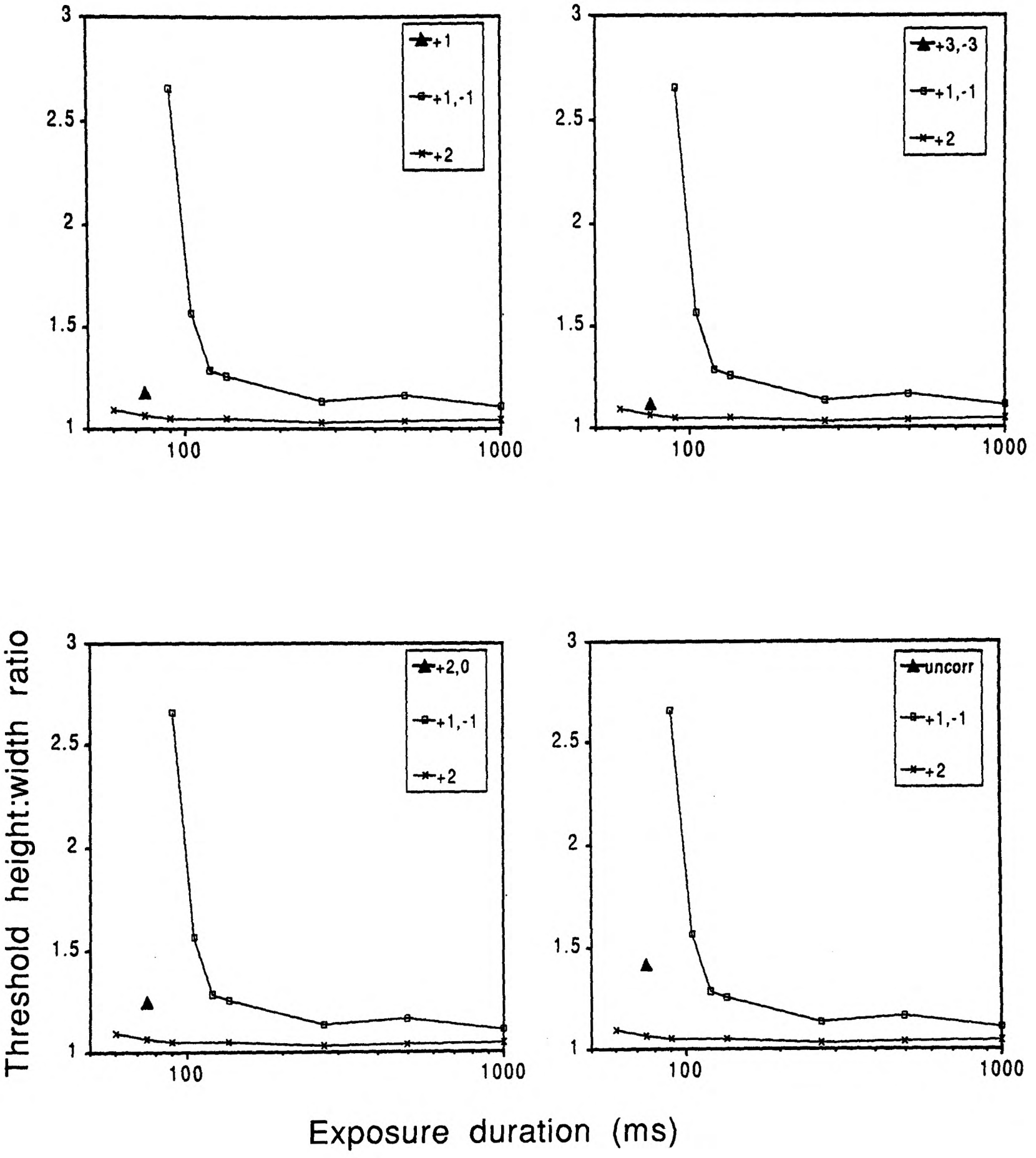


Fig 4.7b

Results for experiment III, subject RAE. As in figures 4.5 and 4.6, threshold height:width ratios are plotted against exposure duration (ms). Data is shown for a 75 ms exposure for: a target with a uniform crossed disparity of 1 pixel (top left); a target in which odd rowshad a disparity of two pixels, even rows a disparity of zero (bottom left); a "+3,-3" target (top right); and an uncorrelated target (bottom right). The dotted lines show data from figure 4.5 ("+1,-1" and "+2" targets) for comparison.

perhaps best described as an "experimental model". A preferable solution would be to use an ideal observer model (e.g. Barlow, 1978; Morgan, Hole and Glennerster, 1991). An attempt was made to construct one but the problem proves to be quite difficult. The task is not one of detection but a spatial judgement and many assumptions would have to be made about how this was being done in the visual system. Instead, the approach taken was to present to subjects filtered versions of the stimuli used in the experiment I. Subjects were asked to make the the same judgement as in the exposure duration experiment (i.e. whether the target was horizontally or vertically elongated). The exposure duration subjects were allowed was long and it was assumed that high thresholds reflected poor information in the stimulus. An alternative approach to modelling, which illustrates the position of centroids in stimuli filtered at a coarse scale, is described in section 4.9.

4.8.1 Stimuli for experiment IV

The stimuli were 128 by 128 pixels (i.e. 4.25 by 4.25°, pixel size 2 arcmin). The targets in the pre-filtered stimuli were generated in the same way as in experiment I (except that the vertical and horizontal jitter applied to the target's position was less, i.e. a jitter of up to ± 8 pixels, and target size was not jittered). Seven height:width ratios were used in each experimental run (as in experiment I). For each height:width ratio three different filtered patterns were generated and each pattern was displayed twice, in random order, i.e. the total run consisted of 42 trials. The pre-filtered stimuli were of two types, either uniform 2 pixel disparity or mixed, "+1,-1" disparity targets. (A "+3,+1" stimulus was not used as the results from experiment I for this condition were essentially the same as for the uniform target of 2 pixels.)

The stimuli were filtered using a Gaussian filter with a space constant of 2.8, 5.6, 8.4 or 11.2 arcmin. The equation used for the filters was:

$$\nabla^2 G\,(r,f,\theta) = e^{-r^2/2f^2}$$

where r is the distance and θ the direction from the centre, f is the space constant (standard deviation of the Gaussian). Details of the filtering process are given in appendix B.

The experiment was carried out using two different colour look-up tables and data are presented for both conditions. In the first case, an uncorrected (linear) look-up

table.was used which meant that the screen had a non-linear gamma function (shown in appendix A). In the second case, the colour look-up table was corrected to give a linear gamma function. Appendix A describes how this was done. In each case the filtered image was presented with the same Michelson contrast (maximum and minimum luminance values of 32 cd/m^2 and 0.12 cd/m^2 respectively, i.e approaching 100% contrast) and contained 128 grey levels. The background luminance was 0.12 cd/m^2, as in experiment I.

The exposure duration in the first experiment (uncorrected gamma) was 500 ms. In the second experiment it was up to 2 seconds, although subjects could respond before this time, triggering the next display. A 128 by 128 pixel, 1-bit, 50% density random dot mask (correlated, zero-disparity) was displayed between trials, as in experiment I.

4.8.2 Results

Figure 4.8 shows threshold height:width ratios plotted as a function of the standard deviation of the Gaussian filter used to low-pass filter the stimulus. Note that blur decreases along the abscissa. Data for the condition in which the gamma function of the screen was uncorrected is shown in figure 4.8a, and for the linear gamma function in figure 4.8b. There appears to be no systematic difference between the two conditions.

There is comparatively little effect on thresholds for the uniform disparity stimulus for the range of blur used (from σ = 2.8 to σ = 11.3 arcmin). However, for the mixed disparity "+1,-1" stimulus there is a sharp rise in thresholds over a relatively narrow range of blur. This mirrors the results of experiment I, and suggests that the effects of short exposures may be modelled by assuming that the scale of analysis changes over the first second of viewing.

4.9 Model

In this section the hypothesis that the effects of exposure duration can be explained by a change in the scale of analysis over time is examined. Other possible explanations for the pattern of results in experiments I - III are considered in section 4.10.

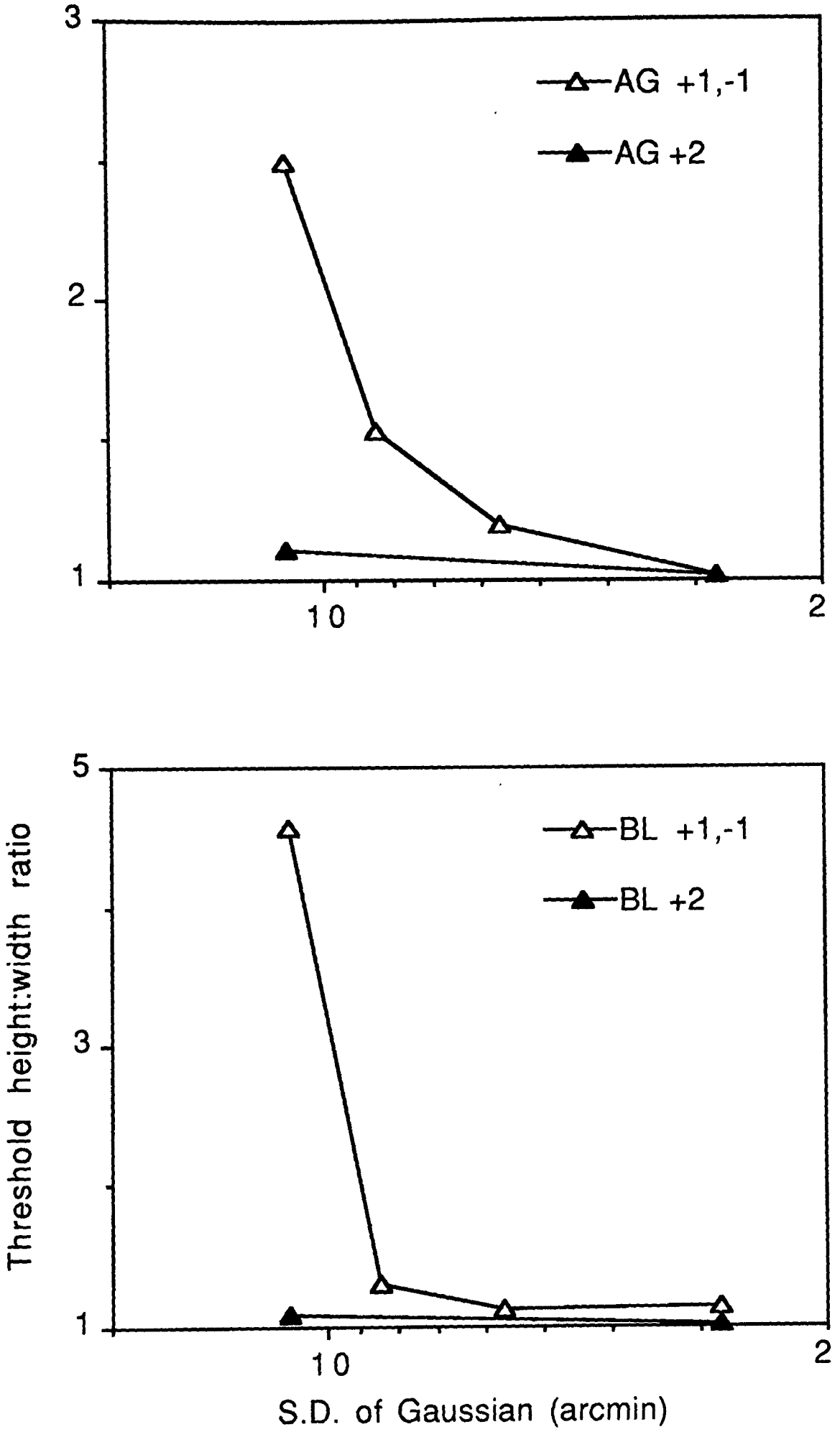
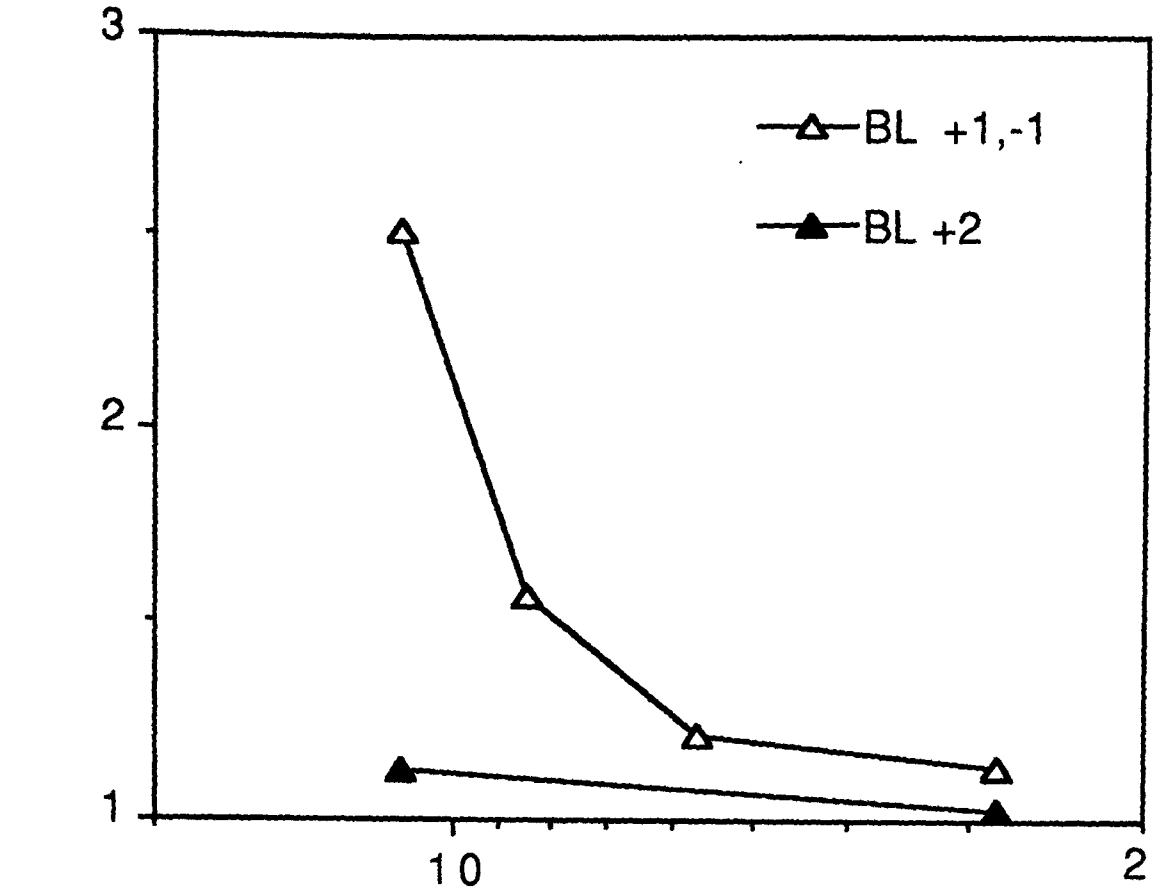


Fig 4.8a

Results for experiment IV for two subjects (BL and AG) using an uncorrected gamma function for the screen luminance look-up table (see methods). Threshold height:width ratio is plotted against the standard deviation of the Gaussian used to filter the image. The (pre-filtered) target was defined either by odd and even rows having a disparity of +1 and -1 pixels or by a uniform disparity of 2 pixels. The exposure duration was 500 ms.



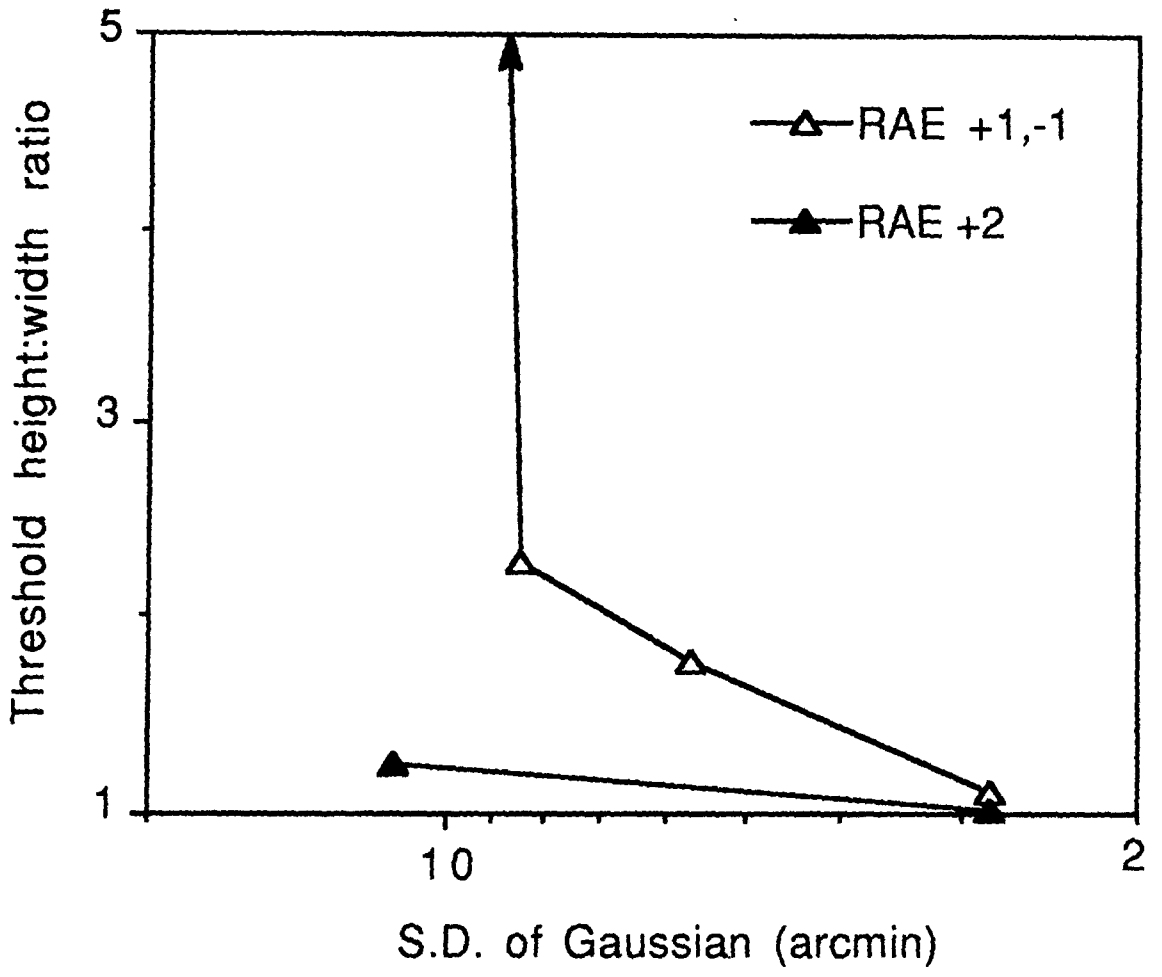


Fig 4.8b

Results for experiment IV for two subjects (RE and BL) using a linear gamma function for the screen luminance look-up table (see methods). Threshold height:width ratio is plotted against the standard deviation of the Gaussian used to filter the image. The (pre-filtered)target was defined either by odd and even rows having a disparity of +1 and -1 pixels or by a uniform disparity of 2 pixels. The exposure duration was 2 seconds. Subject RAE could not perform the task for a blur of 11.3 arcmin (nominally, his threshold was 102).

4.9.1 Spatial modelling

The difference between the "+1,-1" stimulus and a uniform disparity shift when filtered at a coarse scale is illustrated, very schematically, in figure 4.9. On the left is shown a group of black pixels that might appear in the left eye's image. The outline represents some kind of blurring by a coarse filter. On the right is the right eye's view of the same pixels if they formed part of the "+1,-1" disparity target and again their coarse outline is shown. Below, on the left, is a diagram illustrating that the blobs would have a very similar shape and position. The diagram on the right illustrates a normal disparity shift.

A more detailed illustration of the same point is made in figure 4.10. Figure 4.10 (a) shows left and right eyes' images for a target defined by a uniform disparity shift of 2 pixels and a "+1,-1" stimulus. Figure 4.10 (b) shows one of these images filtered at a coarse scale (with a Laplacian of Gaussian filter of space constant of 16 pixels, i.e. 32 arcmin at the viewing distance used in the experiment). The filtered output has been half-wave rectified (i.e. only the positive response is shown) and in white the (1-D) centroids of the "blobs" have been marked. The equation for a centroid is :

$$P_i = \frac{\int_{Z_{c_i}}^{Z_{c_{i+1}}} x \,.\, R(x) \,.\, dx}{\int_{Z_{c_i}}^{Z_{c_{i+1}}} R(x) \,.\, dx}$$

where Z_{c_i} and $Z_{c_{i+1}}$ are the positions of adjacent zero-crossings and R(x) is the response at point x. Details of the calculation of centroids are given in Appendix B. The choice of primitive is not critical, zero-crossings would do equally well in marking the position of blobs at a coarse scale. Figure 4.10 (c) illustrates how, for the uniform disparity stimulus, these centroids are all shifted by 2 pixels within the target region. Above, all the centroids are shown. Those that are in the same position in the left and right eyes' image are drawn in light grey, those that are shifted are shown in black or white (black marks the centroid position in the right eye, white the left eye). Below, only those centroids that are shifted (and hence indicate the presence of the target) are shown. Figure 4.10 (d) shows the "+1,-1" stimulus in the same way. This illustrates very well the point that in this case left

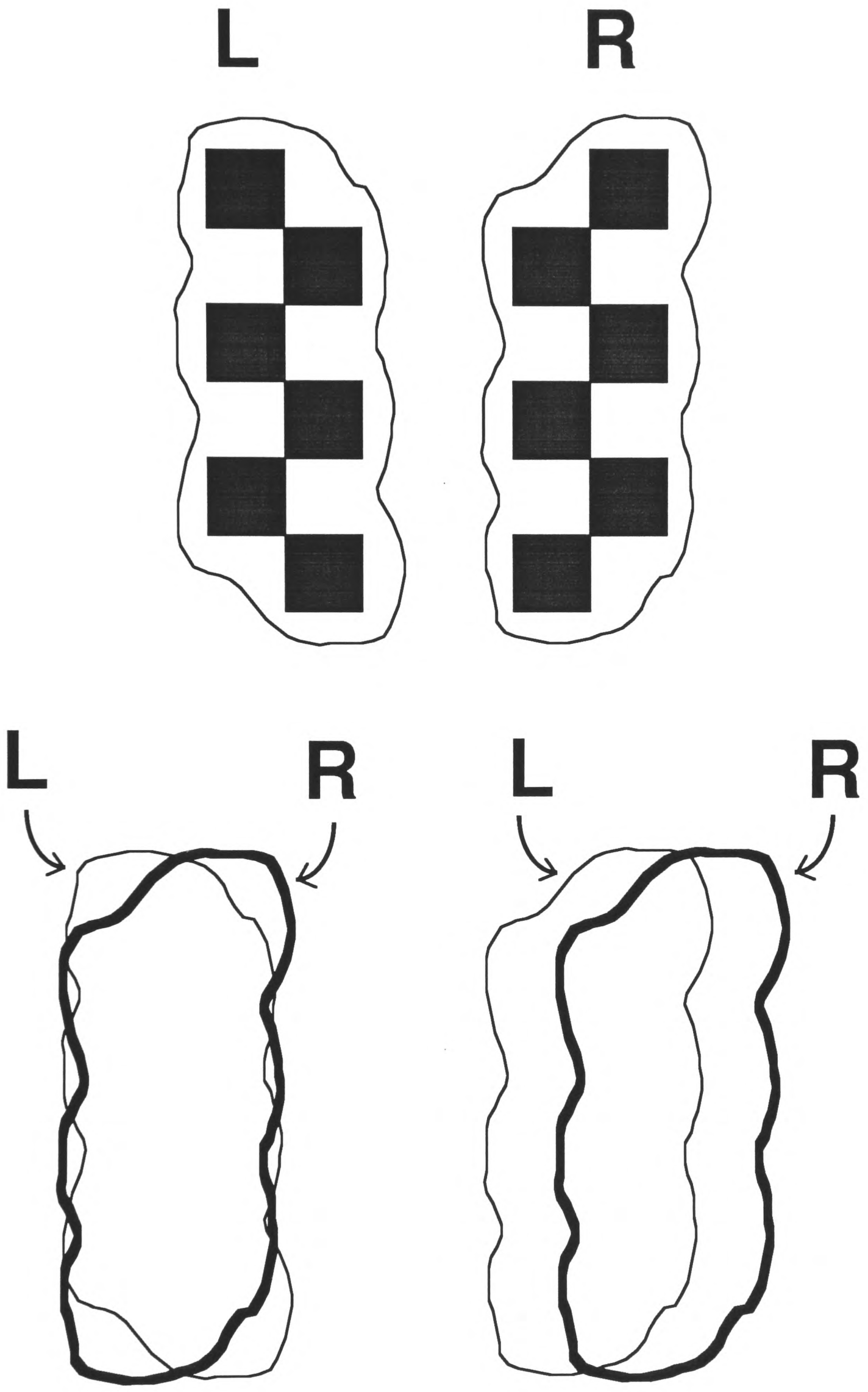


Fig 4.9

At the top, on the left is shown a group of black pixels that might appear in the left eye's image and an outline of their shape at a coarse scale. If a disparity of +1 pixel is applied to even rows and -1 pixel to even rows then the right eye's image, at a coarse scale, (shown on the right) is very similar in shape and position (compared below left). The diagram on the right illustrates a normal disparity shift.

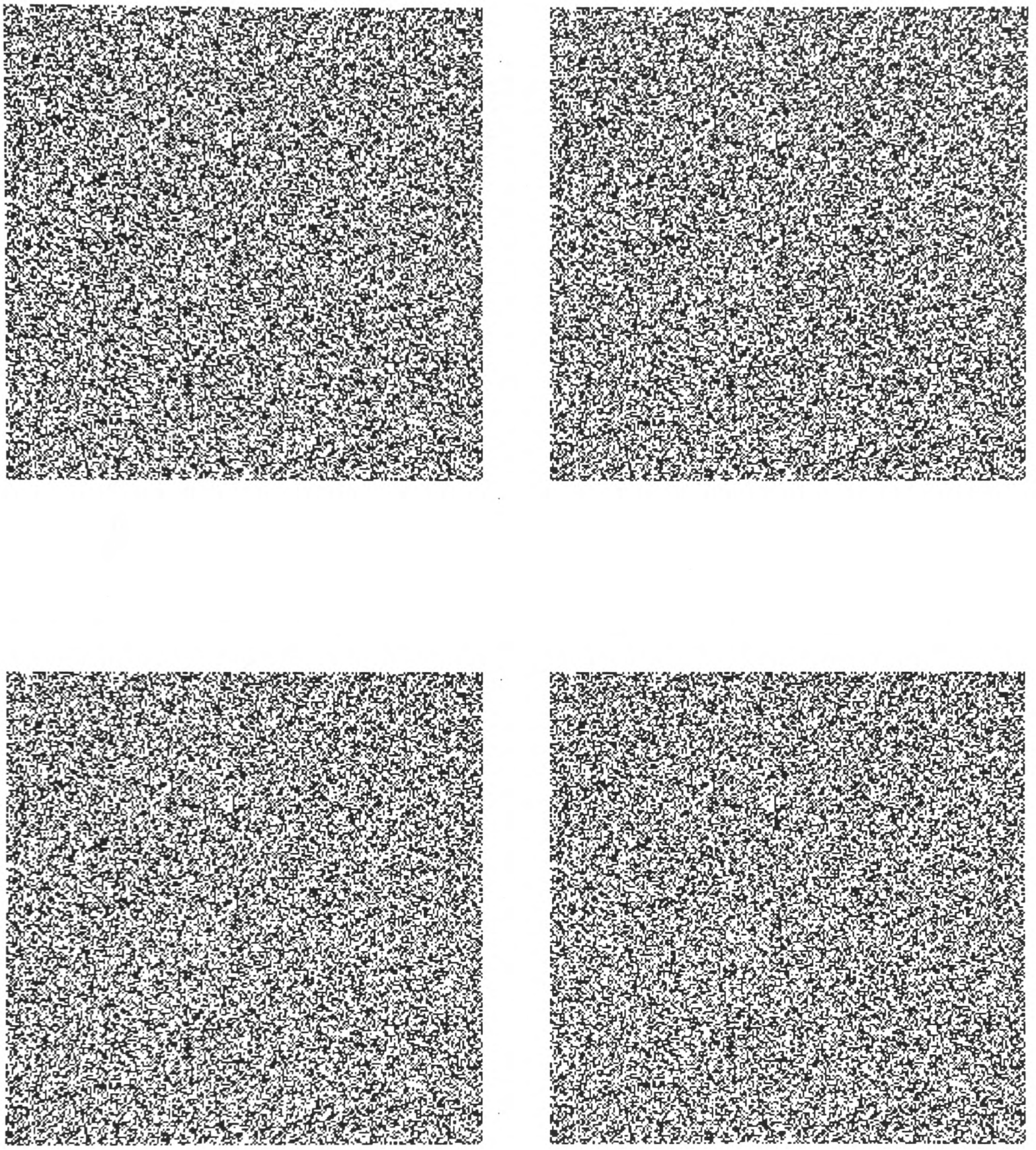


Fig 4.10 (a)

Above, left and right eye's images for a stimulus containing a uniform 2 pixel disparity target. Below, the target is defined by a disparity of +1 pixel for odd rows, -1 pixel for even rows, i.e. a "+1,-1" stimulus. These images were used for the models illustrated in (b) to (d).

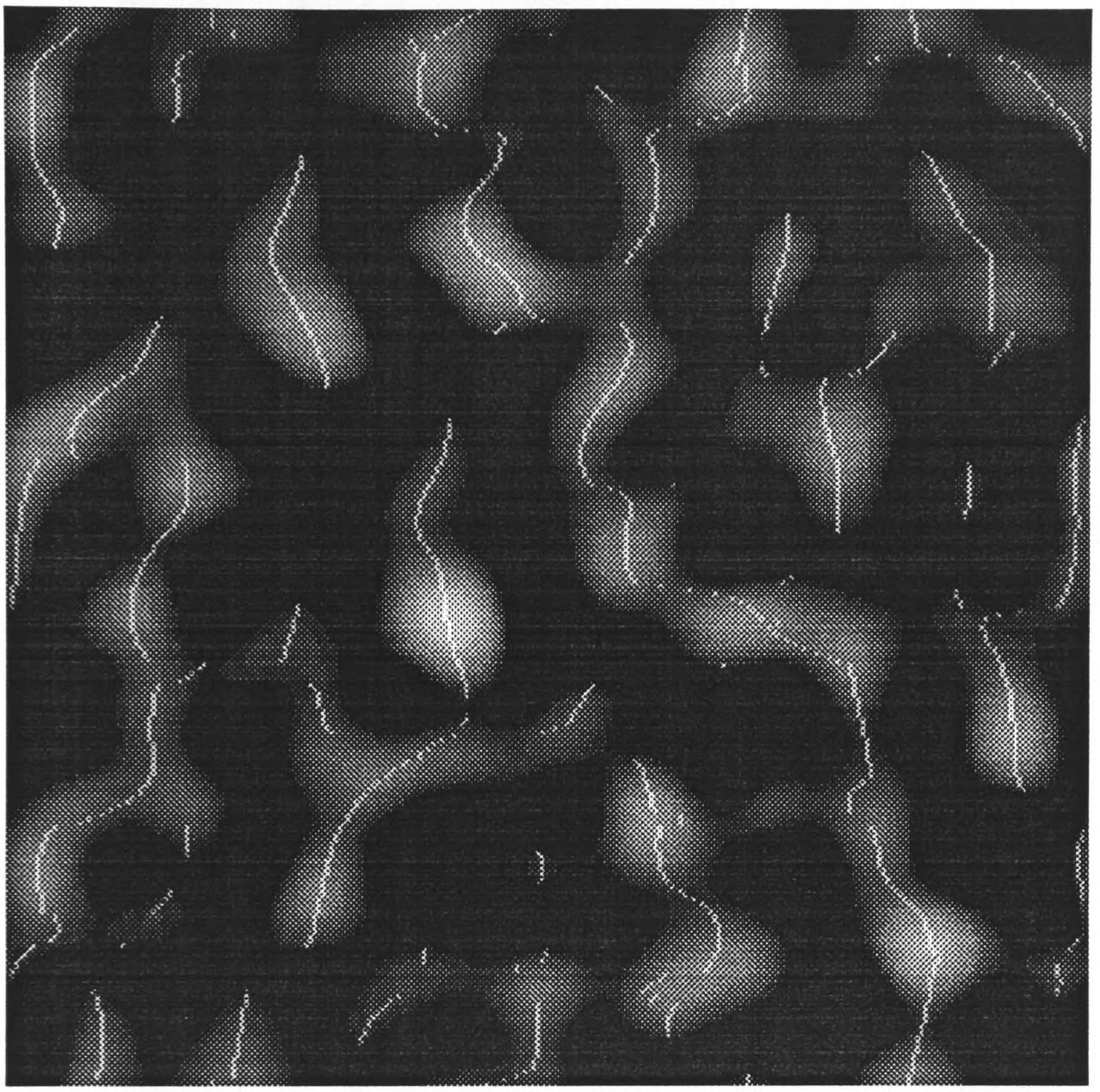


Fig 4.10 (b)

Centroids derived from a filtered image. One of the patterns shown in (a) has been filtered with a Laplacian of Gaussian filter (σ = 11.3 pixels) and half-wave rectified (the positive response is shown here). 1-D centroids were calculated (as described in the text) and have been superimposed on the filtered image. They mark the "backbone" of a blob

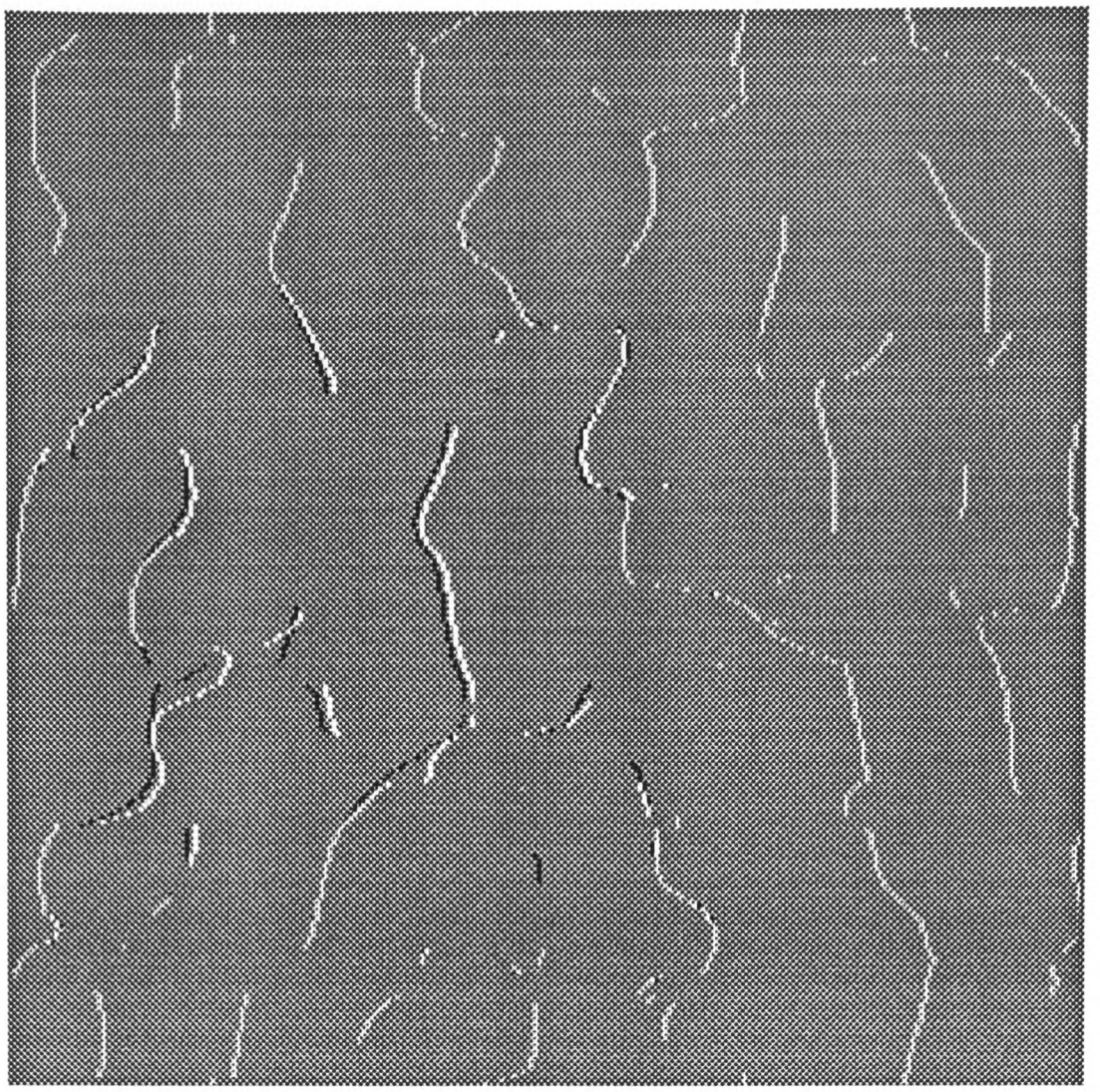


Fig 4.10 (c)

This shows the difference between the centroids derived from the left and right images of a uniform disparity stimulus (shown in (a)) after coarse scale filtering (σ = 11.3 pixels). Light grey pixels (above) indicate the presence of a centroid at that position in both the left and right images (i.e. zero disparity). White pixels indicate a centroid in the left eye's image but not in the right, black pixels indicate a centroid only in the right eye's image. A "snake" of white centroids next to a snake of black centroids signals a blob with a uniform disparity. Below, only those centroids with a non-zero disparity are shown.

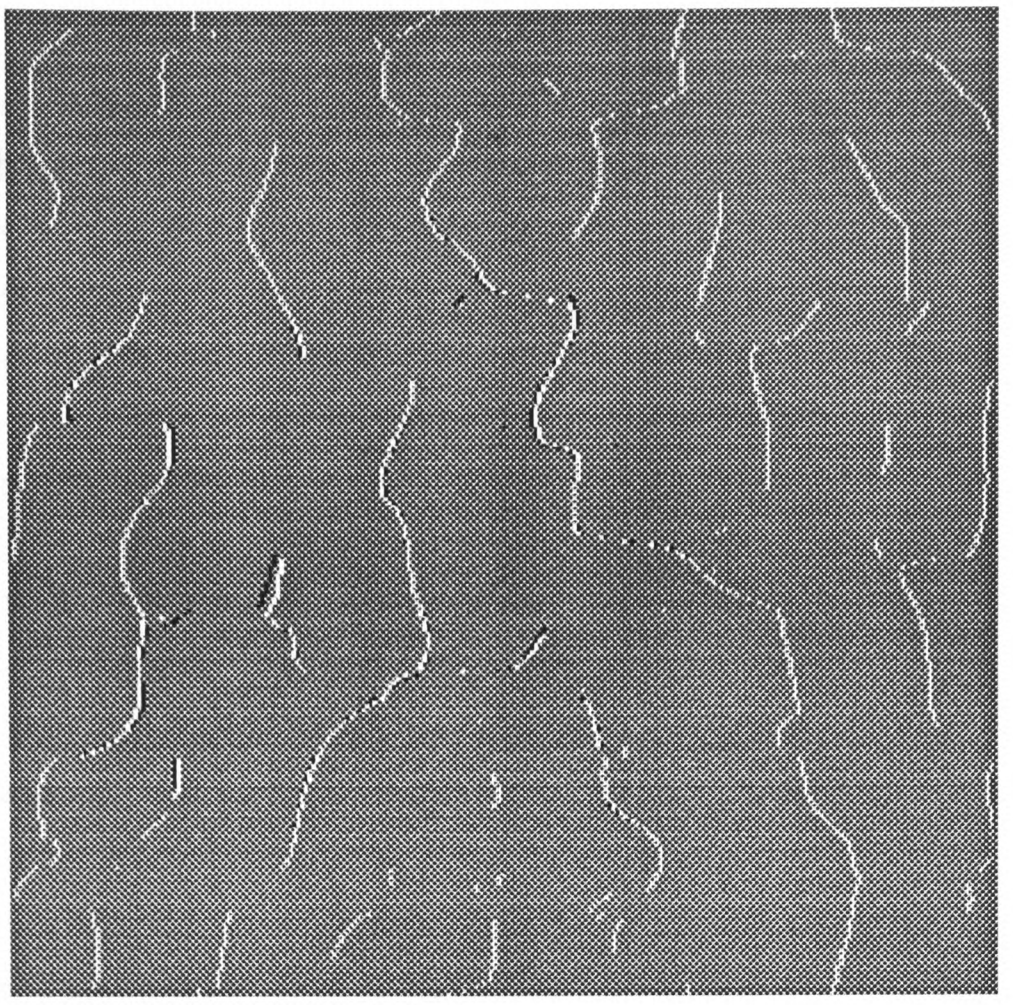


Fig 4.10 (d)

The "+1,-1" stereo pair is shown here in the same way as for the uniform disparity stimulus in (c). For this stimulus, at a coarse scale, the positions of centroids in the left and right eye's images are almost identical.

and right eyes' images are almost identical at a coarse scale. The bottom figure shows that there are differences in the centroid positions in the two eyes' images but they are very small and there is no consistent pattern to them. The long vertical snake that lies almost entirely within the target region provides a good example. In figure 4.10 (c) this "snake" is shifted to the right along its entire length. In figure 4.10 (d) a small number of pixels are shifted either to the left or right, but overall the centroid position in left and right eyes images is the same.

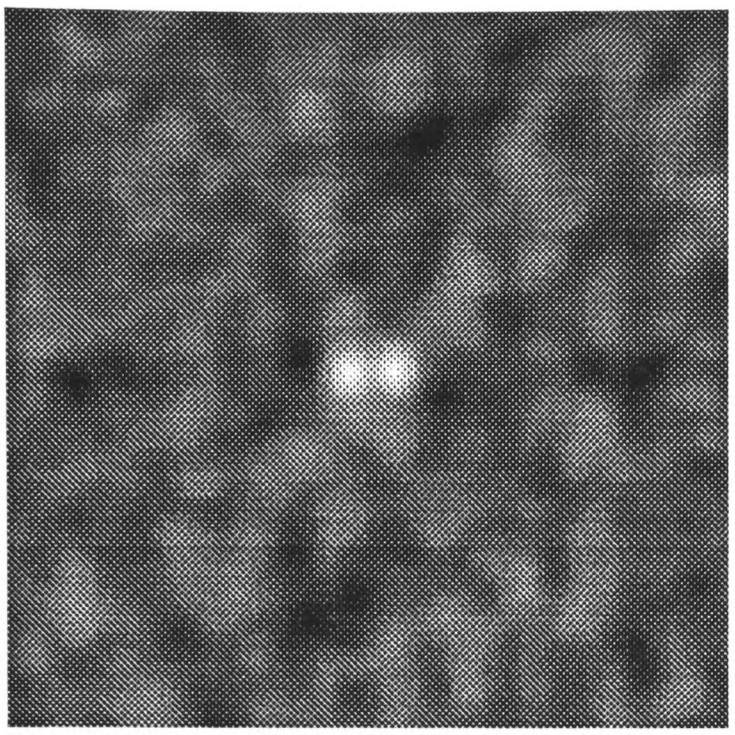
(In figure 4.10 (d) it can be seen that one or two small blobs are shifted wholly to the right or left. These isolated blobs should be detected relatively easily as having a disparity different from the background but they would give only very poor information about the target's shape. This corresponds well with subjects' perception of the "+1,-1" stimulus either at short exposures or when low-pass filtered at a coarse scale: the *presence* of the target was easy to detect, but its shape was very hard to determine. In this respect the task is very different from the tasks subjects were required to perform in the experiments of Parker and Yang (1989) and Stevenson et al. (1989).)

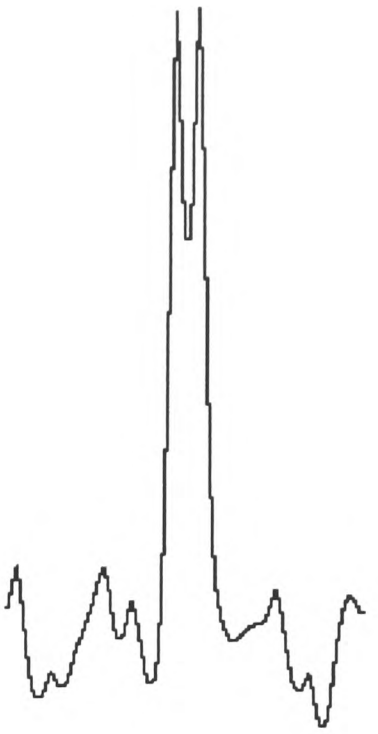
4.9.2 A cross-correlation model

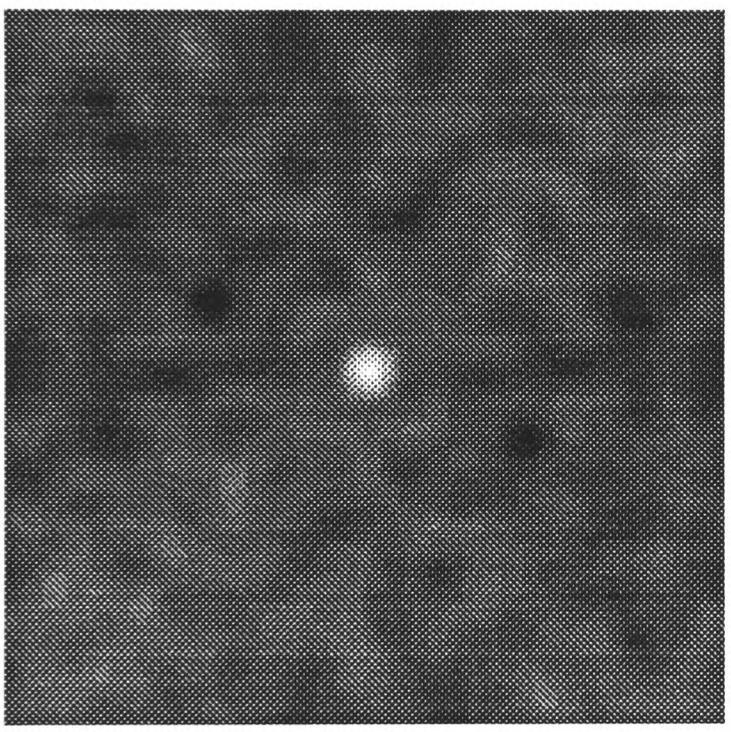
The other way to model these results is, as Parker and Yang did, to consider an area-based correlation of left and right images. Figure 4.11 (a) shows, at the top,

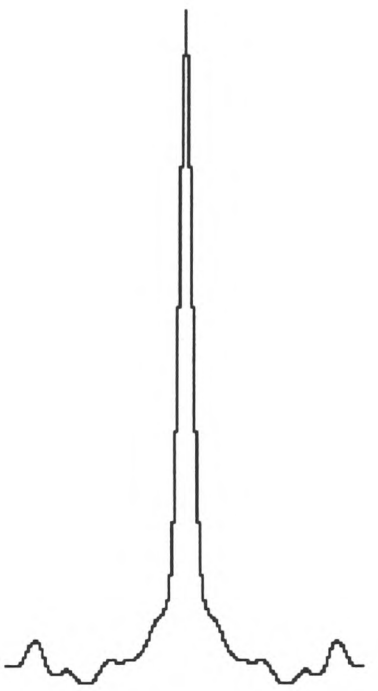
---

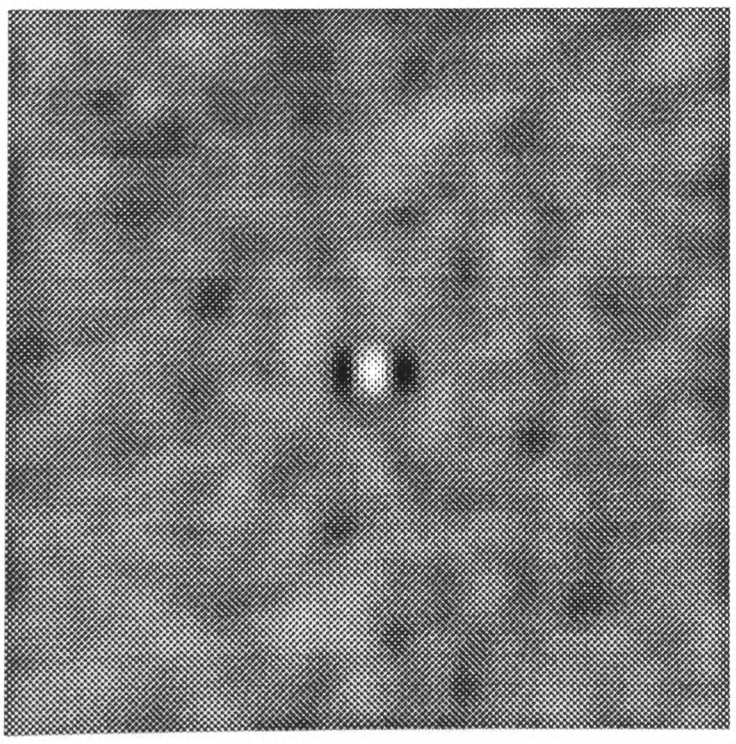
Fig 4.11 (overleaf)
Cross-correlations of "+1,-1" stimuli at different scales. At the top, on the left, is shown the cross-correlation of two images in which odd rows had a disparity of +1 pixel, even rows a disparity of -1 pixel (i.e. an area-based correlation entirely within the target region). The images were filtered with a Laplacian of Gaussian filter (σ = 0.7 pixels). On the right a cross-section through the centre (i.e. zero vertical disparity) is shown. The two peaks are at +1 pixel and -1 pixel, as expected. The peak correlation in this image is about 0.68 (where 1 corresponds to perfect correlation). In the centre, the auto-correlation of the background, filtered at the same scale is shown (the peak height is, by definition, 1). At the bottom, the difference between the above two correlations (i.e for target and background) is shown. The maximum difference between target and background is 0.51.
(Note that, to obtain sub-pixel resolution in this model, "pixels" were in fact made up from 4 by 4 "quarter-pixels").
(b) As for (a), but using a filter size of 1.4 pixels. The peak correlation for the target is 0.74. The maximum difference between target and background is 0.26.
(c) A filter size of 2.8 pixels. The maximum difference between target and background is 0.085.











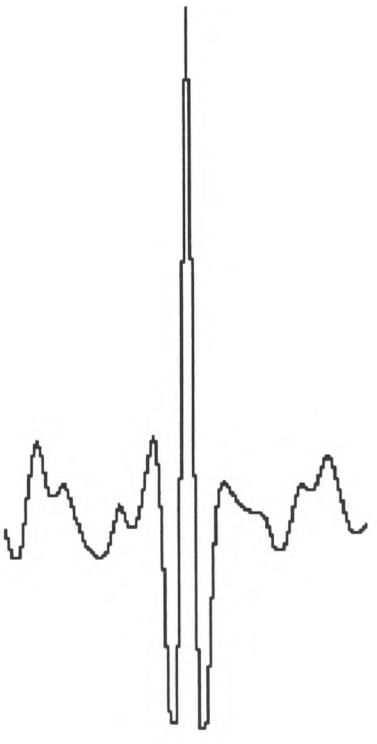
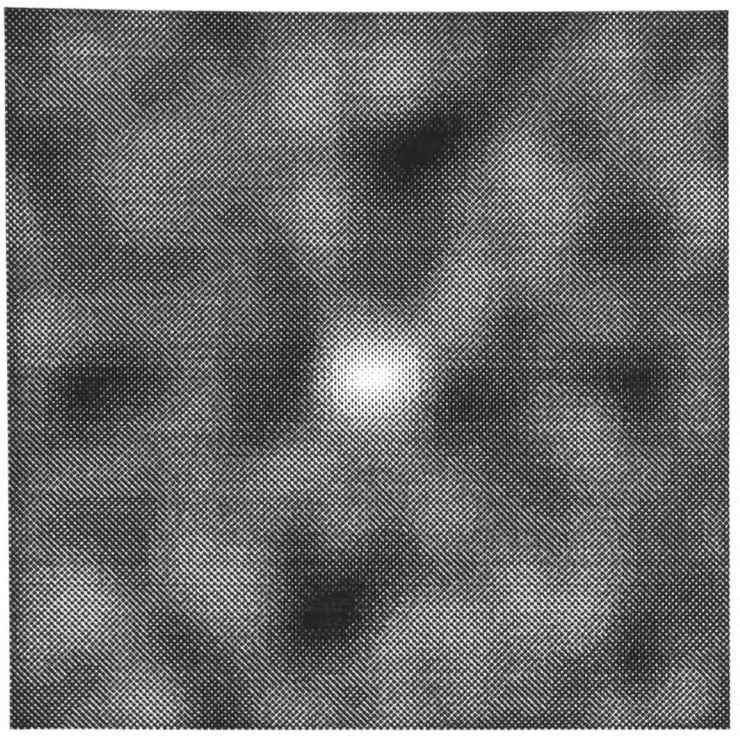
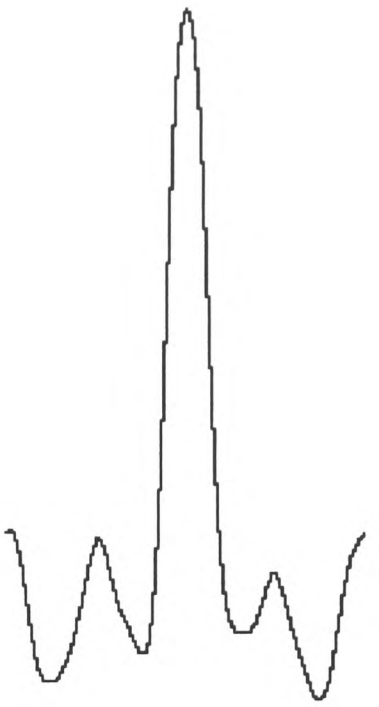
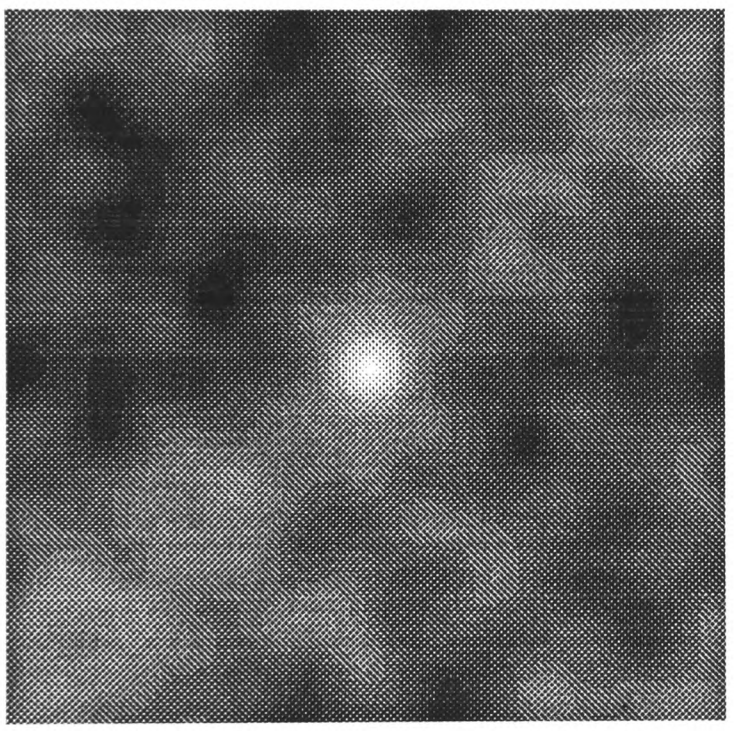
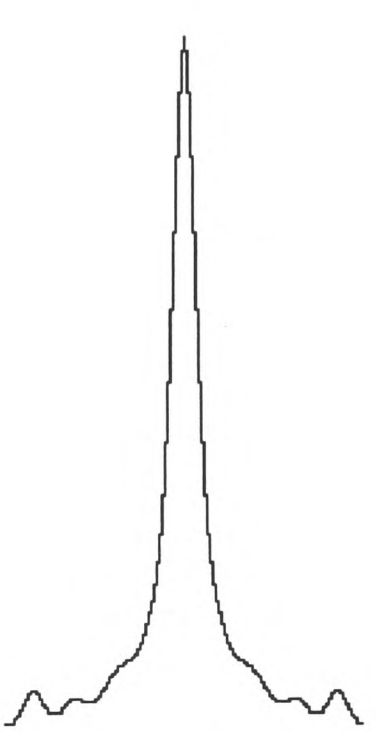


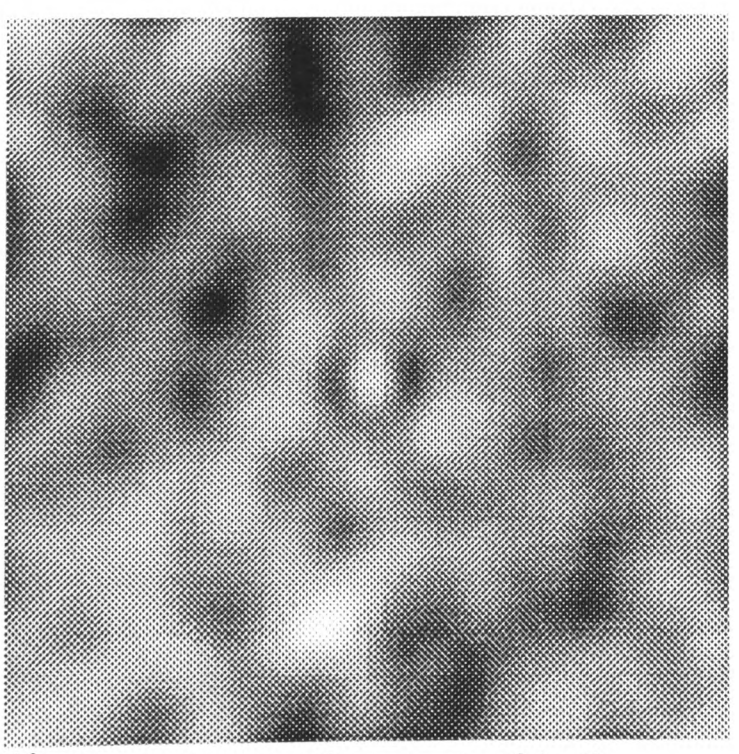
Fig 4.11a (legend on previous page)











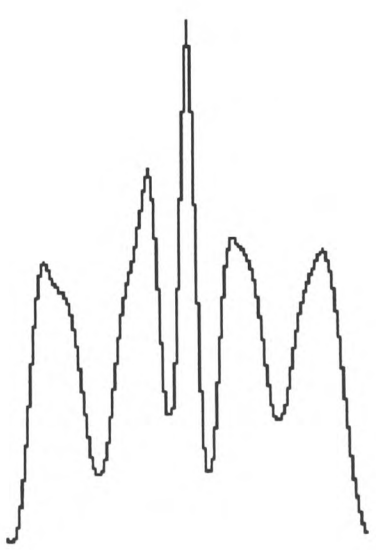
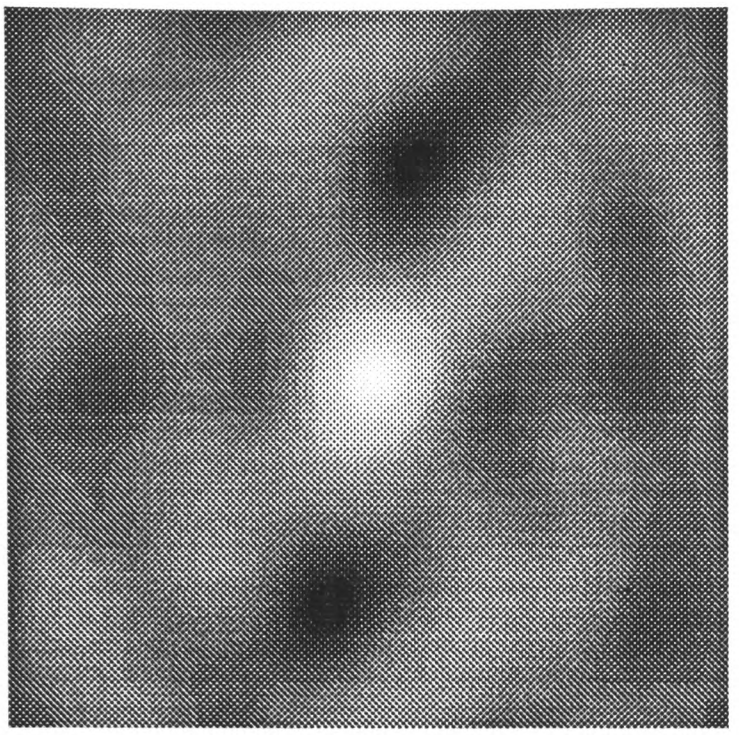
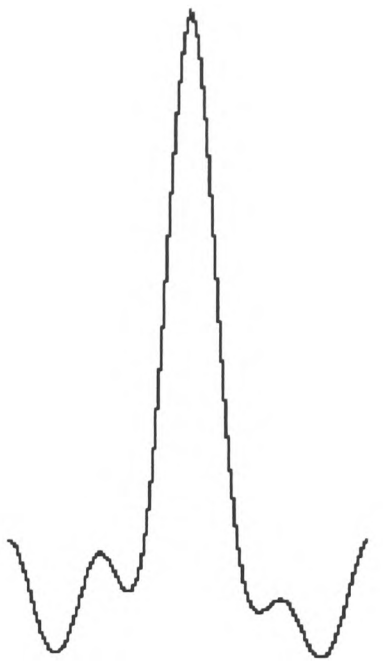
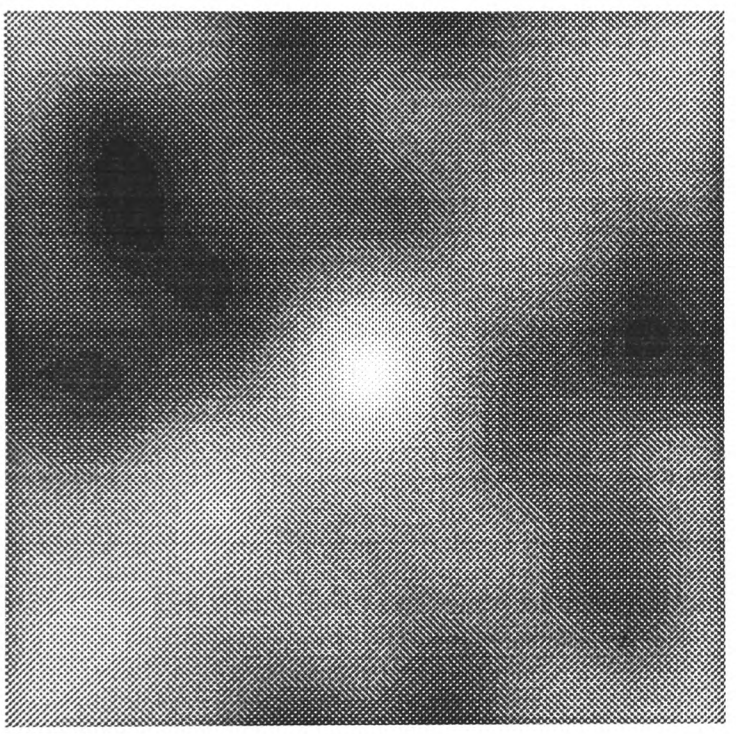
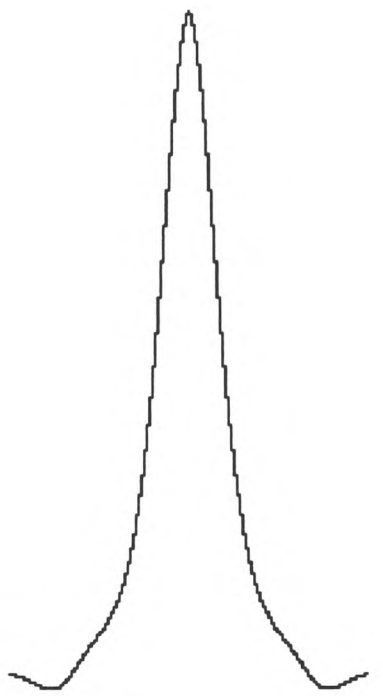


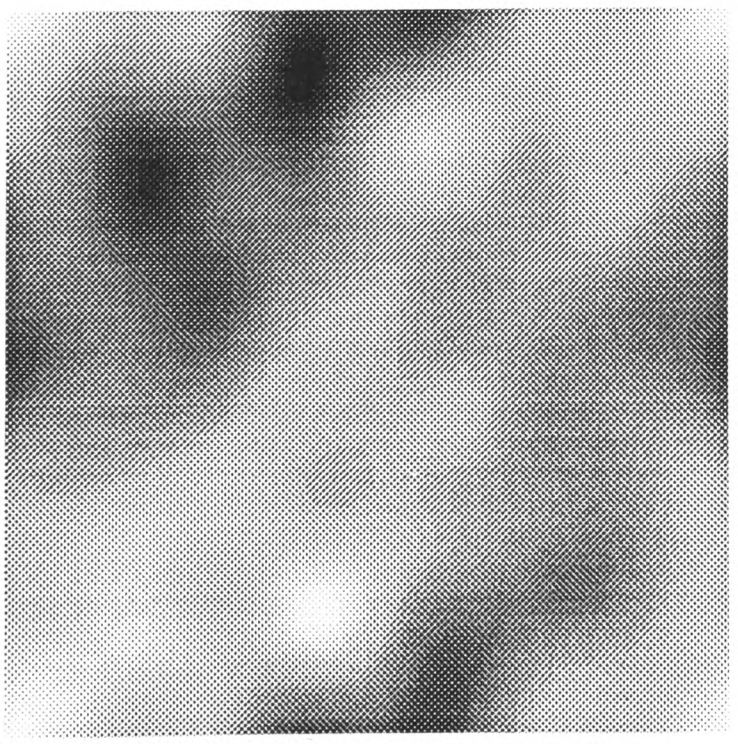
Fig 4.11b (legend on previous page of text)











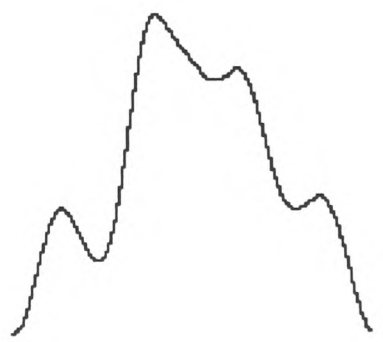


Fig 4.11c (legend on previous page of text)

the cross-correlation of left and right images (for an area within the target region) when the target contains disparities of +1 and -1 pixel. The images were filtered with a small filter (a Laplacian with a space constant of 0.7 pixels) and the two peaks, corresponding to the two disparity planes, can be seen in the centre. In Parker and Yang's experiment, the critical aspect of the modelling was to determine at what disparity the two peaks could be resolved. In the experiment described in this chapter, the subject's ability to do the task depends on the target being distinguishable from the background. The cross-correlation of an area of the background (i.e. its auto-correlation) is shown in the middle panel. There is one peak in the centre corresponding to the single, zero disparity plane. The bottom panel shows the difference between the correlation functions for the target and the background. There is a peak and two troughs in the centre, of much greater amplitude than surrounding "noise". It is one way of showing that at this scale target and background are potentially distinguishable.

Figure 4.11 (b) shows the same thing for images filtered at a larger scale (σ = 1.4 pixels). Target and background still differ significantly, as the bottom panel illustrates, even though the peaks in the cross-correlation of the target (top panel) are not resolved. For a larger filter (σ = 2.8 pixels, in figure 4.11 (c)), the correlation functions for target and background are essentially the same.

This type of modelling makes the same point as section 4.9.1 but in a different way: at large spatial scales the "+1,-1" target is very difficult to distinguish from the background. Both support the hypothesis that the "+1, -1" disparity target is difficult to see at very brief exposures because the scale of analysis is too large. As exposure duration increases the scale of analysis reduces, the individual lines become resolved, and the shape discrimination task becomes possible.

4.9.3 An experimental model

One striking aspect of the results of experiment I (figure 4.5) is the precipitous drop in thresholds over a very narrow range of exposure durations for the "+1,-1" stimulus (between about 80 and 100 ms). Does this correspond to a sudden change in the visual system at around this time after the onset of a stimulus, for instance a sudden change in filter size? There are two reasons for not coming to such a conclusion.

The first is the results from experiment II in which strip height was varied for the "+1,-1" stimulus. Varying strip height alters the disparity modulation frequency of

the stimulus. For wider strips the modulation is of a lower frequency and should be detectable to a coarser scale mechanism, either in the spatial or disparity domain. The fact that doubling the strip height moved the curve to the left (figure 4.6), i.e. the task was possible at shorter exposures fits well with the idea that coarser mechanisms are acting at shorter exposure durations. The fact that the curve was again very steep, but now the sharp change occurred at a slightly *different* exposure duration, suggests that it is not only a change in the visual system that is responsible. Rather it suggests that some *interaction* between the spatial characteristics of the stimulus and a change in the visual system causes the rapid change in performance.

The second and much more direct reason is the results of experiment IV in which the spatial frequency content of the stimulus was varied. The rationale was to model the effects of short exposure durations by artificially removing high frequency information from the stimulus. As figure 4.12 shows, the model is a good one. In particular, the very sudden change in thresholds produced by a very small change in the size of Gaussian filter used to blur the stimulus mirrors the sudden change in thresholds over a narrow range of exposure durations.

The implication of the results of experiment IV is that a sudden change in performance for the "+1,-1" stimulus at one particular exposure duration is compatible with a model in which filter size changes smoothly with exposure duration after the onset of a stimulus. Even more can be gained from these results, as figure 4.12 illustrates. In figure 4.12a the results for one subject from

---

Fig 4.12 (overleaf)
This figure illustrates how the results of experiment I and experiment IV can be combined to give, very approximately, the rate of change of filter size over time. In figure 4.12a (top), results from experiment IV are shown (re-plotted from figure 4.8). Below, results for the same subject for experiment I are shown (re-plotted from figure 4.5). In figure 4.12b (top), the data from these two graphs have been superimposed by transforming the s.d. of Gaussian into a "model" exposure duration (see text). Figure 4.12b (bottom left) shows graphically the relationship between filter size and exposure duration that is derived from superimposing the two data sets. (No data is shown in this plot because it is the abscissae not the data that are being compared). The result is similar to the relationship between filter size and exposure duration derived by Watt (1987) for an experiment on line length discrimination, ((d),below right).

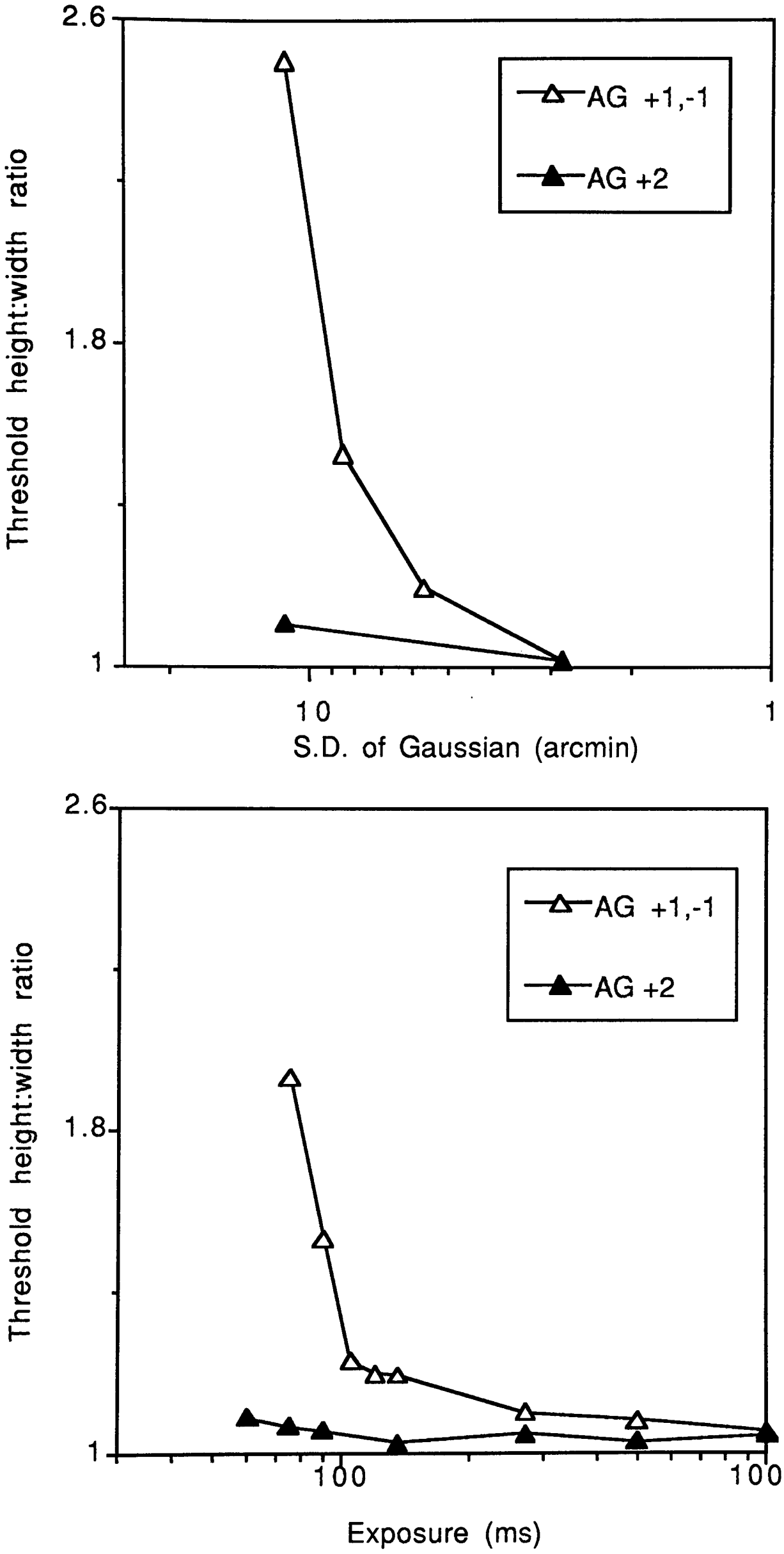
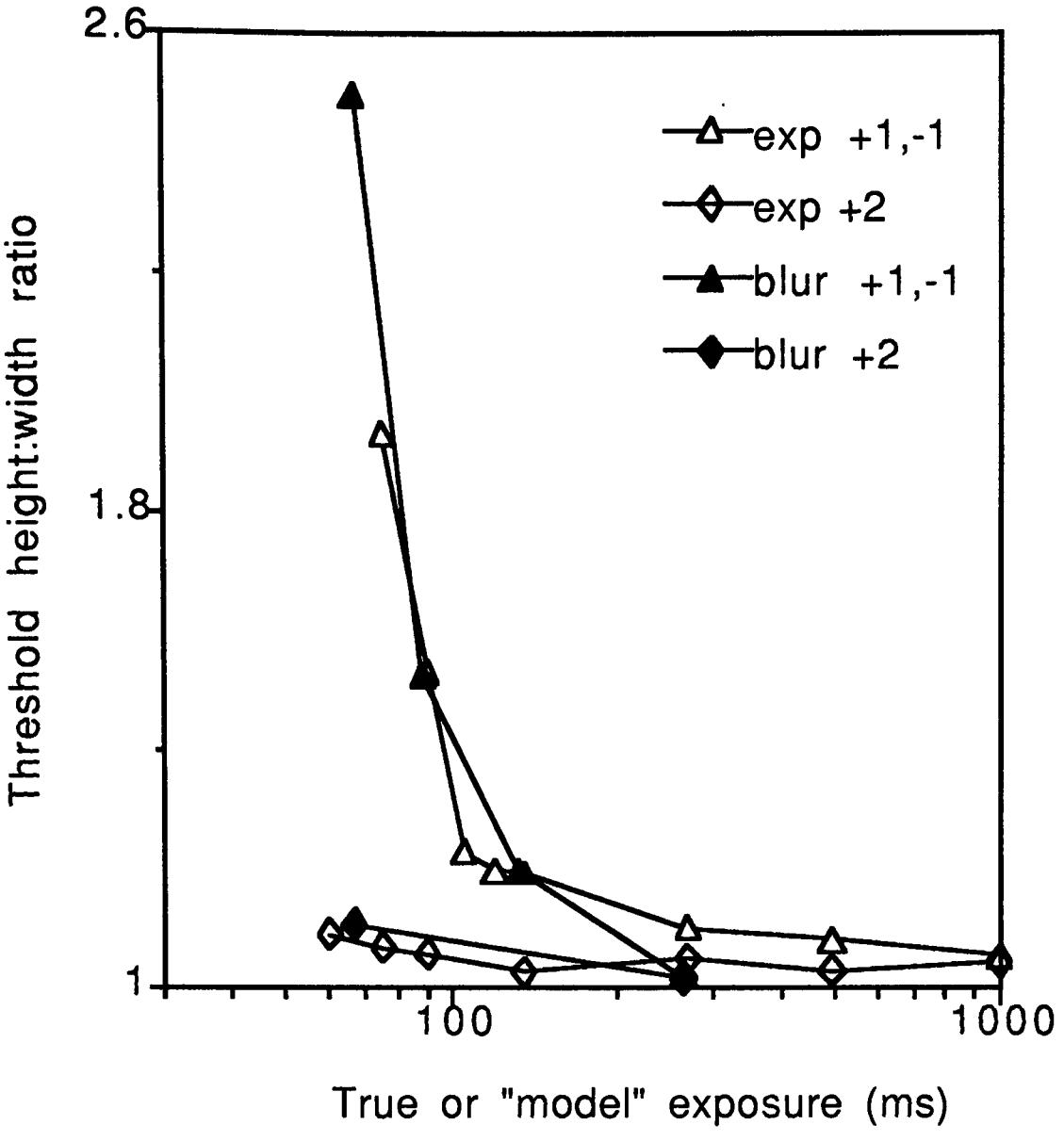
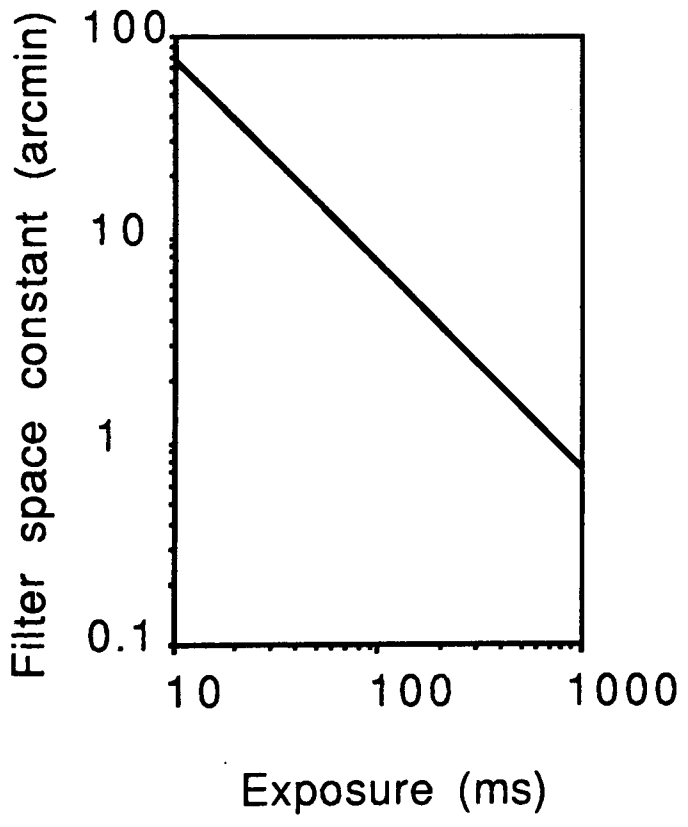


Fig 4.12a (legend on previous page)





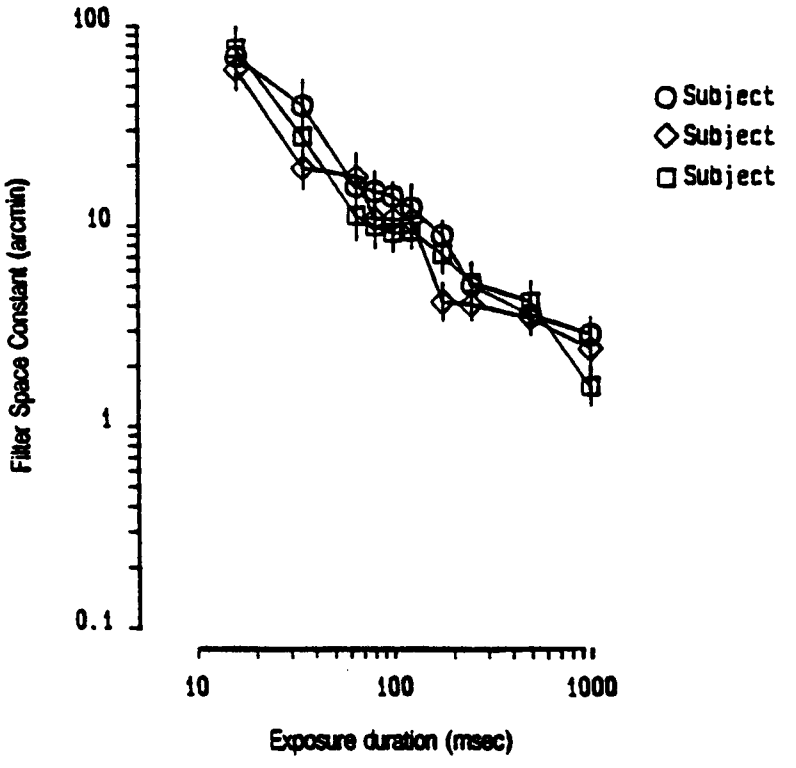


Fig 4.12b (legend on previous page of text)

experiments I and IV are shown together (data re-plotted from figures 4.5 and 4.8a). Threshold height:width ratio is plotted in the ordinate in both cases. The abscissa in one case shows filter size and in the other exposure duration. All the key aspects of the exposure duration data are repeated in the low-pass data. The uniform disparity target is seen easily, and performance is only slightly affected by either exposure duration or filter size. Thresholds for the "+1,-1" target, on the other hand, are too high to measure at exposure durations briefer than about 80 ms or for a filter size above about 10 arcmin. Then, over a narrow range of exposure durations, or a narrow range of blur, thresholds change very rapidly. At longer exposures, or for finer scale blur, performance for the "+1,-1" stimulus is only slightly worse than for a for a uniform disparity target.

Figure 4.12b (top) shows the data from the two plots in figure 4.12a superimposed. To do this one x-axis has to be transformed. The transformation shown here is:

$$\text{"model" exposure (ms)} = \frac{750}{\text{s.d.of Gaussian (arcmin)}} \qquad \text{(i)}$$

This equation can be used to plot model filter size against exposure duration (bottom left). There is considerable variability in the transformations which can be chosen, especially in the "stretch" applied to one axis (and hence the slope of the plot of filter size against exposure). It is possible to determine more precisely the exposure duration and filter size which correspond to the sharp change in thresholds for the "+1,-1" target (e.g. 75 ms and 10 arcmin).

The plot of filter size against exposure duration derived from experiment I and IV can be compared with that derived by Watt from an experiment on hyperacuity judgements at a range of exposure durations (discussed in section 2.4). The two plots are similar. In particular, note that the estimate from Watt's experiment of filter size at 75 ms is in the region of 10 arcmin.

The modelling in this chapter, including the experiment using filtered stimuli is not very sophisticated and has only been applied to two target types (+1,-1 and uniform disparity of 2 pixels). It seems likely that coarse-to-fine hypothesis would be compatible with the results of experiments varying strip height (experiment II) and perhaps also the results of experiment III using small disparity and mixed disparity targets. A better approach would be to develop an ideal observer model which could be used to test the effect of strip height, pedestal disparity, noise (in the spatial or disparity domain) and filter size or different filter combinations.

There are several problems in developing such a model, but it remains an important goal.

4.10 Discussion

In the experiments described here, one stimulus stands out as different from the others, i.e. the "+1,-1" mixed disparity stimulus. As the modelling in the previous section makes clear, in the context of a filtering model this is a special case because at a coarse scale left and right eye's images are almost identical: only fine scale information will distinguish the two images and reveal the target. Hence, the exposure duration data from experiment I have been interpreted as strong evidence for a coarse-to-fine process in human stereopsis. Is this conclusion justified?

4.10.1 Can a local-to-global theory explain the results?

It is important to consider in some detail other possible explanations. After all, most currently popular stereo algorithms with which human performance is compared (e.g. Pollard et al. ("PMF"), 1985; Prazdny, 1985) are fine-to-coarse ("local-to-global") algorithms. In fact, most stereo algorithms are variations on this theme (Julesz, 1971; Marr and Poggio, 1976; Lehky and Sejnowski, 1990; Mitchison, 1988). How would these models account for the time course data presented in this chapter?

There are two main stages in any local-to-global process (discussed in greater detail in section 1.3). The first is to find the disparity of any given feature. In situations where there are many alternative matches some co-operative or iterative process is usually required to determine the correct match for each feature. (The position of each feature in the left and right eye's image is assumed to be known so that, once the correct match is made, calculating the disparity is not a problem.) The second stage is to interpret the disparity of a feature to give its depth. This depends, in part, on the disparity of surrounding features and includes disparity averaging, depth contrast, scaling according to viewing distance etc.

Which of these two steps can account for the time course observed for the "+1,-1" disparity stimulus when compared to other stimuli? It is not likely to be the first stage: the correspondence problem is no more difficult for this stimulus than for others used in the experiment, e.g. the "+3,+1" stimulus. (For most algorithms, co-operativity only applies in the horizontal direction, along epipolar lines, in

which case the correspondence problem is equally simple for all the stimuli. If there is, on the other hand, some vertical co-operativity then the "+3,+1" and "+1,-1" should be equally disrupted and the uniform 2-pixel disparity stimulus much easier to see. This was not the experimental result.)

So, according to a local-to-global model, when the first stage is complete, the disparity of each pixel (or fine scale feature) is known with a given degree of accuracy. Given that thresholds for the uniform disparity and "+3,+1" stimulus are quite low even after an exposure of 60 ms, the reason that equally good performance is not achieved for the "+1,-1" target must be explained.

It is not sufficient to say that in the second stage disparity averaging "obliterates" the target by interpolation, i.e. by drawing an average plane through the target at the same depth as the background making it invisible. This would beg the question of what happens at longer exposure durations to make it visible again?

An obvious hypothesis is that the degree of accuracy with which the disparity of a pixel is known might improve with longer exposure durations. If the populations of disparities within the target region and the background could be compared then the mean disparity of the populations would distinguish the uniform disparity and the "+3,+1" disparity targets from the background (unless the noise was very large) but not for the "+1,-1" stimulus (its mean disparity within the target is the same as the background). As exposure duration increased and the standard deviation of disparity estimates reduced the "+1,-1" target could be distinguished from the background in one of two ways. Either the width of the distributions for target and background could be compared and, at some level of noise, the target accurately distinguished as having a wider distribution than the background; or, at a lower level of noise, the two disparity populations within the target region (+1 and -1 pixel) could be resolved and so distinguish the target.

Note that some degree of spatial averaging is implicit in this proposal. If the analysis could be restricted to one raster line then the "+1,-1" target should be as easy to see at short exposures as a "+1" target, i.e. a uniform disparity target with a disparity of 1 pixel. The results of experiment III show this not to be the case. This example also emphasises that it is not simply the small disparities used in the "+1,-1" target that make it difficult to see at short exposures.

Other results of experiment III are those for a "+3,-3" stimulus at brief exposures, i.e. one in which odd rows were defined by a disparity of +3 pixels, even rows by a disparity of -3 pixels. If noise in the disparity domain explains the effect of exposure duration then this stimulus should be easier to see at brief exposures than a "+1,-1" stimulus, a prediction confirmed by the results shown in figures 4.7. Unfortunately, a coarse-to-fine model predicts this result as well. Left and right eyes' images are much less similar at a coarse scale for a "+3,-3" stimulus than they are for a "+1,-1" stimulus and so the target can be distinguished because it is less well correlated than the background. Figure 4.7 also shows results for an uncorrelated target. Thresholds for this stimulus are much lower than for the "+1,-1" stimulus and are comparable with the "+3,-3" stimulus.

These variations on the main stimuli used in experiment I help to illustrate that, provided some degree of spatial averaging is incorporated, a model in which noise in the disparity domain reduces as exposure duration increases can successfully explain some aspects of the data from experiment I. To summarise, the model would rely on some mechanism whose receptive field lies within the target area and that measures the distribution of disparities in that area without regard for their spatial location. The output, i.e. the distribution of disparities, would be compared to that for another area, the background. As exposure duration increases the amount of noise in each distribution would, according to this model, reduce enabling the target and background to be distinguished.

There are two important caveats. First, in discussing the distribution of disparities in the target region and the background it may not be valid to describe a noise process as simply increasing the variance of the distribution since each pixel does not necessarily provide an independent sample: a local-to-global algorithm such as "PMF" (Pollard et al., 1985) depends on the high correlation between the disparity of adjacent pixels in order to solve the correspondence problem. Any local-to-global model would have to specify what degree of noise in the disparity domain was compatible with a solution of the correspondence process.

The second caveat concerns the data from experiment II, in which the height of strips was altered. In considering only the distribution of disparities within the target region or the background the *spatial location* of pixels within the receptive field is not taken into account. It cannot predict, therefore, the effect of varying strip height unless the model is modified.

4.10.2 Can a *modified* local-to-global theory explain the results?

One solution is to propose a range of mechanisms with different receptive field sizes covering parts of the target. These would pick up "skewed" populations of disparities within the target if their receptive field included more strips with +1 pixel than with -1 pixel disparity, and hence distinguish target from background. The smaller the receptive field the more likely this would be to happen. The results of experiment II could be explained if the receptive fields were large at short exposures and reduced in size as exposure duration increased. (Note that, in the version of the model considered so far, if any change of receptive field with exposure duration were to be proposed it would be the opposite one. That is, the *larger* the receptive field of the mechanism "looking" at the target the more accurate its estimate of the distribution of disparities and hence the better the performance (provided its receptive field was entirely within the target)). Making this change adds a coarse-to-fine element to the model, although it remains nominally in the disparity domain.

The remaining aspect of the results that require explanation in any model is the sharp change in thresholds over a narrow range of exposure durations. In fact, this fits quite well with the model in which the size of the receptive field of the supposed "mechanisms" shrinks over time. In this model, the presence of the target is signalled by a bias or skewed distribution of disparities picked up by some of the mechanisms. This will happen when the receptive field covers more "+1" strips than "-1" strips or vice versa. The maximum difference in the number of "+1"or "-1"strips within a receptive field is one (because the strips alternate) and this will cause a minimal bias to the population when the receptive field is large. For example, as the receptive field size reduces from approximately 32 lines to 16, the ratio of number of pixels with "+1" disparity to those with "-1" disparity changes (at most) from 15:16 to 7:8. (These ratios are, in fact, for receptive fields of 31 and 15 pixels. Receptive fields covering 32 and 16 lines (or any even number of strips) would never receive a skewed distributions of disparities). Neither of these ratios are particularly "skewed", i.e. different from 1. However, as the receptive field size is reduced from about 8 to about 4 lines the ratio of number of pixels within the receptive field area whose disparity is "+1" compared to those with a disparity of "-1" changes (at most) from 3:4 to 1:2, in other words, a much swifter change. Reducing receptive field size still further gives a maximally skewed ratio of 0:1 (the receptive field lies entirely within one strip). In other words, there is a sudden switch from a roughly balanced signal to a completely skewed one and hence a sudden improvement in performance would be expected.

The receptive field size at which this switch occurred would be proportional to the strip height.

So, modifying a local-to-global theory in this way yields a satisfactory explanation of the data in experiment I and II. But can the model any longer be considered to be a local-to-global one?

The modifications suggested in this section are in fact quite profound. First, an assumption must be made that although the disparity of every pixel is calculated, there is no access to the disparity of individual pixels (or, for example, the disparity of pixels along one raster line) but instead the disparities of the pixels within the receptive field are considered as a "population" whose mean and perhaps standard deviation are known. Second, it is assumed that the summation area over which this disparity "pooling" takes place is initially large and reduces progressively over time. Although such a model is nominally in the disparity domain, it is not easily distinguished from a coarse-to-fine model in the spatial domain, as proposed in section 4.9.

4.10.3 Coarse-to-fine or coarse-then-fine?

The results of the experiment described in this chapter are compatible with a model in which the scale of analysis varies with exposure duration, i.e. a coarse-then-fine model. They provide no direct evidence for a coarse-to-fine model, such as that proposed by Marr and Poggio (1979). To do so would require evidence that, under some circumstances at least, the processing of fine scale information depends on the outcome of coarse scale analysis. The issue is discussed in detail in chapter 5, in particular with reference to the work of Mitchison and McKee (1985, 1987a and b).

A coarse-then-fine model would be one in which the matching process was carried out independently at several scales simultaneously. The density of spatial primitives is much lower at a coarse than a fine scale so the correspondence problem might be expected to be solved more quickly at a coarse scale. This might be the reason that coarse scale stereoscopic information is apparently available at an earlier stage than for fine scales. The fine scale process, being independent of coarse scale processes, would be faced with a difficult correspondence problem. It would need to be a co-operative process, like "PMF" (Pollard et al., 1985). Presumably, the larger the disparity of the target the more difficult the correspondence problem would be and hence the longer it would take to reach a

solution. The data from the experiment described in this chapter cannot be used to test this hypothesis. A coarse-then-fine theory lacks computational elegance, but is still a possible model of the data presented in this chapter.

4.10.4 Coarse scale "grouping" and the perception of (2-D) shape

One feature of the "+1,-1" stimulus that has not been discussed is its appearance at long exposure durations (1 second or greater). Despite the fact that for an exposure duration of 1 second performance on the shape discrimination was just as good as for the uniform disparity stimulus (see figure 4.5), the subjective quality of the perception is different. Subjects report that the stimulus appears "less solid" and is "hard to pin down". Why might this be?

A possible explanation, that fits with the description of a hierarchical representation of information about the image described in chapter 3, is that the shape of the target rectangle is recorded at just one scale, the scale at which it is first detected. For a uniform disparity target this would be quite a coarse scale. For a "+3,+1" target this would be an equally coarse scale with, at a finer scale, the modulations between odd and even lines recorded. In other words, the fine scale information would provide an adjustment to the coarse scale signal, it would "paint in the detail" on the coarse scale object. For a "+1,-1" target there is *only* fine scale detail and no coarse scale object on which to "paint it". Maybe this is the reason that it appears "unsatisfactory" or "less solid".

To put it another way, perhaps the visual system is bad at "building up" an impression of large scale objects from fine scale information (even though this is the fundamental principle of a local-to-global theory). The perception of the "+1,-1" stimulus at long exposure durations gives a clue as to how fine scale information about the target may be stored, a theme taken up in greater detail in the next chapter.

4.10.5 Noise or filter size?

The issue of noise was considered in the context of a model in the disparity domain (section 4.10.1) but has not been considered with respect to a coarse-to-fine model. This is an important omission. It is very likely that at short exposures the level of noise will be greater than for long exposures because, if there is summation over time, noise (assuming it is temporally uncorrelated) will sum to zero while the signal will not. Any model of the effect of exposure duration should incorporate some analysis of the change in noise levels over time.

The most important question is whether all the results could be accounted for in terms of a reduction in noise levels over time rather than a change in filter size over time. The reply is very similar to that given in section 4.10.2. If filter size does not change over time then presumably the filter is the finest one. In this case the reason the "+1,-1" stimulus gives such different results from any other is difficult to explain unless some sort of spatial averaging is incorporated in the model. The most difficult aspects of the results to explain, as discussed in section 4.10.2, would be the effects of strip height and the reason for the sharp change in thresholds over a narrow range of exposures.

A more serious challenge is the possibility that *both* filter size *and* noise levels change with exposure duration. This is very likely to be the case. The consequence would be that a smaller range of filter sizes might successfully model the change in thresholds over time. The same argument applies to the experiment carried out by Watt (1987) with which the "derived time course" (figure 4.12) was compared. In that experiment too (discussed in section 2.4), the effects of short exposures were modelled entirely on the basis of change in filter size. If both noise and filter size were assumed to change with exposure duration then the range of filter sizes needed to explain the data might be considerably smaller.

4.11 Summary

The aim of the experiment described in this chapter was to pit two very different theories (or classes of theory) about the stereoscopic matching process against one another. For a "local-to-global" theory, of which there are many variations, fine scale information provides the "building blocks" or starting point for the algorithm. (The correspondence process is assumed to use fine scale information as its input, and, once the correspondence problem is solved, the output is a fine scale disparity "map" (after which stage further processing may or may not occur)). By contrast, in a coarse-to-fine theory, fine scale information adds detail to an approximate solution derived at a coarse scale.

The key stimulus in this experiment was one in which the target was defined *only* by fine scale disparities. In theory, such a stimulus should present no difficulties to the correspondence stage of a "local-to-global" matching algorithm. On the other hand, an algorithm that initially processed the stimulus at a coarse scale would find

this type of target particularly difficult to detect (until the scale of analysis was reduced). The results of the experiment described in this chapter confirm, in line with the predictions of a coarse-to-fine theory, that this type of stimulus (defined only by fine scale disparities) is much more difficult to see at short exposure durations (experiment I) than other types of stimuli. The results of experiment II, which showed that subjects were able to do the task at shorter exposure durations when the disparity modulations within the target were of a lower spatial frequency provided extra support for a coarse-to-fine model. The results of experiment IV showed that blurring the stimulus can mimic the effects of short exposure durations. The results do not fit with the simplest local-to-global algorithms, such as "PMF". They do not rule out a local-to-global matching process altogether, but a coarse-to-fine model appears to provide the most parsimonious explanation .

CHAPTER 5

In the previous chapter, evidence was presented suggesting that stereoscopic images might be analysed first at a coarse scale and then, over the first second of viewing, at progressively finer scales. But the experiment did not address the *purpose* of such a scheme. Marr and Poggio put forward a coarse-to-fine algorithm because it helped solve the correspondence problem. This is certainly one advantage (discussed in section 1.4 and in more detail in the next chapter). Watt's hierarchical scheme has another purpose: it orders or groups information about the image. It is this issue which is examined in the present chapter.

5.1 Hierarchical encoding and stereopsis

According to Watt (1988, discussed in chapter 2), the outcome of the coarse-to-fine analysis, once completed, is a "database" of all the lengths, orientations and

separations of the blobs in the image at every scale. It is a hierarchical database where fine scale information about a blob relates it to other nearby blobs and only via its coarse scale "parents" to other parts of the image. Judgements of length, orientation and separation are then made on the basis of information in this database. In chapter 3 it was proposed that this sort of analysis is carried out independently for each eye's image before the two images are compared for stereopsis.

How is it possible to verify this theory experimentally? The approach taken in this chapter is to investigate a well known illusion in which lines of the same length appear to differ in length (the Müller-Lyer illusion). This illusion may reflect distortions which arise as a result of a hierarchical method of encoding position (a hypothesis discussed in the next section). If this is the case, and the two eyes' images are compared after each has been encoded as a hierarchy of lengths, then several experimentally testable predictions will follow. The one explored in this chapter is the possibility that "illusory" length differences resulting from the Müller-Lyer illusion might yield a perception of slant when the stimuli are viewed binocularly.

The next section examines the reasons why a hierarchical system of encoding position might lead to apparent distortions of length.

5.1.1 The Müller-Lyer illusion

In a hierarchical database of image measurements, some measurements are recorded explicitly and others are not. The way in which the image of a face might be encoded serves as a good example. The information required for most (geometric) judgements is likely to be recorded explicitly at one scale or another. The outline of the face and its relationship to other features in the image would be recorded at a coarse scale. The distance between the eyes and the orientation of the nose would be recorded at a finer scale and at an even finer scale, the separation of the eye lashes. Each of these measurements would be calculated, and recorded, at a single scale, namely the scale at which the blobs (for example, the eyes) were initially resolved.

How might these measurements be used? In theory, the information stored in the hierarchy could be used to reconstruct the precise 2-dimensional Cartesian co-ordinate of every point in the image (since no information has been lost). So, for example, the relative position of two points could be calculated precisely, no matter

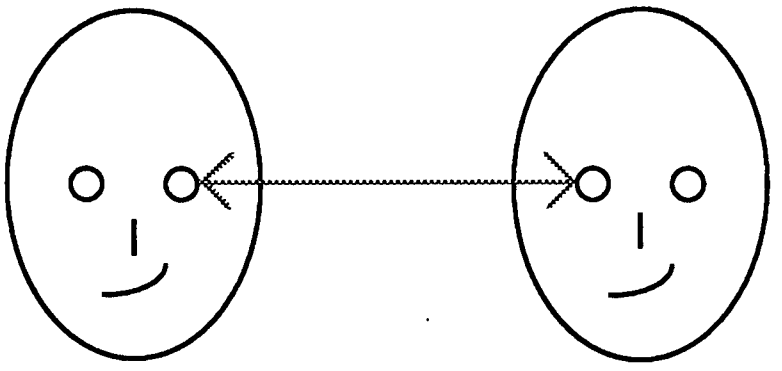
how the two points were grouped in the hierarchy although, if the two points (fine scale blobs) were nested within different coarse scale groups, the calculation of their relative position would require information recorded at several "levels" (or scales) in the hierarchy. The separation of these two points is *implicit*: it can be calculated from a series of explicit measurements.

The hypothesis explored in this chapter is that the visual system may not be very accurate at comparing implicit measurements, that is, making comparisons which require information to be combined from several levels in the hierarchy.

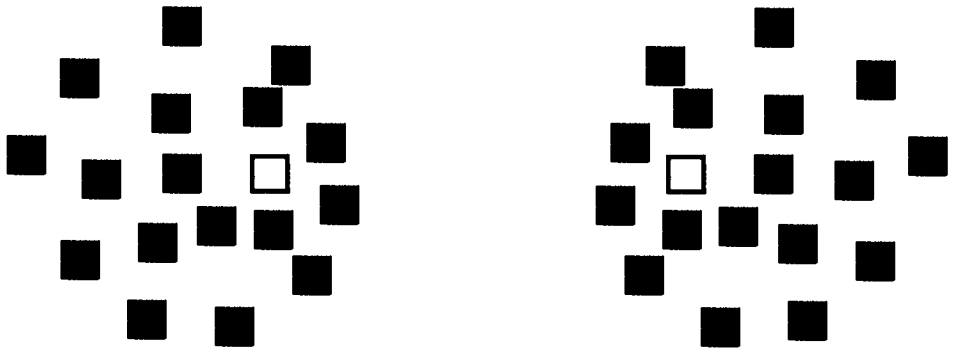
The example of a face, described above, can be used to illustrate this idea. Suppose that the separation of the eyes is an explicit measurement recorded at one spatial scale and that, for example, the separation of the left eye and the nose is also recorded at the same scale. These two measurements could be compared directly. In a similar way, the separation of the eyes in one face might easily be compared to those in another face (see figure 5.1 (a)) because both are single, explicit measurements made at one scale.

On the other hand, the separation of the left eye in one face and the right eye in a second face in the same scene is *not* explicit in a hierarchical model. This is illustrated in figure 5.1 (a). It is a different case because the features are in different groups. The separation can be derived because the distance between the faces is known, but only by combining information recorded at different scales (that is, the distance of one eye from the centroid of the face (a fine scale measurement); the distance between the centroids of the two faces (a coarse scale measurement); and the distance of the other eye from the centroid of that face (a fine scale measurement)).

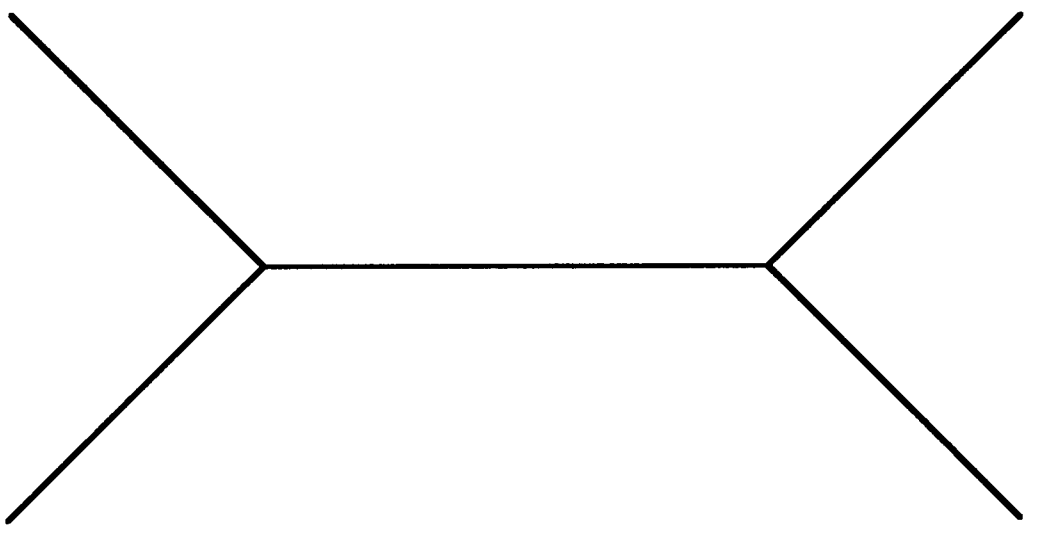
There is evidence that human observers are not very accurate at performing a judgement of this type (they tend to make consistent errors). Morgan, Hole and Glennerster (1990) describe a task which is analogous to the judgement of the distance between eyes in different faces. They presented two clusters of dots each containing a target dot (of a different colour to the other dots in the cluster) as illustrated in figure 5.1 (b). Subjects were asked to compare the separation of the target dots with the separation of two comparison dots, presented on their own, a second later. Morgan et al.. systematically varied the offset of the target dots from



(a)



(b)



(c)

Fig 5.1

(a) In a hierarchical system, the distance between the eyes in these two faces would not be recorded explicitly, because the eyes are within different groups.
(b) An example of the stimuli used by Morgan et. al. (1990) (schematic). They found systematic biases in subjects' judgements of the separation of the target dots (white).
(c) The biases were similar to those which are found for the Müller-Lyer stimuli, such as this "fins-out" figure.

the centres of the clusters. Their results show that, despite being told to ignore them, subjects' responses were always influenced by the clusters. Subjects responded as if they saw the target dots closer to the centre of the cluster than they really were (i.e the measured bias in subjects' responses was always in the opposite direction to the offset of the target dots from the cluster centre). An example of Morgan et al.'s results, is shown in figure 5.2 (a).

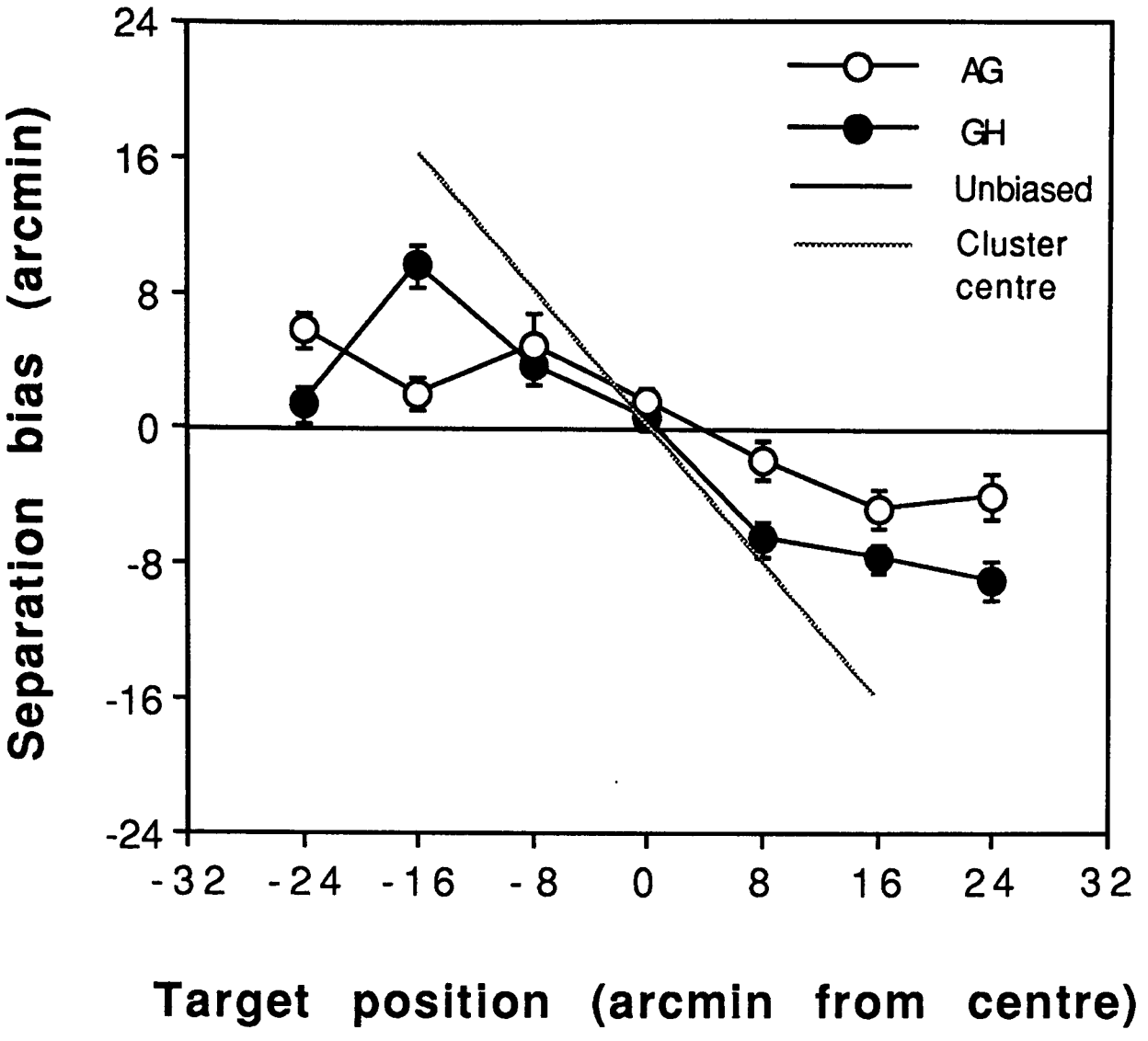
The experiment of Morgan et al. (1990) illustrates an important point about the encoding of position in the visual system. The results of their experiment can best be described by saying that subjects behaved "as if" they saw the target dots much closer to the cluster centre than they really were. But this is not what subjects reported. A control experiment illustrates that subjects could locate the target relative to the cluster very precisely (results shown in figure 5.2 (b)). This supports subjects' claims that in the main experiment they could *see* that the target dots were displaced from the cluster centre, but that despite this they did not (apparently) use the information in judging the separation of the dots. Taken together these results suggest that different information about the target's position is used depending on the task.

If this is the case, then a hierarchical model is a very appropriate one. A hierarchical description of the stimulus (e.g. the one shown in figure 5.1 (b)) would record the separation of the clusters (perhaps of their centroids) at a coarse scale and at a finer scale record the position of the dots, including the target dots, with respect to the cluster centroids. Nowhere in this model would the separation of the target dots be recorded explicitly. This might explain subjects' responses. That is, (assuming the stimulus is represented hierarchically in the human visual

---

Fig 5.2

(a) Data re-plotted from Morgan et al. (1990) for clusters of black dots and white target dots. Data are shown for two subjects. Biases for the separation task (a positive bias corresponds to an over-estimation of the separation) are plotted against the position of the target dots relative to the cluster centres (a positive target position means that the separation of the targets was larger than the separation of the cluster centres). Data would fall along the horizontal line if subjects made a veridical judgement of the target dot separation and along the diagonal line (dotted) if they reported the separation of the cluster centres. Error bars show the standard error of the mean. (b) This plot shows, for the same two subjects and for the same type of cluster as in (a) (i.e. a 16 arcmin radius cluster of black dots with a white target dot), thresholds for localising the target dot relative to the cluster centre. Data for three eccentricities are shown, 100 arcmin (the separation of the clusters in the above experiment), 50 arcmin and foveal viewing.



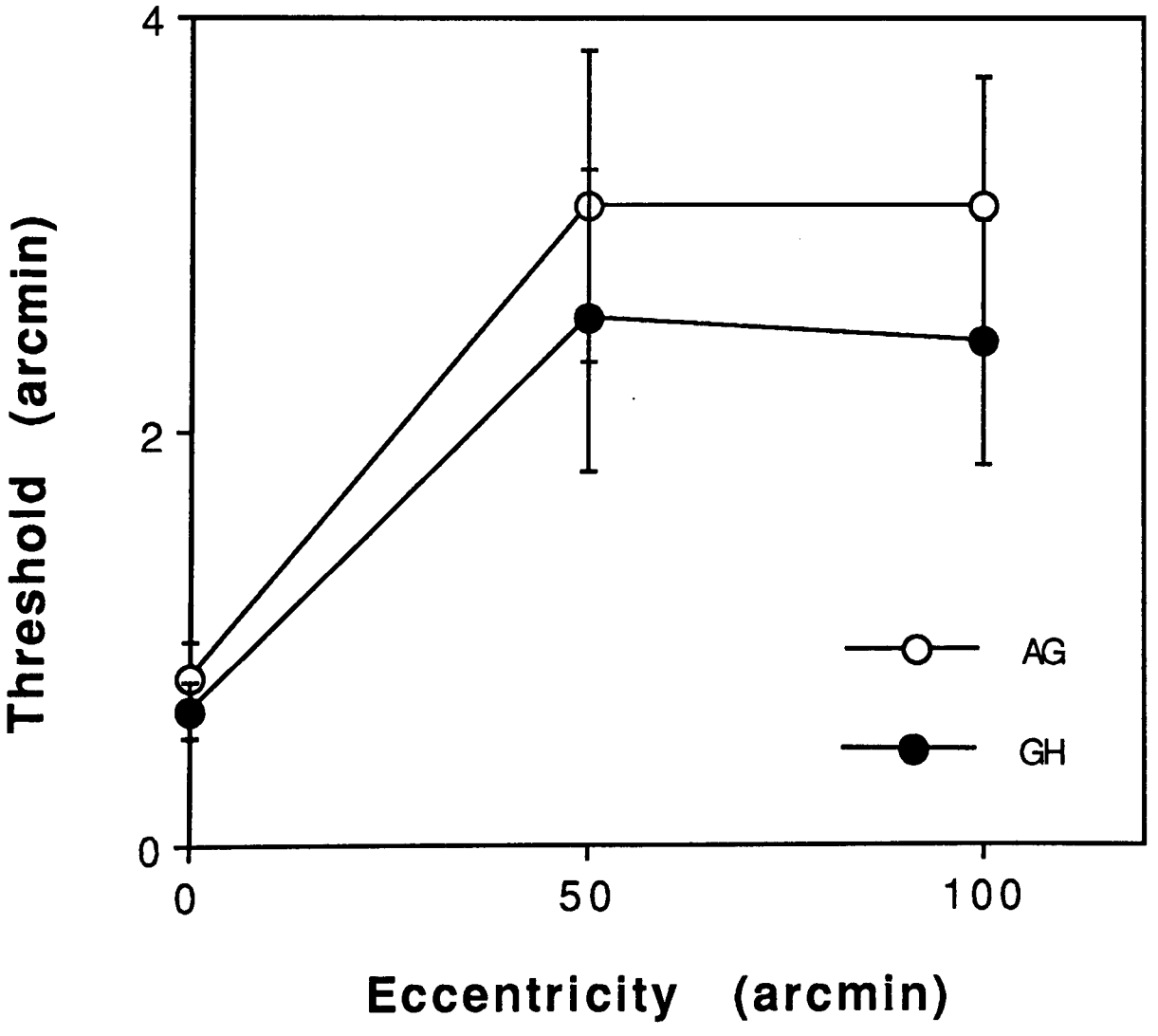


Fig 5.2 (legend on previous page)

system) it may be that distances which are not recorded explicitly cannot be compared accurately (i.e. without bias). When forced to make a comparison between two "implicit" or "compound" distances (i.e. ones made up from measurements made at more than one scale) the visual system may compromise and use as an approximation a measurement made at one (coarse) spatial scale. Note that the comparison of a single coarse scale measurement in each eye's image would not necessarily affect sensitivity despite introducing a significant bias. This is what Morgan et al.. reported for the cluster stimulus.

By contrast, the task in which subjects had to determine the location of the target with respect to the cluster centre does not rely on measurements at more than one scale. The location of each dot in relation to the coarse scale centroid (cluster centre) is recorded explicitly (as a single measurement) in a hierarchical model. It might be expected, then, that subjects would perform well at this task, and the results shown in figure 5.2 (b) confirm this. In summary, Morgan et al..'s results fit well with the predictions of a hierarchical model.

The stimulus Morgan et al.. used can be considered as a version of the Müller-Lyer illusion and in this sense their result is not a new one. (Both are shown in figure 5.1: the target dots are equivalent to the shaft ends, the clusters to the fins.) However, in many ways the cluster stimulus is a simpler stimulus, particularly in relation to its analysis at different spatial scales. It emphasises the coarse grouping and the relation of the target to the coarse centroid measurement.

This is not so clear cut for the Müller-Lyer stimulus. Even so, there is still a coarse length measurement which differs from the veridical or fine scale length (Ginsburg, 1978; Cornell, 1978; and section 5.5). Also, it is reasonable to suppose that at the finest scale the length of the whole shaft is too long (relative to filter size) to be measured accurately by a "fine scale mechanism" and, in a hierarchical representation, that this length would be recorded at a coarse scale. The position of the ends of the fins, then, would be recorded as a fine scale "adjustment" to a coarse scale measurement, in a manner exactly analogous to the cluster stimulus.

In support of this interpretation, psychophysical data on the extent of the Müller-Lyer illusion have been shown to fit very well with predictions made on the basis of coarse filtered versions of the Müller-Lyer figures (Ginsburg, 1978; Cornell, 1978). Their interpretation has been criticised because it fails to explain why

information from high spatial frequency channels is apparently not used in the length discrimination task (e.g. Carlson et al., 1984). Ginsburg himself recognises this problem with the coarse-filtering hypothesis:

> *"..it is reasonable to ask why, if these data suggest that the perceptual space for distortions of illusions is the low-pass filtered image, the details of the illusion are seen as distorted, i.e. the high spatial frequencies of the objects."*
>
> (Ginsburg, 1978, p69)

The answer, he says, must lie in the combination of filter outputs but does not provide one. A hierarchical model, on the other hand, can explain this apparent paradox.

A final point worth considering is why, if the visual system finds it so difficult to combine information across scales it does not make "mistakes", similar to that illustrated in the Müller-Lyer illusion, more often. The reason may be that the task demanded of subjects in comparing the Müller-Lyer figures is, visually at least, a very complicated one, and one it is very rarely asked to make. Morgan et al.. conclude:

> *"The term "illusion" is a value judgement [made] by the experimenter, because the subject has not made exactly the judgement that the experiment intended. We have suggested that...the visual system is highly constrained in the nature of the judgements it is able to make. It is possible to formulate verbal instructions which the visual system is unable to carry out exactly."*
>
> (Morgan et al.., 1990, p1809)

In most situations measurements at one scale probably suffice to carry out visual tasks successfully. It is much more common, for example, to be asked to compare two people's faces and the separation of their eyes than to judge the distance between one person's left eye and another's right eye. And if the Müller-Lyer figures were made of wire we would have no difficulty in grasping them at each end. The hand could be directed towards the coarse scale centroid and then, using fine scale information, guided towards the shaft end. There is some evidence that this is what subjects tend to do with their eyes when inspecting this type of figure (Coren and Hoenig, 1972; Coren, 1986; Findlay, 1980).

5.1.2 Direct evidence for a hierarchical model.

Although the Müller-Lyer illusion and the related cluster illusion (Morgan et al., 1990) can be explained in terms of a hierarchical encoding of length, neither prove that this is the system used by the visual system. Even if it is accepted that a coarse

scale version of the stimulus is used for length comparisons, other reasons for the fact that fine scale information is not used might be put forward, or the issue can be left unanswered (e.g. Ginsburg, 1978).

In order to provide direct evidence for a hierarchical model, it must be demonstrated that processing of fine scale information is influenced by the coarse scale structure of the image. For instance, in stereopsis, experiments demonstrating that the matching of fine scale features, or their perceived depth, depended on their position relative to coarse scale features rather than on their absolute retinal disparity, would be good evidence in favour of a hierarchical model. This type of experiment has been carried out by, among others, Mitchison and McKee (1987a and b) with respect to matching and by Mitchison and Westheimer (1984) with respect to perceived depth. These papers are discussed in the next section.

5.1.3 Mitchison and McKee

As discussed in chapter 1, the work of Mitchison and McKee (1987a and b) poses an important challenge for any theory of stereopsis. The aim of their experiments was to discover the matching rules in human stereopsis. Mitchison and McKee summarise their results and model by saying:

> *"Matching in stereograms made of horizontal rows of dots can be described as follows: characteristic features, such as edges and gaps in rows, have unambiguous matches in the two eyes, and these features are matched first. A plane interpolated between the positions in depth assigned to these features then guides the matching in the intervening sets of regularly spaced points which have potentially ambiguous matches. The intervening points are matched so that their disparity with respect to this interpolation plane is minimized. The "nearest disparity" rule describes matching for slanted interpolation planes as well as for fronto-parallel planes."*
>
> (Mitchison and McKee, 1987a, p285)

Although couched in very different language, these matching rules are very similar to those discussed in chapter 3. In fact, the results presented in Mitchison and McKee (1987a and b) provide strong evidence in favour of a hierarchical matching scheme. Although some of their experiments have been described already (chapter 1), it is worth considering the results again in the context of a hierarchical model.

The stimuli in their experiments consisted of a grid of regularly spaced dots. Mitchison and McKee found that the disparity of the dots at the left and right hand

edges of the grid (which was varied, usually by the same amount) had an important influence on the perceived depth of the internal dots of the grid even though the latter all had matches in the fixation plane. At short exposures (150 ms) the internal dots were perceived as having the same depth as the edge dots, despite the fact that the internal dots had no discrete match at this depth. In a hierarchical model the entire grid of dots could be described as a coarse scale blob. The outline of the blob (and to a lesser extent its centroid) will be determined by the edge dots. At a coarse scale of analysis, which, in a coarse-to-fine model, is appropriate for a short exposure duration, the exact position of individual dots is not yet determined although they contribute to the statistical ("texture") description of the blob. So, it is entirely consistent with a hierarchical model that the disparity of the whole blob (determined by the disparity of the edge dots) should be ascribed to the internal dots as well.

At longer exposures, Mitchison and McKee found that the internal dots were perceived to lie in one of two discrete planes, either the fixation plane or a forward plane consistent with each dot being matched with its nearest neighbour in a crossed disparity direction. In other words a discrete match was found for each dot. The particular match established was determined by the disparity of the edge dots. If the disparity of the edge dots was less than half the size of the dot spacing then internal dots were matched in the fixation plane. If the edge dot disparity was between half and one dot spacing then internal dots were matched in the forward plane.

This result too, fits with a hierarchical model. At longer exposure durations the fine scale dots within the blob will be resolved. The position of these dots will be described in relation to the coarse scale centroid, i.e. in this case relative to the position of the edge dots. Dots are then matched each with its nearest neighbour. When the whole blob has a disparity greater than half the dot spacing, the nearest neighbour of each internal dot (if their positions are described hierarchically) will be the forward match.

The model which Mitchison and McKee propose is slightly different from a hierarchical one. They emphasise the unambiguous matches for the edge dots rather that their effect on the coarse scale outline. They also consider that the reason internal dots are not matched at short exposure durations is because there are several possible matches rather than that they are treated as fine scale texture. But the "nearest disparity" rule they describe for determining which match for

internal dots should be chosen is no different from a hierarchical "nearest neighbour" rule. (In the fixation plane the rules are: choose the match whose disparity is closest to zero; or, choose the match with the most similar x-co-ordinate. Clearly these are equivalent. Off the fixation plane the rules become: choose the match whose disparity is closest to the disparity of the edge dots; or, choose the match with the most similar hierarchical co-ordinate [which is determined by the edge dots]. These are also equivalent.)

Mitchison and McKee' extended their model to matching in slanted planes. They describe experiments using the same grid of dots as before (i.e. all the internal dots had matches in the fixation plane) but in which the edge dots had unequal disparities. The perceived depth of the internal dots (at long and short exposure durations) matched the predictions of a slanted interpolation plane very well. (That is, for short exposures the internal dots were perceived as lying on an interpolation plane between the edge dots. At long exposures the internal dots were seen as lying either in the fixation plane or in the forward-match plane, depending on which was closer to the interpolation plane. See Mitchison and McKee, 1987a, figure 8 and 9.) These results are interesting from the point of view of a hierarchical model. Mitchison and McKee's slanted interpolation plane is equivalent to describing the position of fine scale features in terms of the *width* of the coarse scale blob in each of the monocular images separately. This was illustrated in section 3.2.2.

Finally, Mitchison and McKee demonstrated that the matching of dots in a grid which was *curved* in depth in a horizontal direction (in fact made up of three adjoining planes, their figure 3b and c) can best be modelled by assuming that an interpolated plane is drawn through the edge dots of the grid (their figure 3d). This is also the case when the whole curve is slanted in depth - in this situation a slanted interpolation plane best predicts the matching of the dots subjects reported (their figure 7). In other words, it is not possible to use a curved "interpolation surface" to direct the search for matches. For a hierarchical model this corresponds to the fact that the position of fine scale features cannot be described relative to the coarse scale curvature of a blob (at least in a horizontal direction) because there is no such thing. Instead, fine scale features must be defined in terms of the coarse scale blob width (i.e. slanted interpolation plane)* .

---

* On the other hand, the curvature of a blob in a vertical direction can be determined at a coarse scale (section 3.2). Mitchison and McKee did not investigate the possibility of an interpolation plane slanted or curved about a vertical axis (although this may be equivalent to a series of (narrow) horizontal planes at different y-co-ordinates).

In summary, the two matching schemes, despite being described in different terms, are so similar as to be almost indistinguishable in their predictions, certainly with respect to the stimuli discussed so far. One important difference is that Mitchison and McKee assume that the process of drawing an interpolation plane through unambiguously matched points happens only once. Mitchison (1988), who expands on some of the ideas in Mitchison and McKee, describes the process as a stage of "coarse segmentation" followed by a stage of cross correlation within each segment. This *is* different from a hierarchical scheme: a hierarchical model is like the first stage of Mitchison and McKee's model (i.e. drawing an interpolation plane) but which is then repeated many times at finer and finer scales.

Mitchison and McKee admit that some aspects of their experiments could be explained by a filtering hypothesis (i.e. scale-space or multiple-channel model) but argue strongly against this approach as a general explanation. Mitchison and McKee (1987b) show how a filter model such as the one Marr and Poggio (1979) proposed could not predict their results. This does not necessarily imply that a hierarchical scale-space model, based on relative co-ordinates, is also ruled out.

The basis of their argument is that fine scale information is available to the stereo matching system even at an exposure of 160 ms and that, if this is the case, then, according to Marr and Poggio's scheme, the dots should be matched in the fixation plane.

The two examples of the stimuli they use to make their case are shown in figure 5.3. The first is, as described above, a grid of regularly spaced dots in which the end dots have been given an equal (convergent) disparity The shift of the end dots is 0.7 times the inter-dot spacing and, when tested at 160 ms exposure, the dots in the centre of the grid appeared forward of the fixation plane by about 0.7 of the "full forward match" depth, i.e. by an amount equivalent to the disparity of the edge dots. The separation of the dots in this example was 5 arcmin. To explain this result within Marr and Poggio's model, Mitchison and McKee point out that the scale of analysis would have to be larger than any filter which could resolve dots spaced 5 arcmin apart - if the dots were resolved before any eye movements had taken place then the nearest match would certainly be the fixation plane match.

In their second example, they used much more closely spaced dots and their result illustrates that the scale of analysis must in fact be considerably smaller. The

stimulus they used is shown in figure 5.3 (below). As in the first example, the edge dots were shifted by 0.7 times the inter-dot spacing but one of the central dots was also shifted. This dot was given a very small displacement in the same direction in each eye, i.e. no disparity was added but an "unambiguous feature" (e.g. the wider gap to the left of the dot) was created. This feature should, in Mitchison and McKee's theory, be matched in the fixation plane. The predicted perception, on the basis of Mitchison and McKee's theory, i.e. of matching of unambiguous features first followed by interpolation between these features, is shown in figure 5.3. Also shown is the perception subjects reported, i.e. the dots in the centre of the grid were seen as forward of the fixation plane but not by as much as the edge dots. What this demonstrates is that, at 160 ms exposure, the small "unambiguous" gap in this figure was resolved and influenced the matching process, hence the dots in the previous example must also be resolved, and Marr and Poggio's theory falls.

On the other hand, both Mitchison and McKee's model and a hierarchical one can account for the matching in these situations. This is because they do not depend on eye movements and unlike Marr and Poggio's model, matches in the fixation plane are not necessarily the nearest neighbours. In both examples shown in figure 5.3 the coarse scale match would be found at a disparity of 0.7 times the inter-dot spacing. Then the position of the internal dots would be determined. In the first example, figure 5.3 (top), the nearest neighbour matches would all be forward

---

Fig 5.3

The two rows of dots shown above (a) represent the images presented to the left and right eyes in Mitchison and McKee's (1987b) experiment. (In fact, a whole grid of dots was presented of which only one row is shown.) The dots which are lined up vertically in this figure were presented in the fixation plane in the experiment (i.e. any horizontal shift is equivalent to a disparity). The perception of the dots, given unlimited exposure, is shown in the plan view below. The central dots were seen in front of the fixation plane by an amount suggesting that each dot in the left eye was paired with the neighbouring (left hand) dot in the right eye's image. The edge dots were seen as lustrous, having no match (shown as open circles). For an exposure duration of 160 ms, the grid appeared fronto-parallel at a depth between the fixation plane and the full forward match depth (e.g. at the depth of the open circles). The dot spacing in this example was 5 arcmin.
In the example illustrated below (b) there were eight more closely spaced dots but one inter-dot gap is wider than the rest, providing an unambiguous feature to guide the matching of the dots either side of the gap. A plan view of the perception reported by subjects for this stimulus is shown underneath.

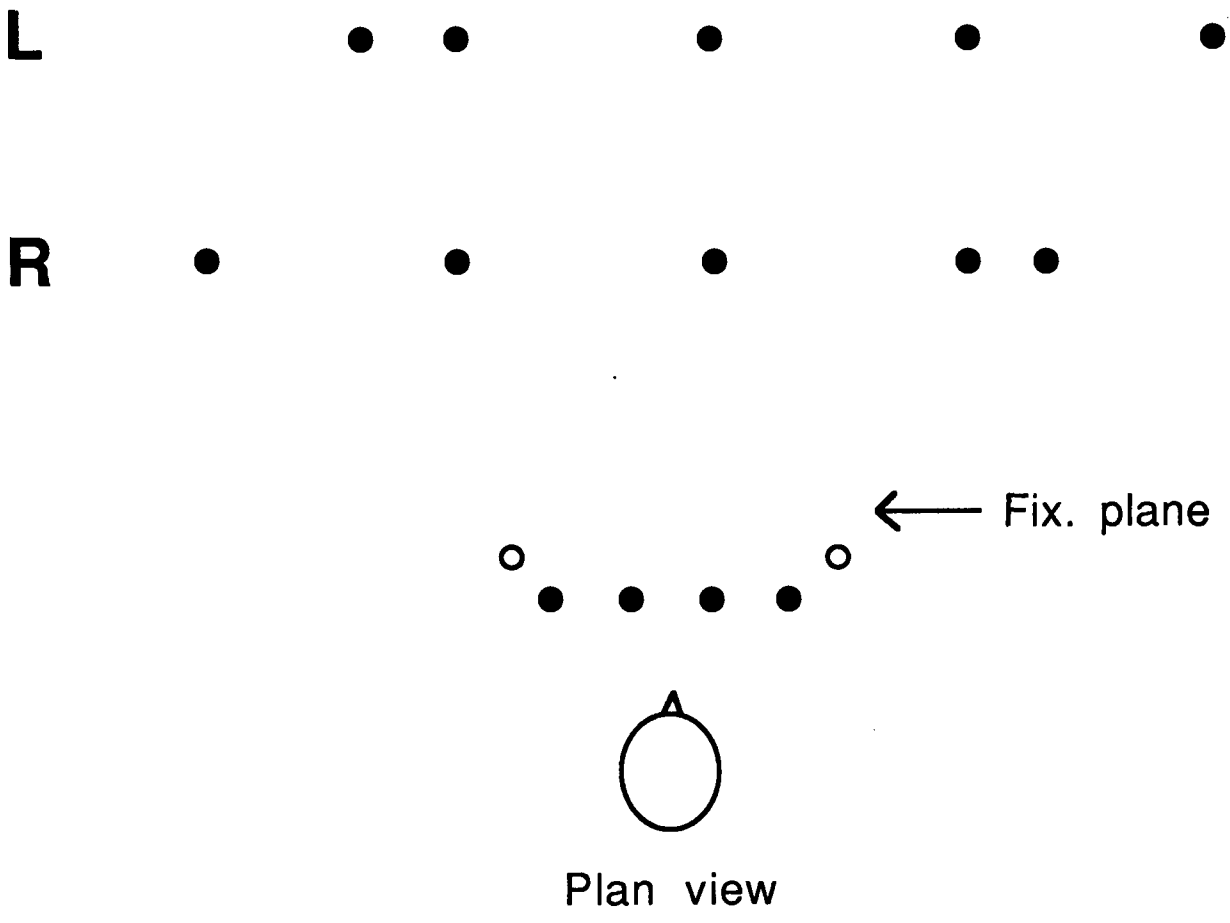


Fig 5.3 (legend on previous page)

matches. In the second example the larger gap in the centre would divide the dots into two "medium" scale blobs which would slant back towards the fixation plane, as subjects reported. At the finest scale individual matches for the dots would be found close to these slanted planes.

Mitchison and McKee's model, because it talks only of "unambiguous matches" cannot distinguish between the edge dots and the "gap" in the centre of the grid, both of which are unambiguous. Why then, according to their model, are the edge dots seen at their true depth at 160 ms but not (yet) the centre of the grid (which was seen at an intermediate depth)? There seem to be degrees of ambiguity. If Mitchison and McKee's model was modified to incorporate interpolation planes drawn at successively finer scales then it could account for the precedence of the edge dots. But it would also make their model formally equivalent to a hierarchical one.

5.1.4 Mitchison and Westheimer

In the previous section it was established that an interpolation plane, as Mitchison and McKee (1987a) describe it, is very closely related to the concept of a hierarchical co-ordinate system. Mitchison and McKee's experiments were primarily concerned with the matching rules used by the visual system and provided good evidence that an interpolation plane (or, equivalently, a hierarchical nearest-neighbour scheme) could predict the matches subjects perceived in their stimuli. Mitchison and Westheimer (1984) investigated the perceived depth of features in simple line and dot figures. They developed the concept of "salience" which is closely related to Mitchison and McKee's idea of an interpolation plane (and to a hierarchical scheme) since the salience of a point is high whenever it lies away from the local interpolation plane. Subjects were asked to judged the relative depth of two vertical lines (which, in most of the experiments these separated by 12 arc min). In one experiment, Mitchison and Westheimer found that adding a single line to one side of and at a different depth from the test pair caused a bias in subjects' judgements. In that experiment, the test lines were seen as fronto-parallel when their disparity was close to that of the plane joining the edge dots (the interpolation plane), not when their disparity was close to zero. Many of their findings were consistent with a model in which depth judgements are made relative to a local interpolation plane and not on the basis of the absolute disparity of features. The most extreme example they gave of insensitivity to absolute disparity differences was an experiment in which the two vertical lines were shown in front

of a grid of regularly spaced dots which were slanted about a vertical axis. Subjects were insensitive to the slant of the grid and judged the vertical lines to be fronto-parallel when they had approximately the same slant as the grid. The grid appeared to act as a reference plane against which precise judgements of relative slant could be made.

The interpretation Mitchison and Westheimer give does not include a discussion of spatial scale, but their data show that the perceived depth of features is strongly influenced by the disparity of surrounding features. In many cases the results can be interpreted in terms of disparities measured with respect to a reference plane determined at a coarse scale of analysis. This, as discussed in the previous section, is the essence of the hierarchical model put forward in chapter 3.

Neither Mitchison and McKee nor Mitchison and Westheimer considered a hierarchical spatial scale model in discussing their results. Mitchison and McKee supposed that an interpolation plane was derived from features which could be paired unambiguously with features in the other eye's image. The experiment described in this chapter is similar in some ways to Mitchison and McKee 's experiments on slanted grids (as discussed in section 5.1.3) but, whereas Mitchison and McKee's results could be interpreted *either* in terms of a coarse-to-fine hierarchical model *or* in terms of matching of unambiguous features, the Müller-Lyer figures used in the experiment described in this chapter fit much better with a filtering model than with Mitchison and McKee's model. Thus, the experiment described in this chapter builds on the findings of Mitchison and McKee but seeks to distinguish their theory from a hierarchical scale-based one as put forward in chapter 3.

5.1.5. Wilson, Blake and Halpern

A recent paper by Wilson, Blake and Halpern (1991) claims to provide evidence for a coarse-to-fine constraint on disparity processing. They used filtered stimuli rather than dots and lines as Mitchison and McKee (1987) and Mitchison and Westheimer (1984) had done. The patterns Wilson et al. used were D6 patches (a one-dimensional sixth Gaussian derivative), similar to gabor patches. They measured stereoacuity and diplopia thresholds for the D6 patches under a range of conditions. They found, in line with the results of Schor, Wood and Ogawa (1984), that diplopia thresholds increased as the centre frequency of the D6 was reduced. They also found that superimposing on the pattern a cosine grating whose frequency was two octaves lower that the D6 reduced diplopia thresholds to about

a quarter of the value when no grating was present. However, a grating with a frequency *4* octaves below that of the D6, or two octaves higher than the D6, had no significant effect on diplopia thresholds.

Wilson et al. interpret their results as evidence for an interaction in the disparity domain between different spatial frequency tuned mechanisms. They do not discuss the possibility that an interaction between filter outputs occurring before the correspondence process might also account for their results. There is some evidence in their paper in favour of such a hypothesis. For instance, the example they give of a stimulus in which the D6 is perceived as diplopic (their figure 1B) shows the D6 in one eye's image on a bright stripe of the cosine grating and the corresponding D6 in the other eye's image on a dark stripe. This is not the case for either of the stimuli which can be fused (figure 1A and C), i.e. the patterns with no grating or a much lower frequency grating.

For a D6 to be in the centre of a bright strip in one eye's image and in the centre of a dark strip in the other eye's image requires it to have a disparity $\lambda/2$, where λ is the period of the cosine grating. The minimum disparity leading to diplopia might be expected to be smaller than this value. Wilson et al. measured the diplopia threshold for patterns like that illustrated in their figure 1B using relatively high frequency gratings (their figure 3) and relatively low frequency gratings (their figure 4). In both cases diplopia thresholds were about $\lambda/5$.

Additional evidence comes from the experiment Wilson et al. report using a cosine grating slanted in depth about a vertical axis. (The frequency of the grating was 2 octaves below that of the D6.) The range over which the D6 was fused was centred on the plane of the grating rather than the plane of fixation. This is compatible with a theory in which the difference in the relative phase of the D6 and cosine grating in the two eyes' images is important in determining diplopia thresholds.

The form of the coarse scale MIRAGE response to a D6 and a cosine grating depends on their relative phase, and is very different when the gabor lies within a bright stripe than when it lies within a dark stripe. An experiment by Rentschler and Treutwein (1985) uses stimuli in which the relative phases of high and low frequency components is varied. It is discussed by Watt (1985) who shows that the grouping of fine scale features in the MIRAGE response can be quite different for different phase relationships.

In summary, the claims made by Wilson et al. must be treated with some caution since the assumptions they make, in particular that the cosine grating and the D6 are processed and matched in independent channels, are not necessarily valid.

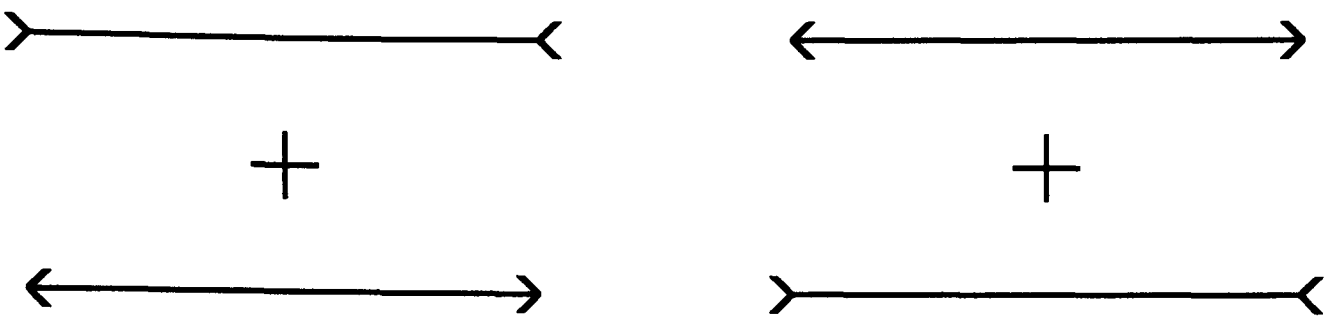
5.2 The rationale of this experiment

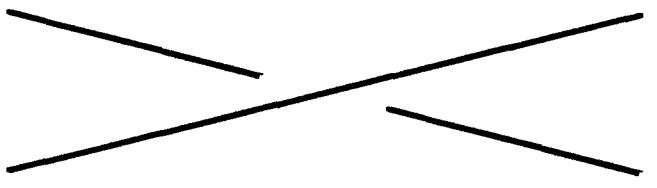
The hypothesis underlying the experiment described in this chapter is that left and right eyes' images are analysed independently at a range of spatial scales and the information organised into two parallel hierarchies. According to this model, length comparisons between the two eyes' images are easily made at one scale but comparisons which require information at several scales are not so easily made. If approximations are made by using measurements recorded at only one spatial scale (a coarse scale) then "distortions" result. All of these characteristics would apply to comparisons of length for stimuli (databases) which were separated in space or in time just as much as they would apply to comparisons between the two eyes.

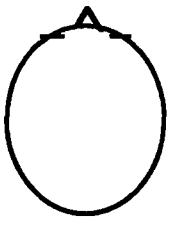
5.2.1 A new illusion

The main effect is illustrated in figure 5.4. In this figure, as in our experiment, the two halves of the Müller-Lyer illusion are shown separately to the two eyes (for example, the "fins-out" figure to the left eye, the "fins-in" figure to the right.). For those readers who can free-fuse, the horizontal shafts drawn in figure 5.4 should appear to be slanted in depth in opposite directions. It is *as if* one eye were presented with a long line and the other a short line. (Also shown in figure 5.4 is an example of a stereogram in which the lines presented to the two eyes really are of different lengths. When fused the lines in this example should be seen as slanted in depth by roughly the same magnitude as in the illusory case.)

The fact that a slant is perceived for this stimulus fits well with the predictions of a hierarchical model. If each eye's image were encoded hierarchically then similar distortions might be expected to affect comparisons of length between the two eyes' images as occur for comparisons of lengths in different parts of one image. The main purpose of the experiment described in this chapter was to explore in a more detail the "3-D" Müller-Lyer illusion and to make a quantitative comparison with the classical 2-D Müller-Lyer illusion.







Plan View

Fig 5.4

The two eyes' images are drawn on the left and right at the top of the page. Stereoscopic fusion of the images should result in a perception of the horizontal shafts as slanted in depth in opposite directions. (For cross-eyed fusion the top shaft should appear slanted with the left end nearer to the observer, as shown in the plan view.) The fins are diplopic. Note that very short fins have been used in this demonstration, which was found to produce the most convincing illusion. The shafts are all the same length, just as they are in the 2-D Müller-Lyer illusion. At the bottom is drawn a stimulus without fins in which the shafts really are of different lengths and hence when fused should appear slanted in depth like the sketch above it. The lines should slant by roughly the same magnitude as in the illusory case.

5.3 Experiment I: Comparison of the 3-D and 2-D Müller-Lyer illusions.

5.3.1 Subjects

The subjects were four adults with normal (6/6) or corrected-to-normal vision.

5.3.2 Apparatus

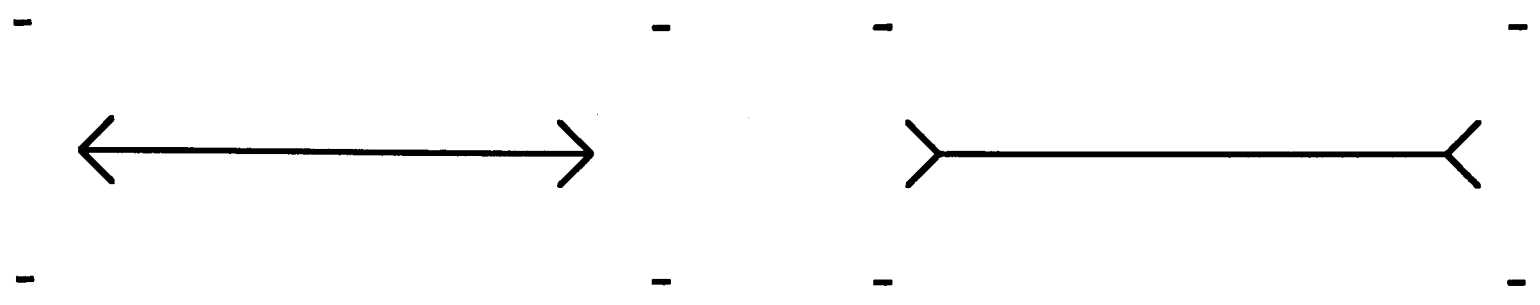
Stimuli were generated on a Macintosh II computer and displayed on two monitors viewed in a Wheatstone configuration at a distance of 57 cm (as described in section 4.4). In the 2-D experiment the displays on left and right screens were identical.

5.3.3 Stimuli

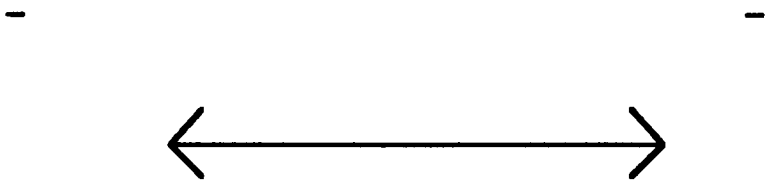
Figure 5.5 illustrates the stimuli. The lines of the figures were 2 arcmin (1 pixel) wide. The horizontal shaft was, on average, 400 arcmin long (see below for details of the psychometric procedure). The fins were drawn at 45° to the shaft. In the 3-D experiment a "fins-in" figure was drawn on one screen, a "fins-out" figure on the other. Four surrounding dots were drawn at the corners of a rectangle of height 200 arcmin and width 600 arcmin to provide a fronto-parallel reference plane (zero disparity). In the 2-D experiment, a "fins-in" figure and "fins-out" figure were shown one above the other. The vertical separation of the figures was constant for a given fin length (approximately 200 arcmin - in fact, 200 arcmin plus twice the height of the fins). The four surrounding dots were drawn at the corners of a rectangle of height approximately 400 arcmin (in fact, 400 arcmin plus four times the height of the fins) and width 600 arcmin. The luminance of the white line figures was 25 cd/m^2, the background 0.1 cd/m^2. The room in which the experiment was carried out was dimly lit.

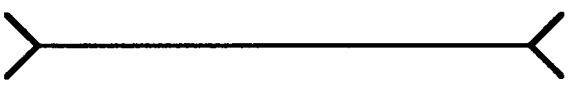
5.3.4 Psychometric procedure

A 2-alternative forced choice procedure was used. In the 2-D experiment subjects were asked to judge whether the top or bottom figure was longer. In the 3-D experiment they were asked to judge whether the shaft was slanted "left end towards me" or "right end towards me". That is, in both cases, the information which enabled subjects to make the judgement was the difference in length of the "fins-in" and "fins-out" figure (this difference in length was the independent variable in all the experiments). To prevent subjects from using the mean length of



3-D experiment





2-D experiment

Fig 5.5

In the 3-D experiment the fins-in figure was shown to one eye, the fins-out figure to the other as illustrated above. The four surrounding dots were drawn in the plane of the screen as a reference. In the 2-D experiment the figures were drawn one above the other.

the set of stimuli to make their judgements, a "jitter" of up to ±12 arcmin was added to the length of both figures (fins-in and fins-out) on any given trial. The horizontal position of the figures was also jittered to eliminate the possibility that subjects could use the vertical alignment between the ends of the shafts as a cue. In the 3-D experiment, similarly, the overall disparity of the figure was "jittered" from trial to trial by up to 12 arcmin so that subjects could not use the disparity of one end of the shaft with respect to the screen to determine the slant. In the 2-D experiment, the relative position of the fins-in and fins-out figures ("above" or "below") was varied randomly; likewise for the 3-D experiment, which figure appeared on the left screen and which on the right was altered randomly. This meant that any bias of the subject to respond in one way, for instance to see slant in a particular direction, would not affect the measurement of the extent of the illusion (although it would affect the slope of the psychometric curve). The range and mean of the independent variable (length difference) was determined from a pilot run of 50 trials so that, in each experimental run of 150 trials, length differences were approximately balanced about the point of subjective equality.

The stimuli were presented for up to 2 seconds after which the screen remained blank until the subject responded, triggering the next display. If the subject responded within 2 seconds of the stimulus onset the screen was blanked for 500 ms before the next trial.

5.3.5 A definition of the "extent of the illusion"

Figure 5.6 shows an example psychometric curve from the 3-D experiment. The proportion of responses given as "fins-in stimulus longer" is plotted on the ordinate. (The subject's responses were in fact "left end towards me" or "right end towards me" but the "fins-in" stimulus might be seen by the left or the right eye.) The actual difference in length of fins-in and fins-out stimuli is shown on the abscissa.

A cumulative Gaussian has been fitted to the data by probit. Its mean (50% point) is shifted to the right by 25 arcmin, i.e. the fins-in figure was in fact longer than the fins-out figure by 25 arcmin when the stimulus was seen by the subject as fronto-parallel. This shift, which is a measure of the extent of the illusion, is plotted against fin length in subsequent plots.

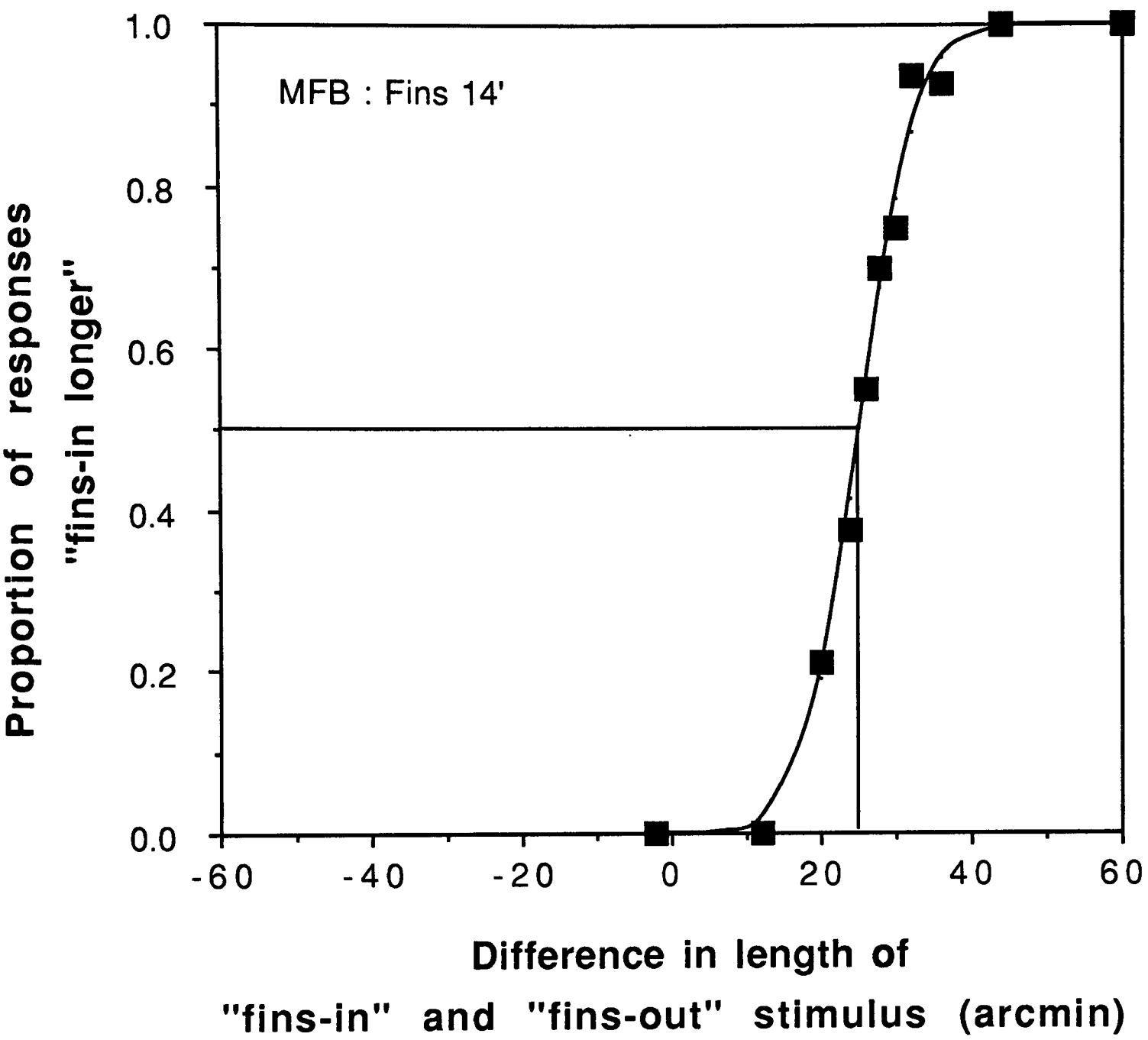


Fig 5.6

An example psychometric curve for the 3-D experiment for one subject and for a fin length of 14 arcmin. The proportion of responses given as "fins-in stimulus longer" is plotted on the ordinate. The actual difference in length of fins-in and fins-out stimuli is shown on the abscissa. Data points are fitted by probit.

5.3.6 Results

Figure 5.7 shows the extent of the illusion for the 2-D and 3-D experiments for three subjects. The finding that the 2-D illusion increases linearly with fin length and flattens off for very long fins is an old one (e.g Lewis, 1909). The extent of the 3-D illusion also increases linearly with fin length, flattening off at shorter fin lengths and returning to zero for very long fin lengths. (In the latter situation, the shaft is clearly and easily fused: the diplopic fins at each end of the shaft are seen to meet sharply, like a cross. This is not the case for intermediate fin lengths: here the perception is unstable and the shaft ends appear to slide over one another as the subject moves his eyes. For these fin lengths the variability of the responses was large, in most cases too variable for a reliable psychometric function to be obtained*.)

For fin lengths shorter than about 30 arcmin the extents of the 2-D and 3-D illusion are very similar. Figure 5.8 re-plots the data in this range to show each subject's results separately, enabling a comparison to be made between their performance on the 2-D and the 3-D task. The agreement is good. Possible reasons for this similarity are discussed in the modelling section.

5.4 Experiment II: Comparison of the cyclopean and 2-D Müller-Lyer illusions.

The implication of the 3-D Müller-Lyer illusion is that the processing of the image which gives rise to the apparent difference in length of the stimuli occurs before the point at which information from the two eyes is combined. The quantitative results of experiment I provide extra support for this idea.

5.4.1 Papert's demonstration

Julesz (1971) came to exactly the opposite conclusion when discussing the "cyclopean" version of the Müller-Lyer illusion (figure 5.9), originally published by Papert (1961):

---

* The slant judgement is unlike many psychophysical tasks in that increasing the "cue", the length difference, does not, beyond a certain stage, make the task any easier. This means that it was not possible simply to increase the range of cues in order to obtain a reliable psychometric function.

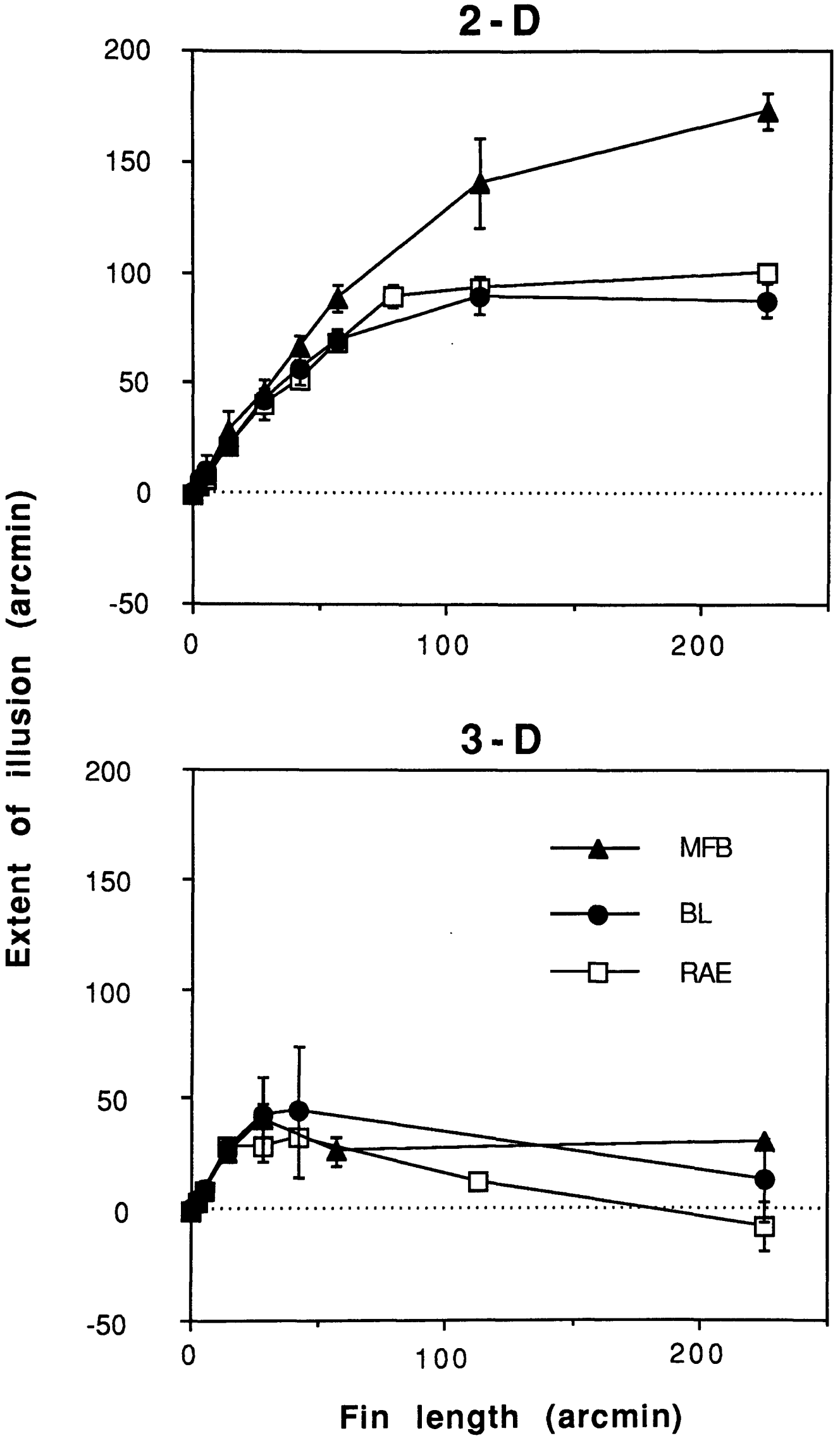


Fig 5.7

Data for the 2-D experiment are at the top, for the 3-D experiment below. The extent of the illusion (defined in the text) is plotted against fin length for three subjects in each case. Error bars show the standard error of the mean.

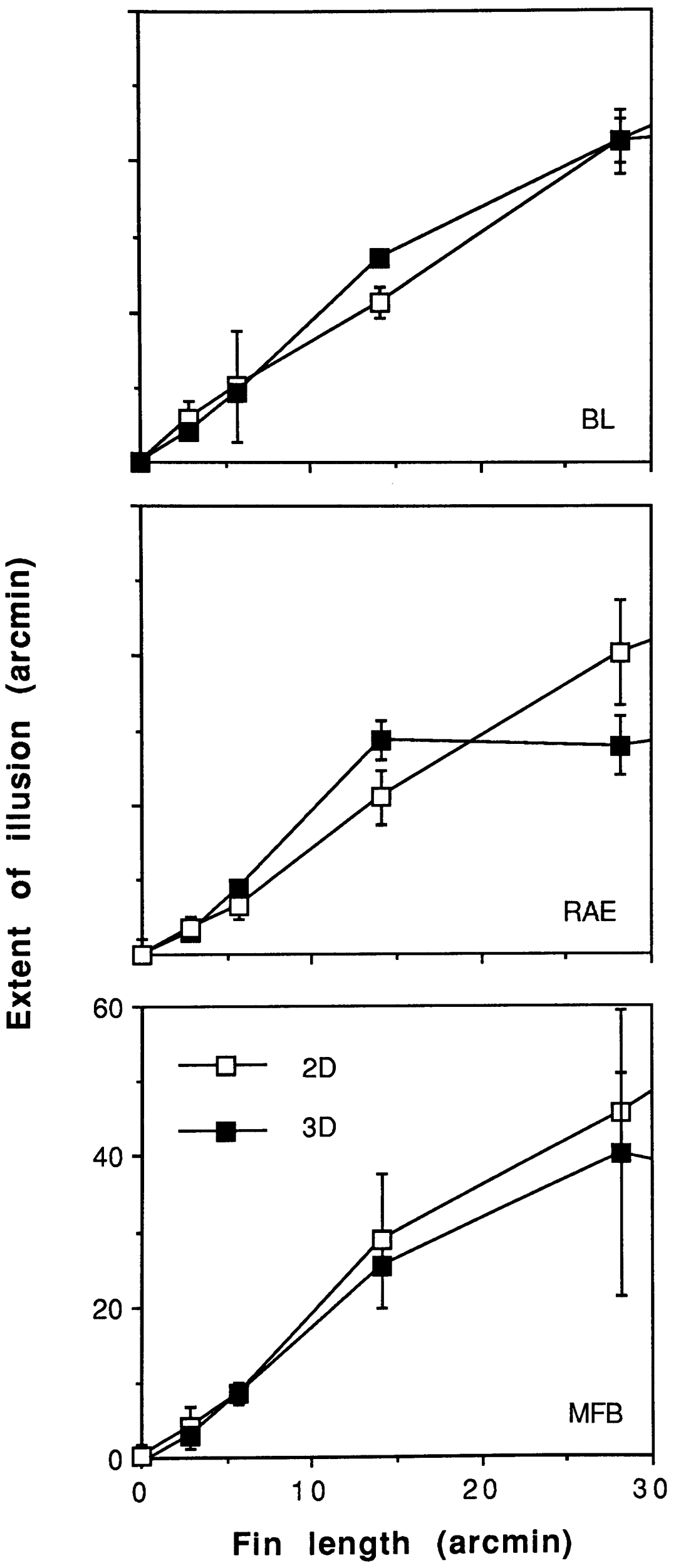


Fig 5.8

Data from figure 5.7 are replotted for short fin lengths showing 2-D and 3-D results for each subject separately.

> *"Suppose that when our cyclops looked at a [random dot stereogram] it saw the familiar Müller-Lyer arrowheads. And suppose that (as is indeed the case) it saw one line as illusorily longer than the other. Then we would know that the process responsible for that illusion can occur farther along in the visual system than is often thought. We would have added to our knowledge of the internal structure of the black box: the illusion occurs beyond the point where inputs from the two eyes are combined"*
>
> (Julesz, 1971, page xii)

This sounds like a paradox. It seems to imply that there must be two "sites" of the Müller-Lyer illusion, one before stereoscopic matching of the two eyes' images (to explain the 3-D illusion) and one *after* stereoscopic matching (to explain the cyclopean illusion).

But such a conclusion is not necessary. Julesz assumed that any effect which can be demonstrated in the form of a random dot stereogram must be due to a non-luminance-based mechanism ("beyond the point where inputs from the two eyes are combined"). This is because Julesz' (1971) model of matching in random dot stereograms was a "local-to-global" one (it used the individual dots as the starting point for the algorithm) and hence the figure (e.g. Müller-Lyer illusion) could only be detected once fine scale correspondences had been achieved. If this assumption is accepted, it rules out any possibility of "luminance" blurring as an explanation of the cyclopean effect.

For a coarse-to-fine matching algorithm the situation is different. Random dot stereograms can be matched at a coarse scale (i.e. using the output of coarse scale luminance filters), as discussed in the previous chapter. It is possible that matching of coarse scale blobs would reveal differences in length of the Müller-Lyer figures and hence lead to a hierarchical model of the cyclopean illusion as well. This idea is pursued in more detail in section 5.5.

In the experiment described in this section, the extent of the cyclopean Müller-Lyer illusion was measured for a range of fin lengths. The fins-in and fins-out figures were depicted as random dot figures standing out in front of a random dot background (see figure 5.9).

5.4.2 Stimuli

Left and right screens displayed random dot patterns with a 50% density in which, as illustrated in figure 5.9, two regions were given a crossed disparity with respect

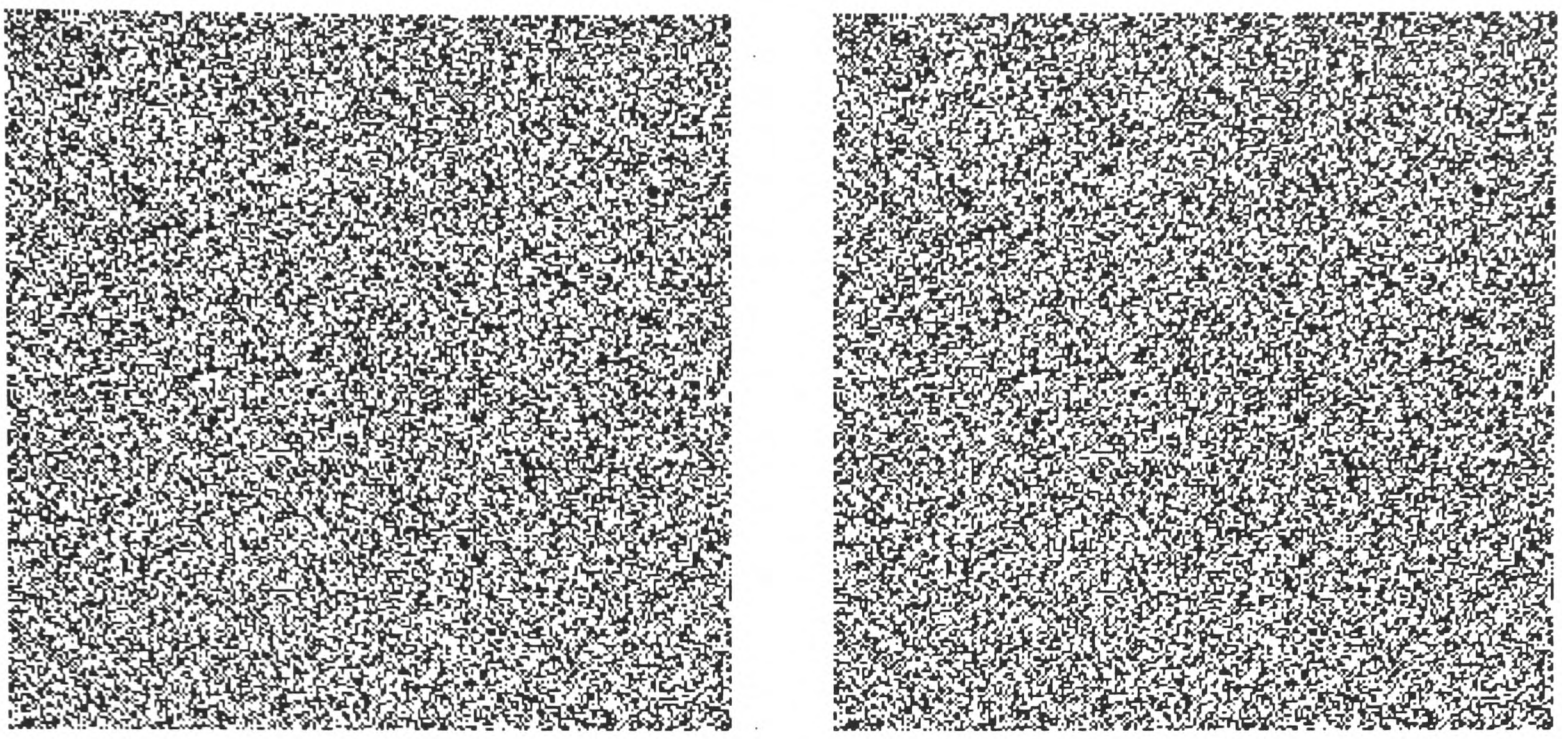


Fig 5.9

The two eyes' views are drawn at the top. When viewed stereoscopically the classical Muller-Lyer figures should appear as illustrated below. This type of stimulus was used in experiment II.

to the background, one in the shape of a "fins-out" figure and one a "fins-in" figure. Otherwise the stimulus parameters and task were as similar as possible to the 2-D task in Experiment I.

The random dot patterns were made up of 160 by 160 dots, each dot 4 arcmin square (2 by 2 pixels). The figures had a disparity of 8 arcmin. As in the previous experiment, one figure ("fins-out") was 400 arcmin long, plus or minus a small jitter of up to 12 arcmin, and the difference in length between "fins-in" and "fins-out" figure was varied between trials. Both the shafts and the fins (at 45° to the shaft) were 16 arcmin wide. (Narrower shafts and fins were used for the shortest fin length condition.) The vertical separation of the shafts was increased with fin length as in Experiment I so that the figures did not overlap. The horizontal position of each figure was jittered randomly within a small range (±12 arcmin).

The stimuli were presented for up to 2 seconds after which a random dot "mask" was displayed (also 50% density, correlated and zero disparity). The masking pattern remained until the subject responded and the response triggered the next display. If the subject responded within 2 seconds of the stimulus onset the mask appeared for 500 ms before the next trial.

5.4.3 Results

Figure 5.10 shows the results. As before the extent of the illusion is plotted on the ordinate, fin length on the abscissa. Data for the 2-D illusion are re-plotted here as a dotted line for comparison and, as Julesz has noted (1971), the agreement is good. Julesz took this as strong evidence for a single process responsible for the Müller-Lyer illusion whose input could either be defined by luminance or disparity. In the next section an alternative (hierarchical) model is considered.

5.5 Model

How far can a hierarchical model explain the results of these two experiments? This section examines how the length of the Müller-Lyer figures used in the above experiments is affected by filtering at a coarse spatial scale, taking as a criterion for "length" the separation of various spatial primitives such as peaks or zero-crossings at the ends of the shaft. As was pointed out in the introduction, it is

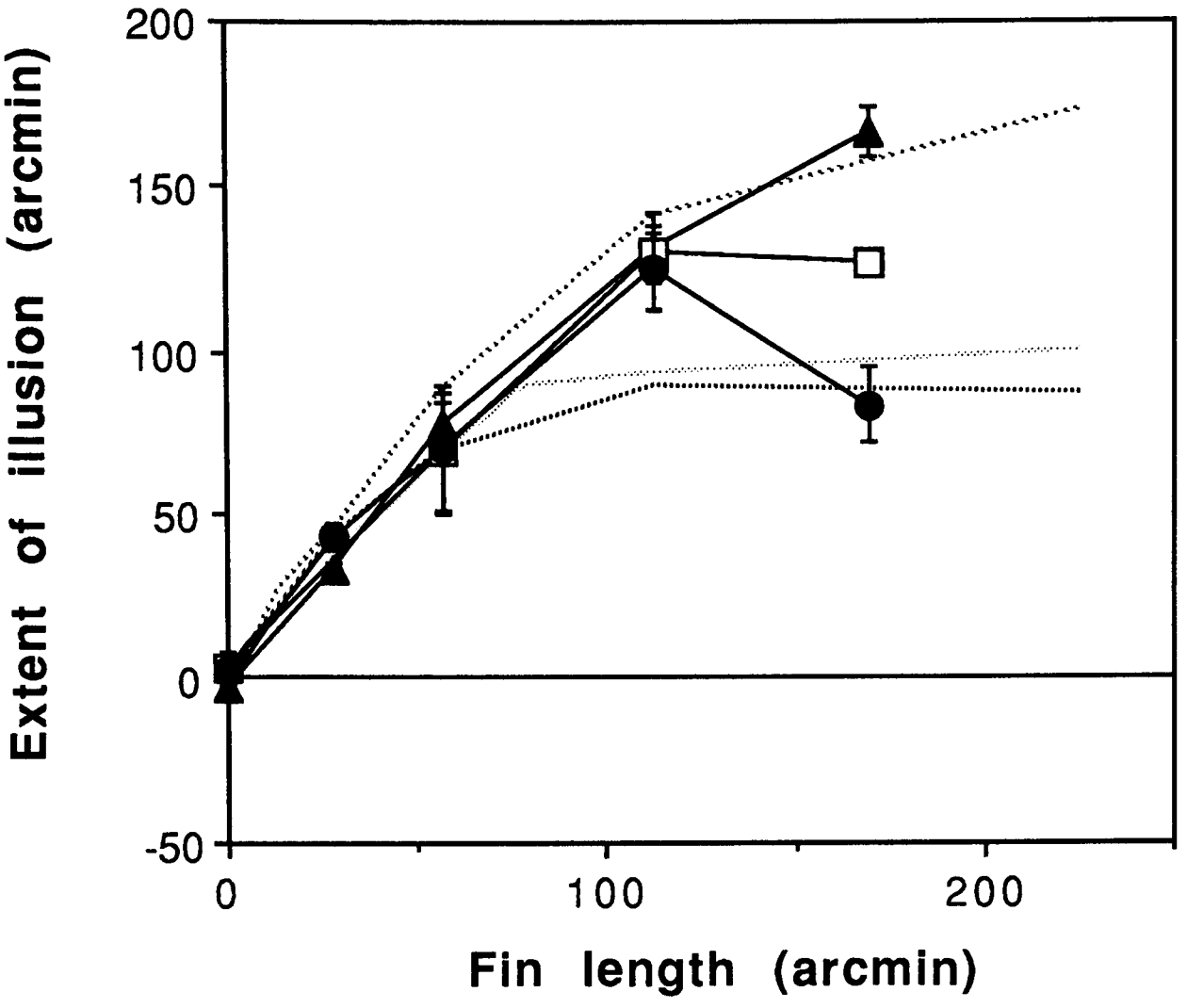


Fig 5.10

Data for the cyclopean experiment are shown for three subjects. Data for the 2-D experiment are also shown as dotted lines for comparison. As in figure 5.7, the extent of the illusion is plotted against fin length.

supposed that the stimuli are filtered at a whole range of spatial scales by the visual system but that large lengths, such as that of the shaft, are likely to be compared at a relatively coarse scale (which would apply to comparison between the two eyes' images, across space within one image or across time).

5.5.1 A filtering model for the 2-D and 3-D illusion

Figure 5.11 shows ingoing and outgoing fins filtered at two scales and for a range of fin lengths. Only the positive filter response is shown and each image is scaled to a constant contrast because it is the position of features, rather than their contrast, which is important for the model. The outgoing fins are shown on the left of each figure. The zero-crossing and peak in the filtered image, which signal the end of the shaft, move further and further out (to the left) as fin length is increased (down the page). This trend does not continue indefinitely. For the longest fin length and the smaller of the filters (bottom left) it can be seen that the tips of the fins are beginning to be resolved from the shaft end. At this stage, which happens at different fin lengths for different filters, the position of the zero-crossings and peaks in the filtered image is no longer affected by increasing fin length. This accounts for the flattening off in the predicted extent of the illusion shown in figure 5.12.

The ingoing fins are shown on the right of each figure and beneath each the filtered version. The position of the peaks and zero-crossings change very little as fin length is increased. This suggests (assuming a coarse-filter model) that the ingoing fins make only a small contribution to the magnitude of the illusion. When the extent of the 3-D Müller-Lyer illusion was measured psychophysically with fins from one figure omitted the results showed exactly this pattern (Table 5.1), i.e. that the slant caused by adding by outgoing fins was usually much greater than for ingoing fins. One other interesting result shown in this table is that, when a shaft with short ingoing fins was presented to one eye and a shaft with no fins to the

---

Fig 5.11

The result of filtering Muller-Lyer figures with coarse-scale Laplacian filters. The original figure is shown beneath each filtered version (only the positive filter response is shown here). The fins are out-going on the left of each figure and in-going on the right. The filter used for the images on the left has a space constant of 22.6 arcmin and is shown at the bottom. On the right the filter has a space constant of 45.3 arcmin. It can be seen that as the out-going fins are lengthened the zero-crossings move out but this does not happen to the same extent for in-going fins. The peaks behave similarly to zero-crossings.

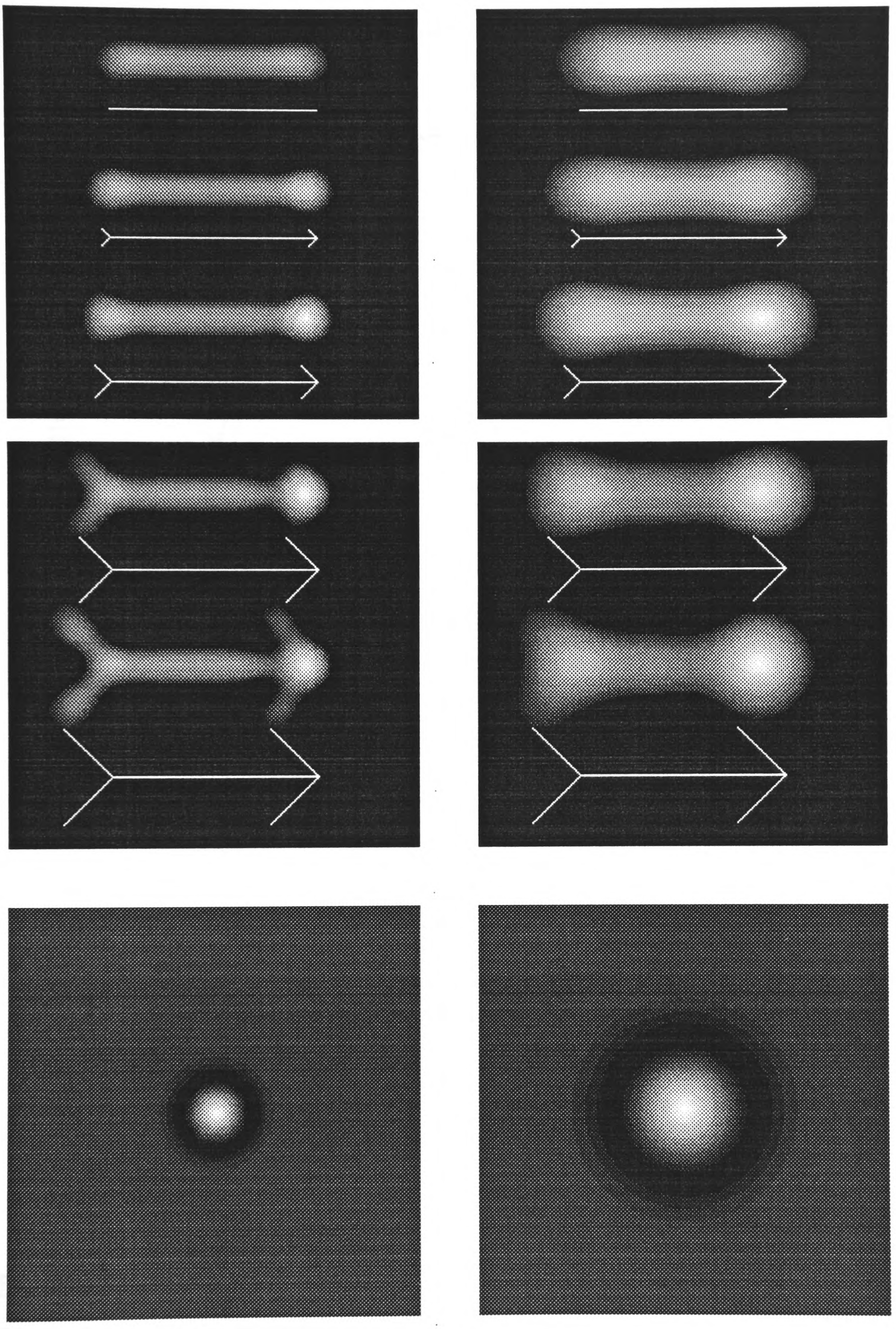


Fig 5.11 (legend on previous page)

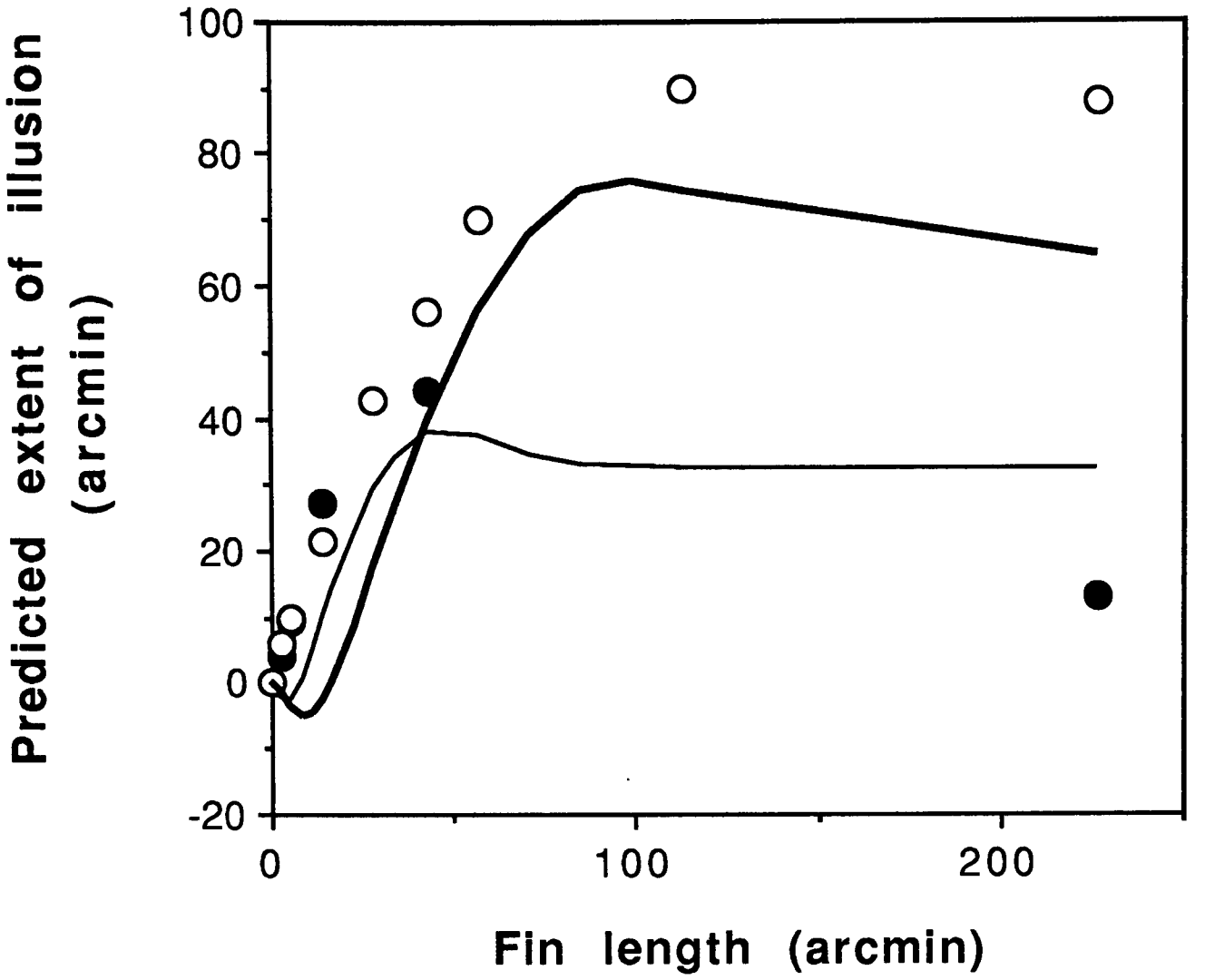


Fig 5.12(a)

The predicted extent of the illusion, i.e. the difference between the length of the "fins-in" and the "fins-out" figure, where the length of each figure is calculated from the separation of the peaks in the filtered output (see figure 5.11). The two filters used are the same as those illustrated in figure 5.11. Data from experiment I is shown for one subject for comparison.

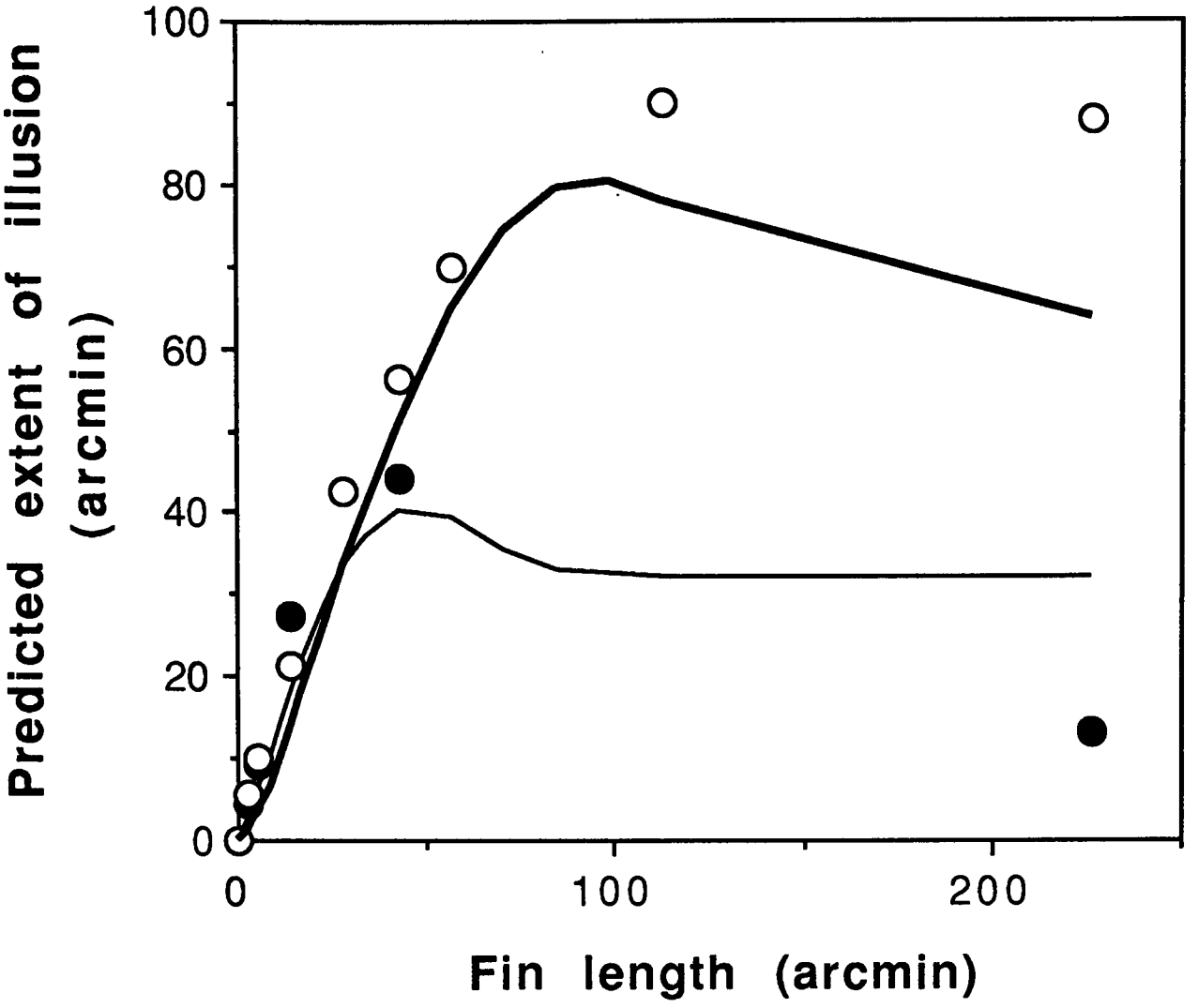


Fig 5.12 (b)

As for (a) but using zero-crossing separation (along the axis of the shaft) to define the length of the figure in the filtered output.

other the perceived slant was consistent with the fins *increasing* the apparent length of the shaft. This is consistent with the fact that the effect of adding ingoing fins is to increase the separation of the peaks along the shaft. (This also accounts for the small dip at short fin lengths in the peaks model shown in figure. 5.12a.)

Figure 5.12 shows the predicted extent of the illusion for Laplacian of Gaussian filters with space constants of 22.6 and 45.3 arcmin and data for one subject for comparison. The smaller of these filters provides a reasonable fit to the 3-D data, the 2-D data would be best fitted by a filter with a space constant slightly larger than the 45.3 arcmin filter illustrated.

5.5.2 A filtering model for the cyclopean illusion

The length of the cyclopean figures *also* depends on the scale of analysis. Of course, there is no length in either eye's image which corresponds to the length of the shaft: the Müller-Lyer figures do not exist in either monocular image. However, as figure 5.13 illustrates, if left and right eyes' images are filtered at a given scale then the outputs can be compared (or, as here, subtracted) to discover the regions over which there are non-zero disparities at this scale. As figure 5.13 shows, this region is longer in the case of the fins-out figure than it is for the fins-in figure. Figure 5.14 plots the difference in length of this region for the fins-in and fins-out figures for a range of fin lengths. (The length of each figure was defined as the length along the axis of the shaft over which the absolute difference between left and right eyes' images, filtered at a given scale, exceeded some particular threshold. A fixed threshold was used which was about 1% of the peak to trough range of the difference image. In fact, the output of the model is very little affected by the value chosen for the threshold, over a considerable range.). The model fits the data reasonably well, (in this case the data are best fitted by a filter with a space constant of about 45 arcmin).

So, a filter of a similar size (about 45 arcmin) can be used to model the data both for the 2-D illusion and the cyclopean illusion. This is different from assuming, as Julesz has (1971), that there is a Müller-Lyer "box" or mechanism whose input can be either luminance or disparity defined. Rather, the models illustrated in figures 5.12 and 5.14 demonstrate that for both the cyclopean and the 2-D figure, the shaft's length varies according to the scale at which it is measured. The choice of

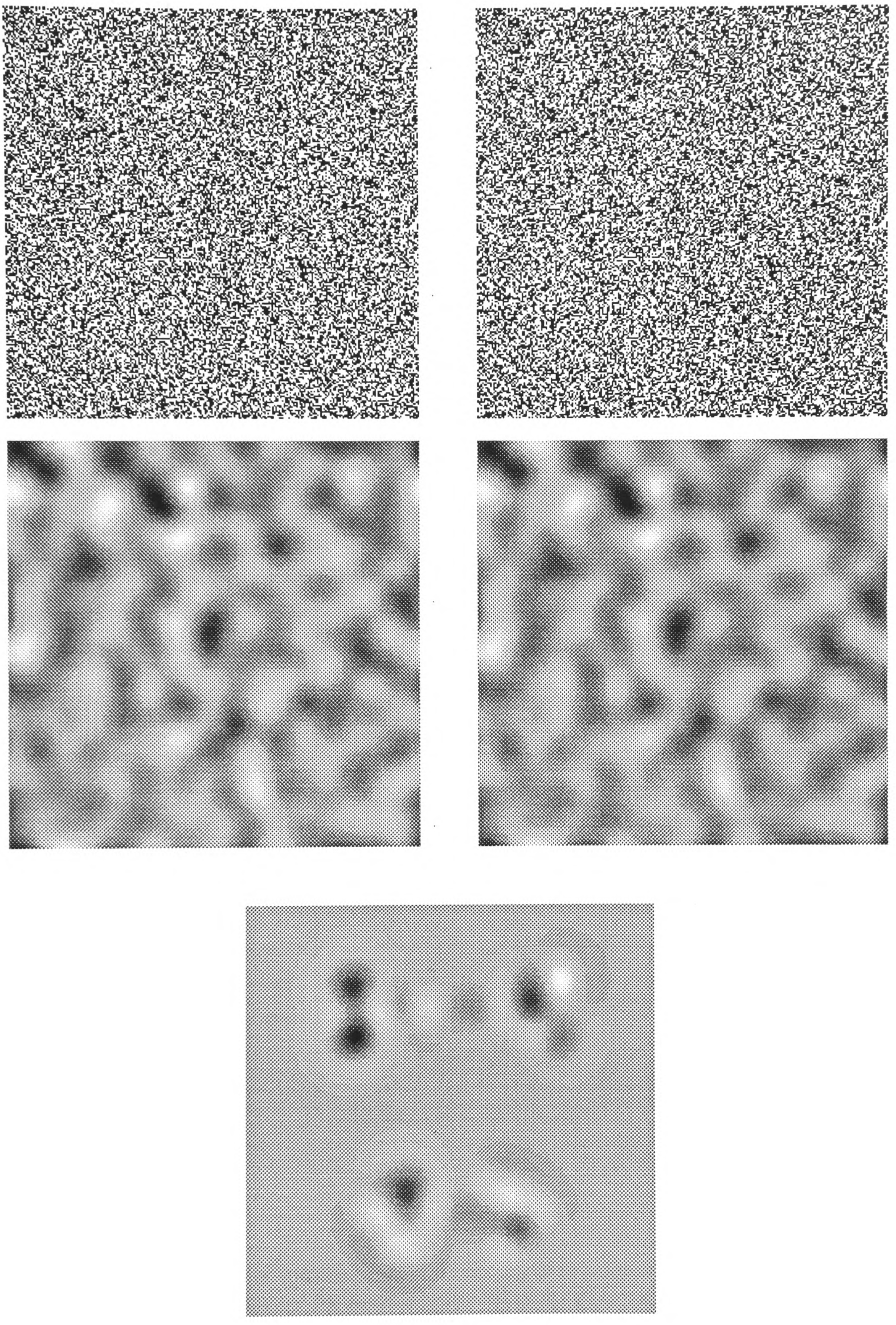


Fig 5.13 (legend on previous page)

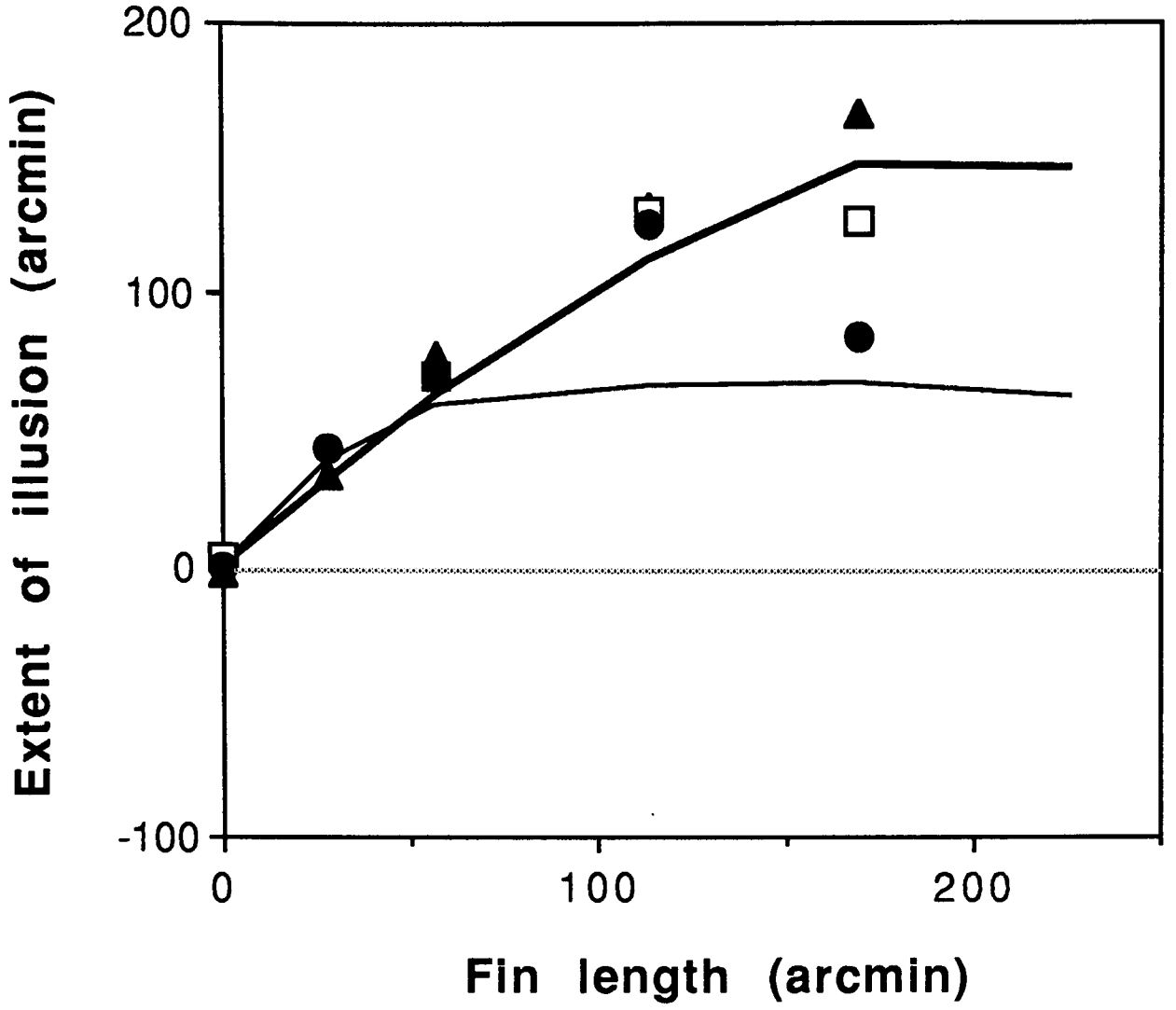


Fig 5.14

The predicted extent of the cyclopean illusion, i.e. the difference in length of the "fins-in" and "fins-out" figures as illustrated in figure 5.13, for a range of fin lengths. (The length of a figure was defined as the distance along the axis of the shaft for which the difference between left and right eyes' images is greater than an arbitrary threshold value). As in figure 5.12, the filters used here have space constants of 45.3 and 22.6 arcmin. Data from experiment II, for three subjects, are replotted here for comparison.

scale at which the length of the fins-in and fins-out figures are compared could depend on several factors, one of which might be their spatial (or temporal) separation (the further apart the larger the scale). This might help explain why a coarser filter is needed to model the 2-D and cyclopean data than for the 3-D data.

The size of the filters used in the modelling here are very large compared to the receptive field sizes of ganglion cells in the retina, neurons in the lateral geniculate nucleus or the psychophysical estimates of the sizes of channels in the fovea (e.g. Wilson and Bergen (1979). It should be pointed out that the model put forward here is a description at the algorithmic level (Marr, 1982) rather than a description of a suggested physiological implementation. Nevertheless, disparity sensitive neurons in the visual cortex with very large receptive fields have been reported, (e.g. Ohzawa, DeAngelis and Freeman (1990) fig 3 illustrates the recordings from cells which respond to stimulation over an area of up to 6°)

The model put forward here is not precise about the scale at which lengths or separations might be compared (across space, time or between the left and right eyes' images). In this respect it differs from Marr and Poggio's model which attempted to specify the size and range of filters at each eccentricity. Instead it is suggested that the scale at which lengths are compared may be, in general, larger for comparisons across larger distances or longer times and for longer lengths. A prediction of this model is that if shorter Müller-Lyer figures were used in experiment I and II, the results would be better fitted by a smaller filter. Of course, unless suitably controlled, a shorter Müller-Lyer figure would fall on less eccentric parts of the retina and be expected on these grounds to be processed through smaller filters. In an analogous argument about whether the Weber relationship for line length discrimination can be accounted for in terms of the eccentricity of the end points, (Levi and Klein, 1989; Morgan and Watt, 1989). Morgan and Watt demonstrated that at a single eccentricity Weber's law still applies. It may be possible to design a similar control experiment using a form of the Müller-Lyer illusion to disentangle these two effects.

5.6 Discussion

The novel demonstration of a "3-D" Müller-Lyer illusion described in this chapter fits well with a system in which length is encoded hierarchically. (The 2-D illusion was discussed in terms of a hierarchical model in section 5.1). The hypothesis is

that each eye's image is encoded hierarchically before being compared for stereopsis. If this is the case then comparisons of the lengths of Müller-Lyer figures separated in space or time (the 2-D illusion) would be expected to show similar characteristics to comparisons made between the images in the two eyes (the 3-D illusion). This is borne out by the results of experiment 1.

Can an alternative explanation of the 3-D Müller-Lyer illusion be given in terms of *retinal* disparity? The ends of the shafts have zero disparity and therefore, at least in any simple scheme, the shaft should be seen as flat, in the fronto-parallel plane. How might the effect of the fins be explained without resorting to a filtering hypothesis?

5.6.1 Are the fins matched instead of the shaft ends?

One possibility is that the overall length of the (unfiltered) figures is matched rather than the shaft ends (i.e. the "point" of the ingoing-fins figure in one eye might be aligned with the tips of the outgoing fins in the other eye's image) but this predicts a much smaller magnitude of the illusion than was found. A second possibility is that the tips of the ingoing fins are matched with the tips of the outgoing fins. In fact, this fits quite well with the observed magnitude of the illusion. (It predicts a slope of $\sqrt{2}$, i.e. 1.41, when the extent of the illusion is plotted as a function of fin length. The slopes of the data in figure 5.8 have a mean of 1.46 (s.e. = 0.07)).

If this were the case, however, a slant should not be perceived when fins from one figure are omitted. Table 5.1 shows results for this situation, and it can be seen that the perception of slant remains. In most cases the slant perceived for a stimulus with only outgoing fins is much greater than with only ingoing fins. As discussed in the previous section, this fits with the position of the peaks and zero-crossings in the coarse-filtered images.

5.6.2 Disparity interactions

A third possibility is that the shaft ends are matched, and their disparity recorded as zero, but that a further stage of processing occurs before this is interpreted to give their depth. The two main types of disparity interaction which have been proposed are depth averaging and depth contrast. In most models (e.g. Westheimer, 1986; Lehky and Sejnowski, 1990; Mitchison, 1992) depth averaging is assumed to take place over very short ranges (up to two or three arcmin) and outside this region,

Subject	Fin length	Fins-in only bias	Fins-out only bias	Total	Fins-in and fins-out (Expt I)
RAE	5.6	-2.10	9.48	7.38	8.6
	14.1	1.82	12.60	14.42	28.6
BL	5.6	-1.40	7.94	6.54	9.2
	14.1	2.88	12.12	15.00	27.4
MFB	5.6	-1.52	5.26	3.74	8.6
	14.1	7.04	1.90	8.94	25.2
BJR	5.6	0.30	6.38	6.03	8.8
	14.1	1.80	5.74	7.54	25.3

Table 5.1

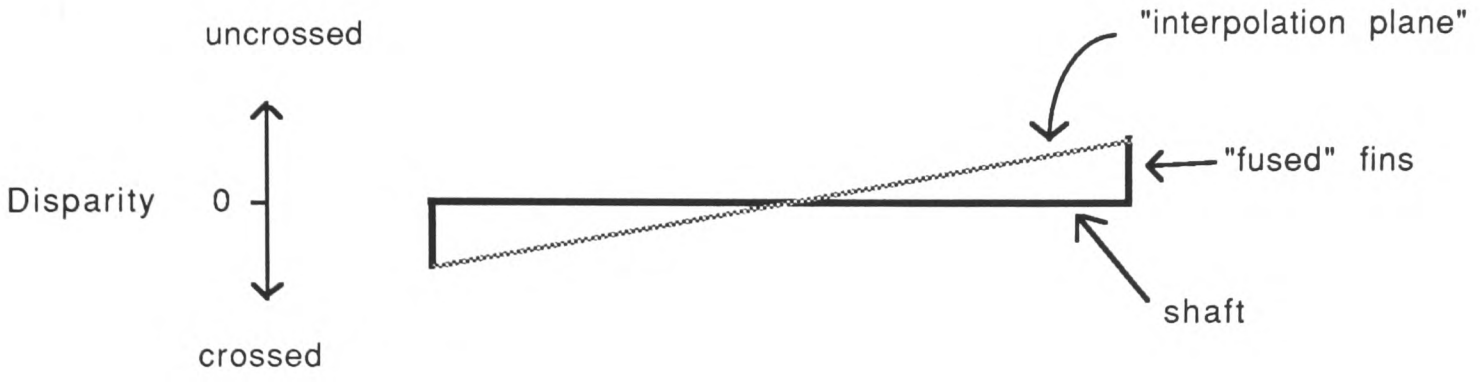
Results for four subjects for conditions in which fins were shown only to one eye (the shaft alone was presented to the other eye). All figures are in arcmin. Two fin lengths were used, 5.6 and 14.1 arcmin. The extent of the illusion (or bias) for the "fins-in" and "fins-out" conditions is shown in the third and fourth columns. (A negative bias means that the "fins-in" figure was shorter than the "fins-out" figure when the stimulus was seen as flat (fronto-parallel). The labels "fins-in" and "fins-out" have been used even when the fin length is zero, i.e the shaft only is shown.) The fifth column shows the sum of the values in the third and fourth columns, i.e. for "fins-in only" and "fins-out only", and the last column shows, for comparison, the extent of the illusion when fins were shown to both eyes.

over a slightly wider field, depth contrast is assumed to dominate. Another type of interaction, interpolation (Mitchison and McKee, 1987, discussed in section 5.1) is, in fact, the most relevant to the 3-D Müller-Lyer demonstration. (Confusingly, disparity averaging is sometimes called interpolation. The interpolation Mitchison and McKee propose is very different.)

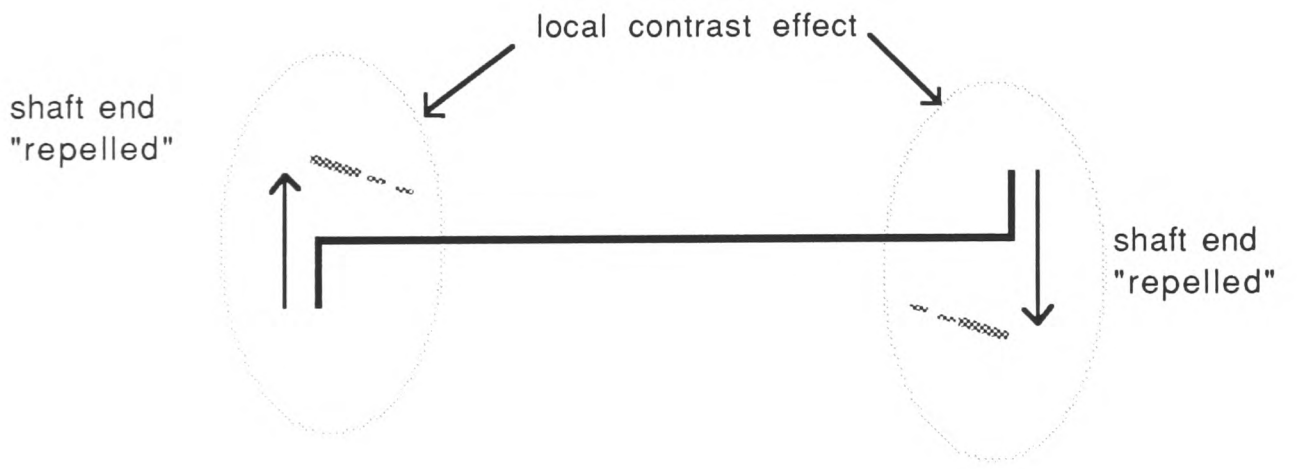
Most subjects, when viewing the Müller-Lyer figures, see the fins as diplopic. If they were fused they would be seen as coming straight out of the screen at one end of the shaft and as going straight into the screen at the other. Even so, they might be expected to give rise to a disparity signal which could interact with the disparity of the shaft ends. Figure 5.15 illustrates the disparities which would be generated by the 3-D Müller-Lyer figures. Also shown in this figure is the slant of the shaft measured in experiment 1 (i.e for fin lengths up to 30 arcmin, when, as discussed above, the perceived slant approximately equalled that of a plane drawn through the *tips* of the fins). It can be seen that the direction of perceived slant of the shaft could be explained in terms of disparity averaging, although it would be difficult to explain the magnitude of the perceived slant with such a model.

Another reason for rejecting depth averaging as an explanation of the slant is that its range is supposed to be short (much shorter than the 400 arcmin shaft length) and certainly shorter than the range of depth contrast e.g. Westheimer, 1986; Lehky and Sejnowski, 1990). But it has been pointed out (Mitchison, personal communication) that evidence of depth contrast can be observed in the 3-D Müller-Lyer stimulus (illustrated in figure 5.15 (b)). When fixating on just one end of the shaft and judging its depth relative to the screen (or fixation marker), for some subjects at least, the shaft end appears to have the opposite depth to that expected from the slant judgement (i.e. if a subject saw the slant of the shaft as "left end towards me" they might nevertheless judge the left shaft end to be farther away than the right end when both were scrutinised carefully). This can be interpreted as a depth contrast effect induced by the disparity of the fins. It operates only locally, in the region of the shaft ends, and can co-exist, paradoxically, with a perception of the whole shaft as slanted in the opposite direction.

Thus, whatever process is giving rise to the perceived slant of the shaft appeared to be a larger scale effect than the process giving rise to the depth contrast. Disparity averaging, on the other hand, should be a smaller scale effect. Mitchison and



(a)



(b)

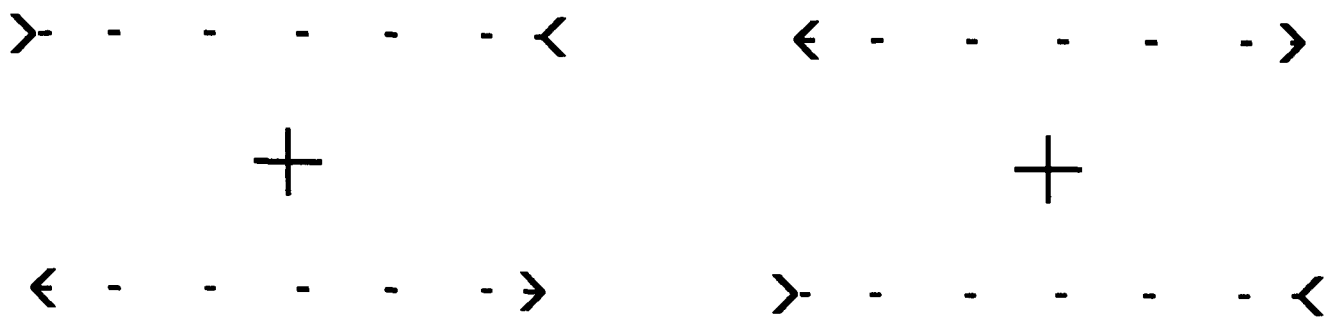
Fig 5.15

An illustration of possible disparity interactions. Above (a), the shaft and fins are shown (bold) in disparity space (as if in plan view). The shaft is fronto-parallel (shown as horizontal); the fins, if fused, would point straight towards or straight away from the observer. If an "interpolation plane" were drawn (in disparity space) between the tips of the fins then it would slant in depth as shown (dotted line). The magnitude and direction of the slant would fit quite well with the observed results (experiment I). A disparity averaging hypothesis would predict, if anything, a smaller slant.
Below (b), a simultaneous contrast effect is illustrated. The fins, if fused, might be expected to cause the shaft ends to be perceived at a depth opposite to that of the fin tips.

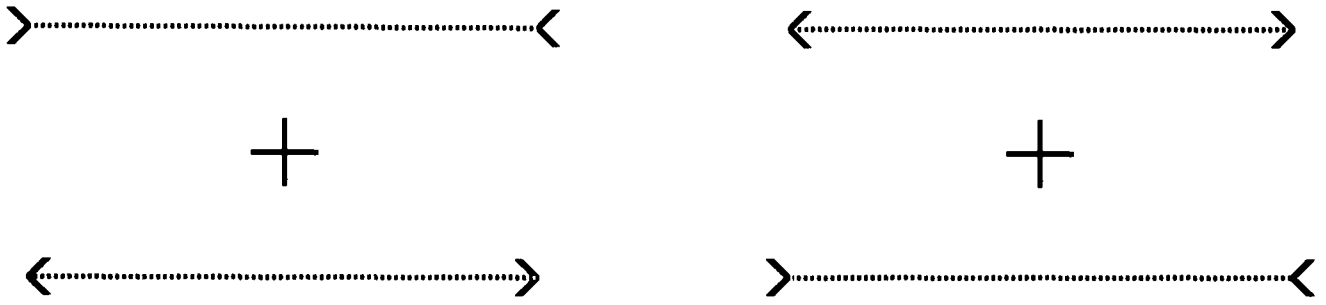
McKee's (1987) "interpolation", however, can take place across relatively large areas and has been applied to stimuli very similar to the 3-D Müller-Lyer illusion. Their experiment and theory have been described in section 5.1. To recap, Mitchison and McKee (1987) propose that the first step in matching left and right eye's images is that features with unambiguous matches in the other eye's image are matched and their disparity calculated. Then, an "interpolation plane" is drawn through these points (in disparity space) and this guides the search for any unresolved (ambiguous) matches, the rule being to chose the match with the disparity closest to the interpolation plane.

One stimulus they used (Mitchison and McKee, 1987, their figure 8) was very similar to the 3-D Müller-Lyer illusion. A grid of regularly spaced dots was shown to each eye. The centre dots all had matches in the fixation plane but the perception of these was strongly influenced by the disparity of the dots at the end of the rows. If these dots had unequal disparities (i.e. the plane joining them was slanted in depth) then, for brief presentations and close dot spacing, the central dots appeared to lie along this "interpolation plane". For longer exposures (2-5 s), the central dots appeared to lie in one of two fronto-parallel planes, consistent with the discrete matching of each dot. Figure 5.16 shows a version of the 3-D Müller-Lyer figure in which the shaft is drawn as a dotted line. For widely spaced dots and long exposures, the illusion of slant is almost abolished, as in Mitchison and McKee's experiment. For closely spaced dots, consistent with Mitchison and McKee's results, most subjects find that the perception of slant remains. For an unbroken shaft, as used in experiment I, the slant is largest and most stable of all.

Despite the close analogy, there are a few problems in applying Mitchison and McKee's model to the results of the present experiment. Mitchison and McKee specified that an interpolation plane was drawn through points which could be matched unambiguously. The most obvious unambiguous points for determining the slant of the shaft are the shaft ends and this would not predict any perceived slant at all. Only if the tips of the fins, or the coarse filtered primitives described in section 5.5.2, were used would the correct slant be predicted. Also, Mitchison and McKee's model provides no explanation for the effect of monocular fins (see Table 5.1). The same criticism applies to any explanation which seeks to describe the effect in terms of interactions in the disparity domain.



(a)



(b)

Fig 5.16

The 3-D Muller-Lyer demonstration redrawn with the shaft as a dotted line. Above (a), the dot separation is large and the dots appear flat, in the fronto-parallel plane. Below (b), the dot separation is small and the illusion of slant remains.

In summary, models of disparity interactions can explain some aspects of the perception of the 3-D Müller-Lyer figure, but not all. One of the most interesting observations about this stimulus is the fact that one feature, e.g. the shaft end, can be seen to have two depths at once depending on where the subject is looking and the judgement they are trying to make. This sort of phenomenon is more easily explained within the framework of a hierarchical model than one based on disparity interactions. (The interpretation of simultaneous contrast effects in terms of a hierarchical model is discussed in section 7.2.)

5.6.3 Balanced dots

The explanation for the Müller-Lyer illusion (2 and 3-D versions) which has been put forward in this chapter is based on differences in length of the coarse filtered versions of the figures. This part of the argument is old (e.g. Ginsburg, 1978; Cornell, 1978) and has been countered in the past by experiments which show that the illusion remains when low spatial frequencies have been removed from the stimulus. For example, Carlson, Moeller and Anderson (1984) made up a Müller-Lyer stimulus from "balanced dots" (i.e. dots containing almost no low spatial frequencies) and showed that the extent of the illusion was not significantly reduced when compared to a stimulus made up of white dots.

Several counter-arguments have been raised. García-Pérez (1991) has argued that although one balanced dot contains almost no low spatial frequency information the same cannot be said for lines of balanced dots. He shows a low-pass version of Carlson et al..'s (1984) stimulus, which does indeed have some energy, but the zero-crossing in this coarse filtered version is very close to the shaft end both for a fins-in and fins-out stimulus and could not explain the persistence of the illusion. The argument that the signal could be rectified (either full-wave, e.g. Chubb and Sperling, 1988; or half-wave, e.g Watt and Morgan, 1985) falls foul of the same problem, unless it is proposed that a *second* stage of filtering occurs after the rectification. A more promising, but less quantitative, model was put forward by Morgan et al.. (1990). They suggested that "texture detectors", with large, widely spaced receptive fields (i.e. coarse scale) were responsible for the illusion and, because they gathered signals from a range of cells of different types and scales, were equally well stimulated by balanced and unbalanced dots.

The problem with using balanced dots, or any high-pass stimulus, to investigate the visual system has been discussed in section 3.3. High-pass filtering does not eliminate the coarse structure of the stimulus and there are many ways, as

discussed above, in which a non-linear processing stage can recover the coarse signal. In fact, Carlson et al.. (1984) showed that the extent of the illusion was sharply reduced by lowering contrast for the balanced dots but not for the white dots, implying that non-linearities at some stage in the visual system had an important part to play.

5.6.4 Historical precedents

Many experiments have described the effect of showing the two "halves" of an illusion to the two eyes (e.g. Springbett, 1961; Ohwaki, 1960) but mostly these have been concerned with the size of the *2-D* illusion (which is typically reduced). Only Lau, in 1922, attempted a similar demonstration to that described here. He used the Zöllner illusion and claimed that observers perceived the long parallel lines as slanted out of the page alternately in opposite directions. However, neither Ogle (1962) nor Julesz (1971) were able to reproduce the effect and all attempts at reproducing the result in this laboratory have failed. Even Lau himself did not get consistent results with the most straight forward presentations, i.e. either showing parallel lines to one eye and parallel lines with cross hatches to the other or showing parallel lines to both eyes with the cross hatches drawn in opposite directions in the two eyes. (It is not clear exactly what the stimulus configuration was which brought him the "gewünschten Erfolg" (desired result, p3) in his third experiment. The long parallel lines were at 45° and the cross hatches either vertical or horizontal, except that in one eye's view the cross hatches were "at a slightly sharper angle" (p3). It is unclear from Lau's description whether this meant that there was a slight orientation disparity between the cross hatches in two eyes and whether this was perceived by subjects. In all, though the idea was good, his results are not. Julesz (1971) comments:

> *"It is really too bad that [Lau's experiment] did not yield convincing results. Had Lau succeeded in his effort, I think cyclopean psychology would have been developed right then."*
>
> (Julesz, 1971, p236)

No satisfactory explanation of the Zöllner illusion in terms of coarse filtering has been put forward (indeed if one steps back from a Zöllner figure until the cross hatches begin to blur together, the illusion disappears) and it may be that this is the reason for Lau's failure.

5.7 Summary

In this chapter the purpose of a coarse-to-fine analysis, in organising information about the image, has been discussed. The main idea is that the position of each blob is recorded first, relative to the centroid of the coarse scale ("parent") blob and second, relative to other blobs at the same scale (within that coarse scale blob). These measurements are "explicit". Using these explicit measurements it is theoretically possible to calculate the relative position of any two features in the image, even though each measurement is local, because the information is organised hierarchically.

However, the hypothesis explored in this chapter was that for the purposes of a large scale judgement the exact position of two points is not calculated but only the separation of the coarse scale *group* of features containing each point. This does not mean that the visual system can only make coarse scale judgements but rather that it may be poor at combining information recorded at several scales (or "levels" in the hierarchy).

The Müller-Lyer illusion was described in terms of a hierarchical encoding of position as an example of a situation in which using the coarse scale grouping to judge the separation of two points would result in a detectable bias. For this reason, the illusion was used as a tool to investigate whether hierarchical encoding of each eye's image precedes stereoscopic matching. A "3-D Müller-Lyer illusion" was described in which, when the fins-in figure of the illusion was shown to one eye and the fins-out figure to the other eye, the binocularly fused shaft was perceived as slanted in depth (section 5.2). The results from experiment I showed that, for fin lengths up to about half a degree, the extent of the 2-D Müller-Lyer illusion (apparent length difference) and the 3-D Müller-Lyer illusion (apparent slant) were very similar. This evidence supports a hierarchical model in which comparisons of length between the two eyes' images (slant judgements) show similar properties to comparisons of length across space.

A hierarchical model may also account for the cyclopean (random dot) version of the Müller-Lyer illusion since length differences between the fins-in and fins-out figures exist when the two eyes' images are compared at a coarse scale. A similar filter size was shown to be appropriate for modelling the data on the cyclopean illusion (experiment II) as was used to model the classical 2-D illusion.

The Müller-Lyer illusion, particularly the version described by Morgan et al.. (1990), is an example of an image with a well-defined hierarchical structure. The output of fine filters lies within the boundaries of coarse scale filter outputs and this leads to a simple hierarchical organisation of the image. The perception of stimuli with a much less well-ordered hierarchical structure is explored in the next chapter.

CHAPTER 6

In chapter 2 the MIRAGE algorithm for combining filter outputs at a range of scales was examined in detail. In chapter 3 several examples of images and the MIRAGE representation of each were illustrated including those of a high and low density random dot pattern. The form of the representation, in particular the spacing of blobs, was quite different for different dot densities. The experiment described in this chapter investigates the detection of large disparities in random dot patterns covering a wide range of densities. The results are relevant to a discussion of whether some combination of filter outputs, such as that proposed in the MIRAGE algorithm, precedes stereoscopic matching.

In the next two sections some experiments are reviewed that have sought to determine the upper disparity limits of stereopsis, and the upper displacement limit in two-frame apparent motion sequences.

6.1 Detecting a large disparity

6.1.1 Low density patterns

The criteria for measuring the maximum detectable disparity are not well established. Problems with some candidate measures such as diplopia have been discussed earlier (section 1.5). Objective measures, using a two alternative forced choice procedure, have been used to determine the upper limit of stereopsis in experiments using simple line or spot stimuli. Westheimer and Tanzman (1956) used a brief exposure duration (10 ms flash) and a randomised procedure. They tested six subjects, the best of whom could reliably discriminate the depth of a spot as in front or behind the fixation plane (i.e. errors of 10% or less) for convergent or divergent disparities of up to 10°. On average, subjects could discriminate the depth of the spot with errors of less than 20% for disparities of up to about 4°. A criticism of this experiment is that subjects could theoretically do the task with one eye closed, although the fact that errors tended to increase with disparity suggests they did not use this strategy. Blakemore (1970b) carried out a similar experiment with line targets (which were given a random lateral displacement). He obtained comparable results for central vision and found, for one subject, a disparity range of about ±14° when the tests were repeated at eccentricities of 5 and 10°.

6.1.2 Filtered random dot patterns patterns

Mowforth, Mayhew and Frisby (1981) also used an objective method to test the upper disparity limit in filtered random dot stereograms. Figure 6.4 (a) shows, in schematic form, how the stimuli were made (originally described by Julesz, 1960). One frame of dots was created in which each pixel within a square boundary had a 50% probability of being plotted black or white. The second frame was created by shifting all the dots in the first frame by a given amount. (In these experiments the shift was relatively large, e.g. 1° for a patch size of 5 by 5°.) All those dots that fell outside the square boundary were not plotted and new random dots filled the "gap" on the other side of the square. Thus, there was a strip of uncorrelated dots in each frame. The two frames were then filtered using a narrow band isotropic filter. Mowforth, Mayhew and Frisby found that the disparity that could be matched in filtered random dot patterns (i.e. initiate a vergence movement in the correct direction) was smaller for high than for low frequency patterns. The largest disparity they tested was 56 arc min. Even for the low frequency stimuli, the upper limits of disparity that could initiate an appropriate vergence eye movement were small compared to those used in the case of line stimuli. They were also small compared to the disparities at which Fender and Julesz (1967) found that depth

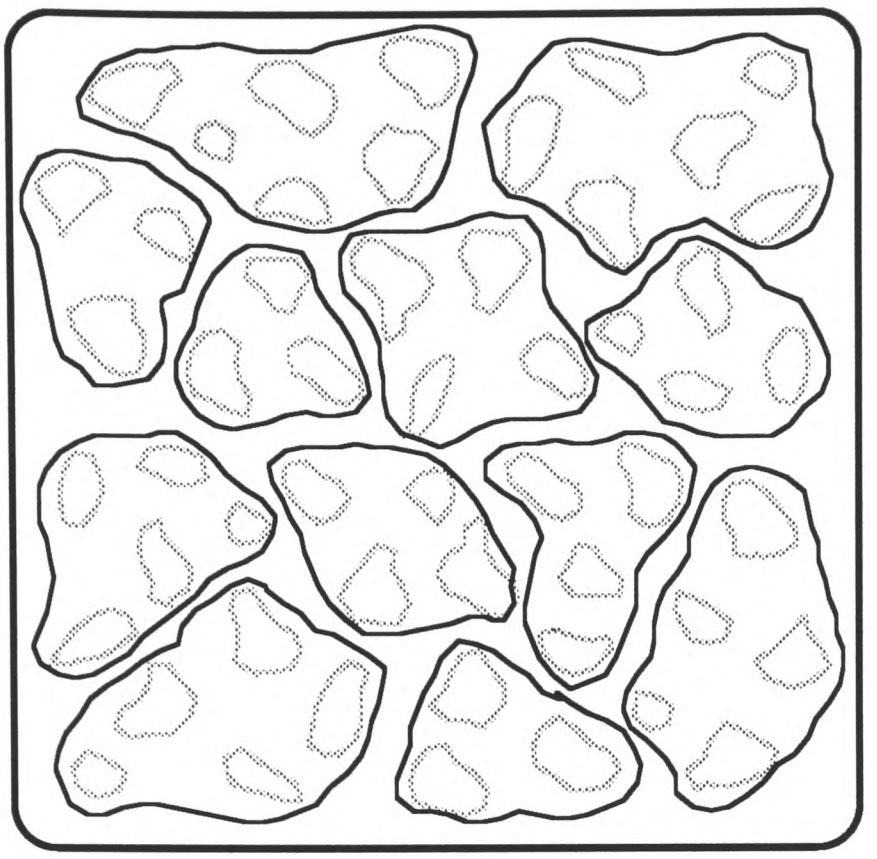
could be perceived in random dot stereograms for stabilised images. There is an important distinction between the manipulation carried out by Fender and Julesz and that used by Mowforth, Mayhew and Frisby. In one case the whole random dot pattern was given a disparity, in the other the disparity of the frame remained constant while the dots were shifted (and new, uncorrelated dots replaced them). Blakemore (1970), in discussing Fender and Julesz's results brings out this point:

> *"The pattern contained a central square of dots that was shifted into 8 min of convergent disparity, relative to the background. It was necessary to introduce 2 deg of absolute disparity before the sensation of depth was lost and the pattern seemed to fragment. Rather than a determination of Panum's fusional area under conditions of stabilization, as suggested by the authors, this experiment may perhaps be better considered as an estimate of the relative disparity threshold (8 min) at an absolute divergent disparity of 2 deg."*
>
> (Blakemore (1970) p608)

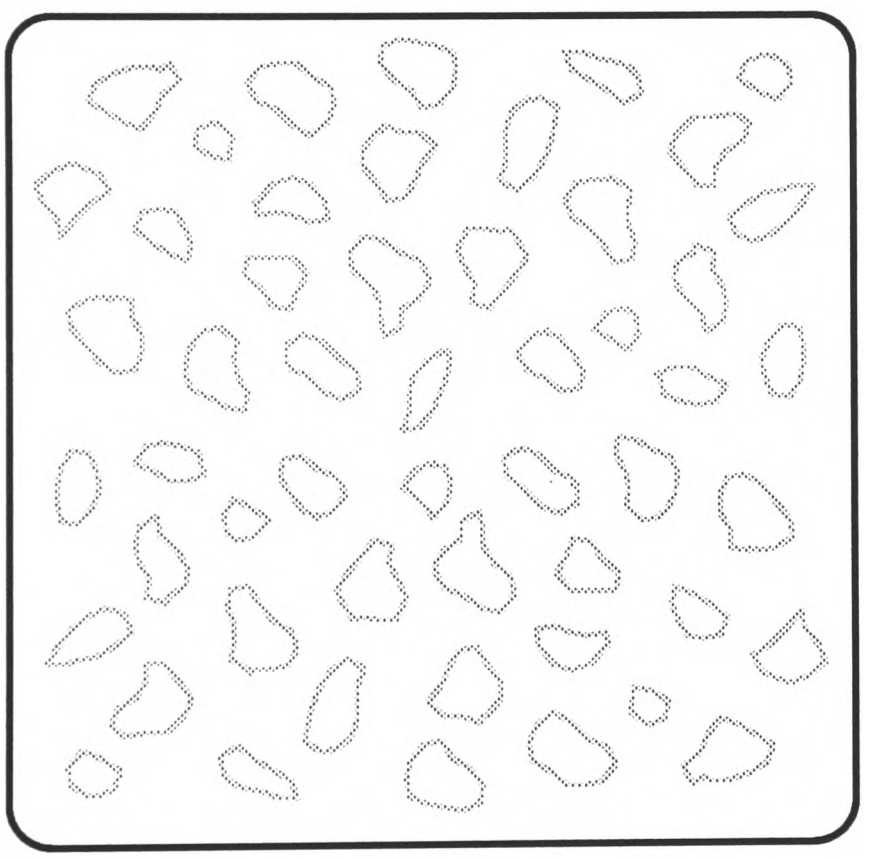
According to this terminology, Mowforth, Mayhew and Frisby (1981) introduced a relative disparity of up to 56 arcmin while the absolute disparity of the pattern (i.e. the outline) remained at zero.

An interpretation of these experiments within a hierarchical model would be very similar. The type of stimulus used by Mowforth, Mayhew and Frisby (for instance, a high frequency band-pass stimulus with a large disparity) "re-introduces" a correspondence problem that, in a hierarchical system, is rarely encountered. At a coarse scale the whole patch is matched in the fixation plane. As the scale of analysis is reduced a large number of fine scale blobs "appear". (The same would apply to both Marr and Poggio's scheme and a hierarchical model.) Usually a much smaller number of blobs at an intermediate scale would be discovered and these would "group together" the fine scale blobs. In the absence of any medium scale grouping, the positions of the fine scale blobs must be defined relative to the coarse scale centroid. As a result, the "hierarchical" disparities of the fine scale blobs are the same as their retinal disparities and these can be large in relation to the spacing of the blobs. Figure 6.1 illustrates these points schematically.

As an aside, it is interesting to note that, if the whole patch were given a convergent disparity, but the fine scale elements within the patch remained in the fixation plane, then the "hierarchical" disparities of the elements (i.e. relative to the coarse blob) could be large while their retinal disparity was zero. Mitchison and



(a)



(b)

Fig 6.1

Shown in (a) is a coarse scale blob (square outline) containing several medium scale blobs each of which contain, in turn, several fine scale blobs (dotted lines). In the case shown below (b), there are no medium scale blobs and hence no hierarchical grouping structure (this represents a high frequency filtered pattern). Natural images (such as that illustrated in figure 1.4) have a 1/f spectrum and an image structure more like that shown in (a) than in (b) (illustrated in figure 3.6).

McKee (1987b) describe a stimulus along these lines and show that the matching of the fine scale elements tends to be forward of the fixation plane (i.e. follows the patch disparity) even in the absence of eye movements. They use this as evidence against a theory based on retinal disparities and in favour of a model equivalent to a hierarchical one. (The experiment is discussed in detail in section 5.1.)

The important point about these stimuli is that fine scale blobs can, unusually, have a large hierarchical disparity. It is possible that in this situation a very simple "nearest neighbour" rule for matching, which is normally adequate, may break down. It would break down when false matches were nearer neighbours than correct ones, i.e. the upper disparity limit would depend on the spacing of blobs.

Although some evidence has been gathered using large disparity random dot stereograms, both filtered and unfiltered (e.g Mowforth et al., 1981; Frisby and Mayhew, 1980), a more thorough investigation of the limitations of the correspondence has been carried out in the motion domain, using 2-frame random dot kinematograms. Some of these experiments are discussed in the next section.

6.2 Detecting a large displacement (in the motion domain)

The correspondence problem for a two-frame apparent motion sequence is formally equivalent to that for a pair of images viewed stereoscopically. Each feature in frame one (or the left eye's image) must be matched with the corresponding feature in frame two (or the right eye's image). It could be argued that this type of correspondence problem is rarely encountered by the motion system because in natural viewing conditions the image is sampled continuously over time. Nevertheless, quite large instantaneous displacements in a two-frame sequence give rise to a perception of motion which shows that in these circumstances the correspondence problem can be solved. In fact, the limits of the correspondence process have been explored more thoroughly in motion than they have in stereo.

A two-frame apparent kinematogram is the closest possible motion analogue of a random dot stereogram. The questions addressed by the experiments using these stimuli are all relevant to stereopsis: How is the maximum displacement that can be detected affected by the spacing of elements in the image? What is the effect of

spatial frequency, and can this be distinguished from element spacing? How does patch size affect the maximum detectable displacement?

6.2.1 The effect of dot density

Two recent experiments have investigated the effect of dot density on the perception of motion at large displacements using two-frame kinematograms (Baker and Braddick, 1982, and Eagle and Rogers, 1991). Subjects were shown two frames of random dots as an apparent motion sequence (100 ms exposure for each frame, no inter-stimulus interval) and had to judge the direction of motion as "up" or "down". As described above for a random dot stereogram, the dots were shifted so that some disappeared from the frame and others moved into the frame as if the surface of dots were moving behind a stationary window. For large displacements coherent motion is no longer be perceived. The displacement at which 20% errors in direction discrimination were made was defined as "d_{max}". (Braddick (1974) introduced the concept of an upper displacement limit that was determined by errors in a two-alternative direction discrimination task. Baker and Braddick (1982), and many authors since, have used the 20% error point as a definition of d_{max}.). The same paradigm was used by both Baker and Braddick (1982) and Eagle and Rogers (1991).

Baker and Braddick (1982) claimed that d_{max} for motion was independent of dot density. As figure 6.2 shows, this is a legitimate conclusion given their results (shown as a dotted line). However, Eagle and Rogers (1991) used a whole range of patch sizes and their results (also shown in figure 6.2) lead to a very different general conclusion.

When the patch size was sufficiently large (e.g. 25 by 25°), d_{max} rose as dot density was reduced, reaching a value of 6 or 7 degrees for the lowest density patterns, i.e. displacements that Braddick suggested were typical of "long range" motion (Braddick, 1974). In fact, rather than postulating a separate long and a short range "mechanism", Eagle and Rogers' results suggest it may be more appropriate to consider the separate factors that limit the correspondence process for motion, e.g. patch size and dot density.

The effect of patch size appears to be to set a "ceiling" on d_{max} of about 1/5 of patch height (i.e. in the direction of motion - other experiments on rectangular patches

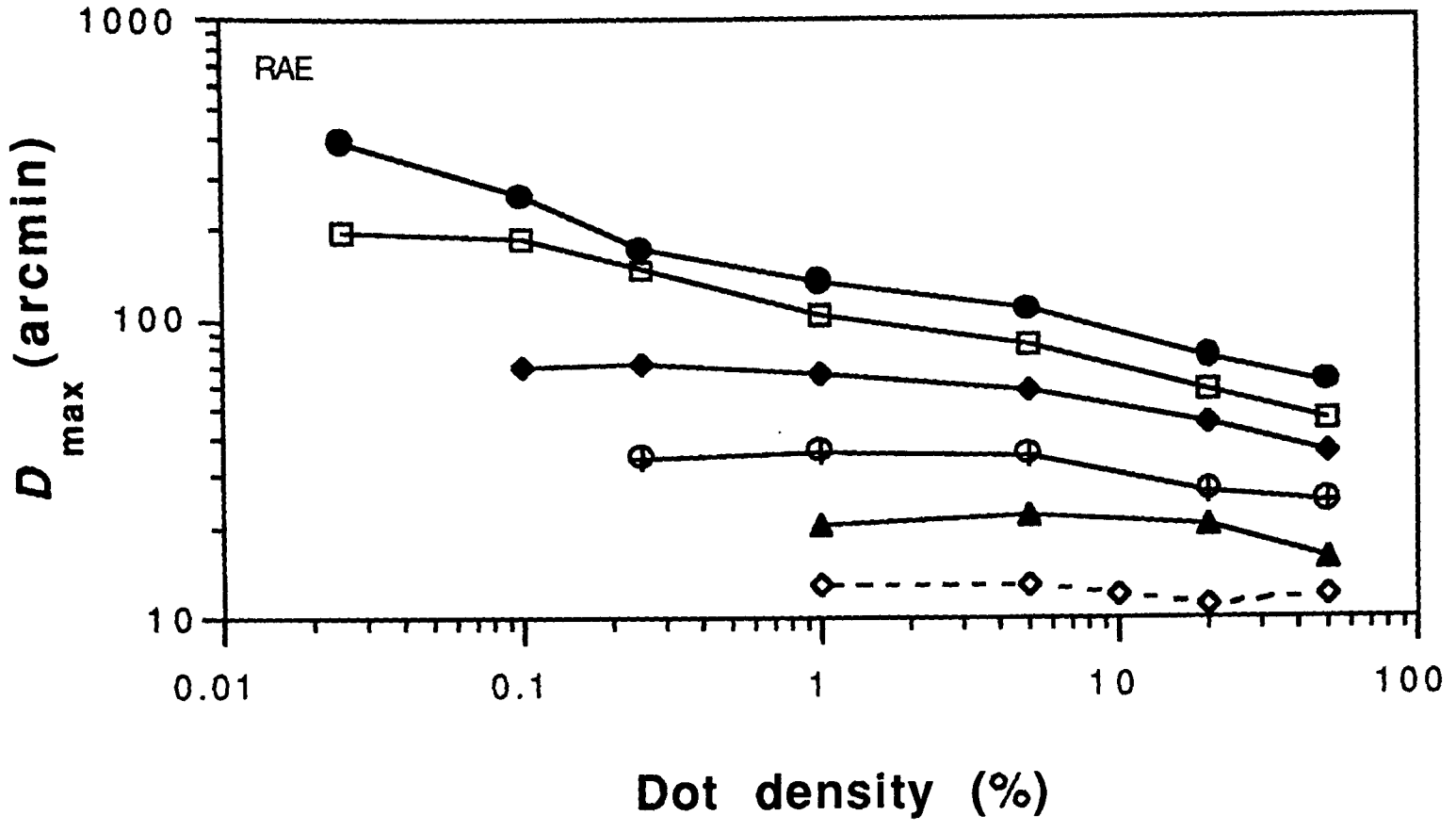


Fig 6.2

Results from Eagle and Rogers (1991). Dmax for a two-frame apparent motion task is plotted against dot density for a range of patch sizes. (The width of the patch, which was square, is shown in the key below.) Results from a similar experiment by Baker and Braddick (1982) are shown for comparison. (They used a rectangular patch whose dimensions are given in the key.) The slope of the best fitting power function for the 25.4° patch size data is -0.23.

confirm this (Eagle, 1992)). Provided patch size is sufficiently large, d_{max} is limited instead by dot density. The gradual rise in d_{max} as dot density is reduced suggests that the upper displacement limit for motion may reflect the spacing of false targets in the image rather than a fixed spatial limit, as Baker and Braddick (1982) supposed.

6.2.2 The effect of dot size

A similar conclusion has been reached by Morgan (1992) using slightly different stimuli. He showed that, for a two frame apparent motion task using 50% random dot patterns (patch size of 5° by 5 °), d_{max} increased in proportion to the size of the dots (provided they were larger than a 15-20 arcmin). This result is not new (e.g. Cavanagh, Boeglin and Favreau (1985) and Sato (1990) have reported similar findings) but his interpretation of the results was, i.e. that d_{max} reflects the spacing of primitives or false targets in the image.

Two effects occur when dot size is manipulated, only one of which is the change in primitive spacing. The second is that, as dot size is increased, the spectrum of the image changes (there is more energy at low and less at high spatial frequencies) and this might be expected to affect d_{max}. The same ambiguity occurs for studies of d_{max} that have used band-passed stimuli (Chang and Julesz, 1983; Cleary and Braddick, 1990a, Bischof and Di Lollo, 1990). The main finding of these experiments, that d_{max} changes inversely with the spatial frequency of the filter, has been interpreted by the authors as evidence for spatial frequency tuned mechanisms each with a different d_{max}, but the results fit equally well with a model based on primitive spacing.

6.2.3 D_{max} depends on density not spatial frequency

Eagle and Rogers (1992) sought to disentangle these two components (spatial frequency and primitive spacing). The stimuli used in their experiment were band-pass filtered versions of the variable dot density stimuli used in Eagle and Rogers' (1991) experiment. Figure 6.3 illustrates their results. In the graph at the top, d_{max} is plotted as a function of the density of dots in the stimulus before it had been filtered. Results are shown for three filter sizes. Clearly, the idea that d_{max} depends only on the spatial frequency content of the stimulus can be rejected - this would give three horizontal lines for the three filter sizes. In fact, the data fits very well with a model in which d_{max} depends only on the density of features in the image and

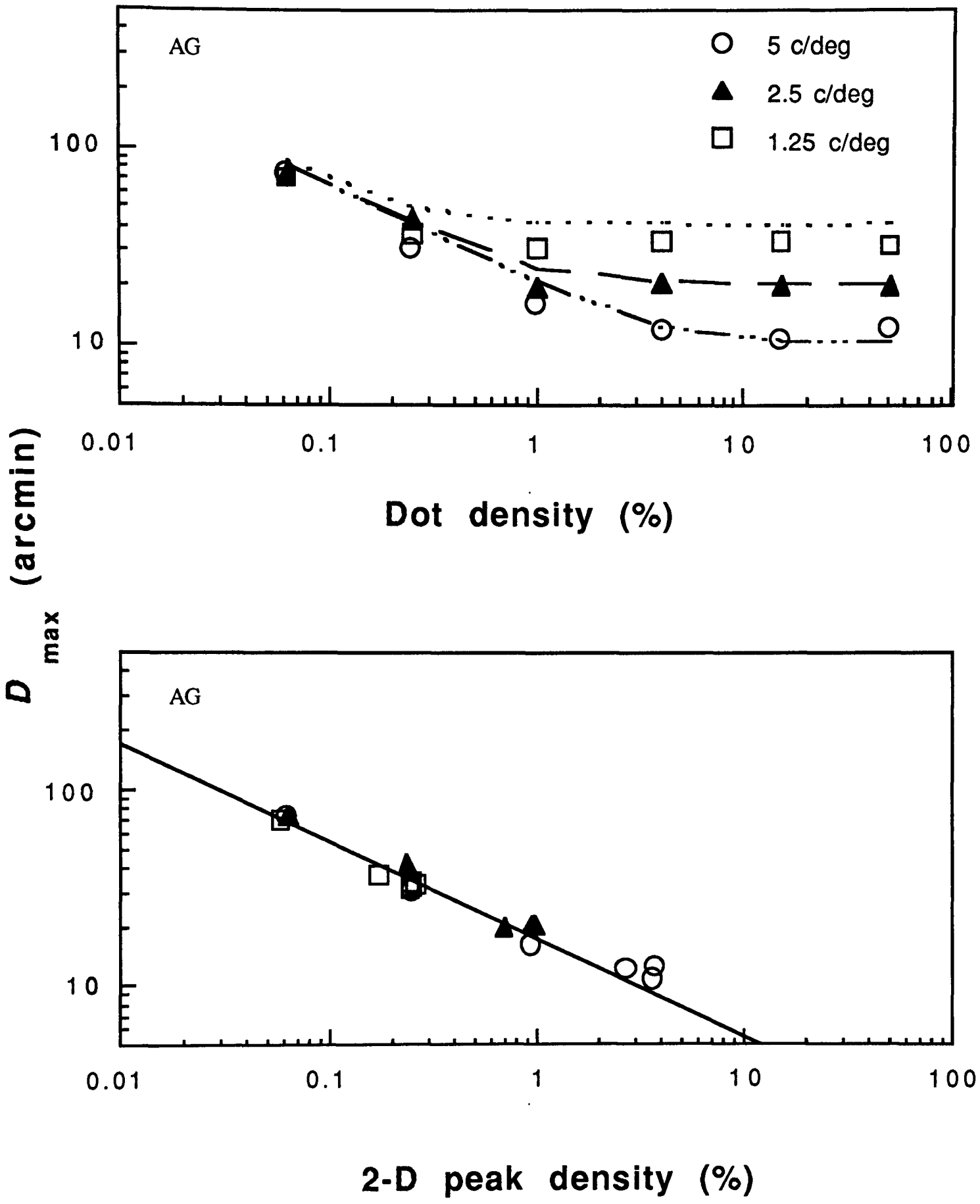


Fig 6.3

Results from Eagle and Rogers (1992). Dmax for a two-frame apparent motion task using bandpass filtered stimuli. Results for stimuli filtered at 5, 2.5 and 1.25 c/deg are shown. At the top, dmax is plotted as a function of the dot density in the pre-filtered image. Below, the same data are re-plotted as a function of the density of (2-D) luminance peaks in the stimulus. If dmax were dependent on the spacing of peaks in the stimulus then the data should fall along the dotted lines in the top graph or along the straight line (slope -0.5) in the bottom graph. Patch size was 7.15° in the direction of motion.

not at all on the spatial frequency content (the dotted lines show the predictions of such a model). This point is emphasised in the second plot shown in figure 6.3 (below). In this figure, the same data is re-plotted against the density of 2-D luminance peaks in the image. All the data points fall along a line whose slope is -0.5, i.e. a square root relationship. This pattern of results suggests that the main determinant of d_{max} is the 1-D spacing of primitives in the image (since the average 1-D spacing of elements, like the data, varies according to the square root of element density).

Thus, a model based on primitive spacing fits with most of the experiments on d_{max} discussed so far, i.e. those in which dot size was manipulated (Cavanagh et al.,1985; Sato, 1990; Morgan, 1992); those using band-pass filtered, 50% random dot patterns (Chang and Julesz, 1983; Cleary and Braddick, 1990a, Bischof and Di Lollo, 1990); and an experiment in which density and spatial frequency were manipulated independently (Eagle and Rogers, 1992). The exception is the experiment in which dot density was varied *without filtering* the images (Eagle and Rogers, 1991). The average 1-D spacing of dots in the stimuli used in this experiment varies with the square root of dot density, predicting a slope of -0.5 when d_{max} is plotted against density on log-log axes. But the data do not fit this prediction. The rise in d_{max} is much more gradual, the slope of the data for the largest patch size, 25.4°, is -0.23, i.e. half the rate predicted on the basis of dot spacing. Care must be taken in interpreting the results of Eagle and Rogers (1991). Whereas in most of the experiments discussed above mean luminance remained constant (with changes in spatial frequency, dot size or density of band-pass dots), when dot density is changed mean luminance changes almost in proportion to the number of dots in the image, i.e. by a factor of up to two log units for the range of densities tested by Eagle and Rogers (1991). Mean luminance is known to affect d_{max} (e.g. Dawson and Di Lollo, 1990). It also affects measures of local contrast (Peli and Goldstein, 1988) which may be important in determining d_{max} (e.g. Morgan and Fahle 1992). However, Eagle (1992) has repeated the dot density experiment of Eagle and Rogers (1991) using stimuli of opposite contrast (i.e. black dots on a white background instead of white dots on black) and obtained broadly similar results. This suggests that mean luminance, which for the black-on-white stimuli changes in the opposite direction as density is varied, is not the main determinant of d_{max} in these experiments. A similar control is used in the experiment described in this chapter and the effects of mean luminance and contrast are discussed in more detail in section 6.7.

If the results obtained by Eagle and Rogers' (1991) are due to the spacing of false targets then the relatively slow rise of d_{max} as dot density is reduced implies that the density of spatial primitives that are used in the motion correspondence process changes as more dots are added but does so much more slowly than the density of dots themselves. This idea is explored in more detail in section 6.7.

6.3 The rationale for this experiment

The main rationale for the experiments described in this chapter was to investigate whether similar limitations apply to the stereo correspondence process as have been found in experiments using a two-frame apparent motion task. The stimuli and procedure used in this experiment were, as far as possible, the same as those in Eagle and Rogers' (1991) experiment except that frame 1 and frame 2 were instead displayed as the left and right eye's images. The subject's task was to discriminate the stimulus as in front of, or behind, the fixation point and the disparity at which the subject made 20% errors was defined as "stereo-d_{max}". Thus, the information available (the input) and the criteria used for measuring the output were the same as that used in motion experiments. Eagle and Rogers' motion experiment was also repeated, as a control, to enable a direct comparison to be made between the experiments in the two domains.

The second aim was to investigate whether a false-targets theory in which primitives are derived from filter outputs at more than one spatial scale could help explain the pattern of results. Other possible explanations, such as the effect of contrast and luminance on d_{max} are also considered.

6.4 Methods

6.4.1 Subjects

Subjects were both experienced psychophysical observers with 6/6 vision.

6.4.2 Apparatus

Stimuli were generated on a Macintosh II computer and displayed on two monitors viewed with a Wheatstone apparatus at 57 cm, as described in section 4.4. At this viewing distance pixel size was 2 arcmin. For the motion experiment the stimuli on the two screens were identical.

6.4.3 Stimuli

The stimuli consisted of random dot patterns in which the density of dots ranged from 50% down to 0.006%, or two dots. In experiment I the dots were bright (32 cd/m^2) on a dark background (0.12 cd/m^2), in experiment II the luminance of the dots and background were reversed. The stimuli subtended 21° (horizontally) by 16° (vertically), i.e. 630 by 480 pixels (virtually the whole screen). Dot size was 6 arcmin square, as in Eagle and Rogers' (1991) experiment (each dot consisted of 3 by 3 pixels).

For each trial, two images of random dots were created, one a displaced version of the other (as illustrated in figure 6.4). Patterns of 1% and greater were plotted probablistically. That is, each point at which a dot could appear (i.e. in the stimulus array) had a given probability (equal to the dot density) of being white (in experiment I) or black (in experiment II). For patterns of lower densities (2 to 128 dots), the exact number of dots was plotted (at random x,y co-ordinates) so that random fluctuations in dot density were prevented. Dots from one image that were displaced outside the "window" were re-plotted on the opposite side of the displaced image, i.e. dots "wrapped round". Although, in theory, the correlation of these dots across the two frames could be discovered, the displacement is so large that they are likely to be treated as uncorrelated. At low densities (2 to 128 dots) any dots that wrapped around were given a new vertical position so that the chance of spurious "backward matches" was reduced.

Figure 6.4 (a) shows in schematic form a pair of frames illustrating the displaced (correlated) dots and the strip of uncorrelated dots in each image (equal in width to the displacement applied). Pairs of images created in this way are shown in figure 6.4 (b). For the motion experiment, the images were displayed as a two-frame apparent motion sequence, each frame lasting 150 ms with no inter-stimulus interval. For the stereo experiment, the two images were presented as a binocular pair (one to each eye). They were exposed simultaneously for 150 ms. In the case of an uncrossed disparity this corresponds to a veridical situation (a surface of dots seen behind a dark window), but for a crossed disparity it does not. (Shimojo and Nakayama (1990) and Howard and Ohmi (1992) discuss the role of "valid" and "invalid" occlusion cues in stereopsis.) This asymmetry does not occur for motion (the surface is seen to move to the left or right behind a dark window). Given

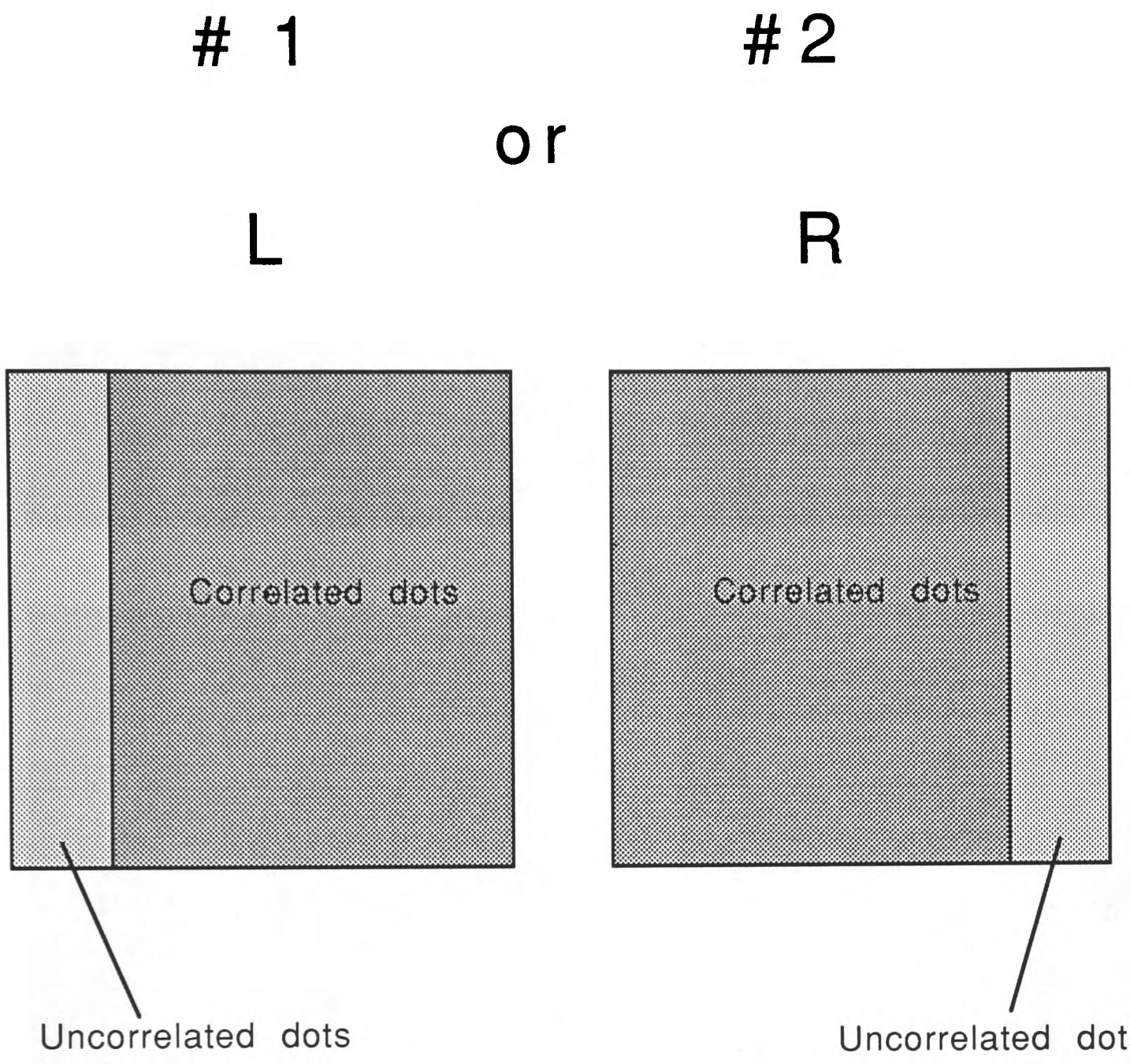


Fig 6.4 (a)

This diagram illustrates how pairs of images were created for display either as a 2-frame apparent motion sequence or as a stereo pair. The correlated dots were all shifted by the same horizontal displacement between frames. Uncorrelated dots filled the strips at the edges, as shown, so that the outline of the patch of dots was unchanged between frames.

Fig 6.4 (b) (see overleaf)

Examples of pairs of frames created in this way are shown for three densities: 0.024% (16 dots), 1% and 50%. The pairs can be fused. For cross-eyed fusion the dots appear in front of the page, illustrating the point that the uncorrelated dots are seen to shimmer. This does not happen for uncrossed fusion, when the dots appear behind a "window" (stereo and occlusion cues are consistent). The disparities of the images are about half that at dmax for each pair.

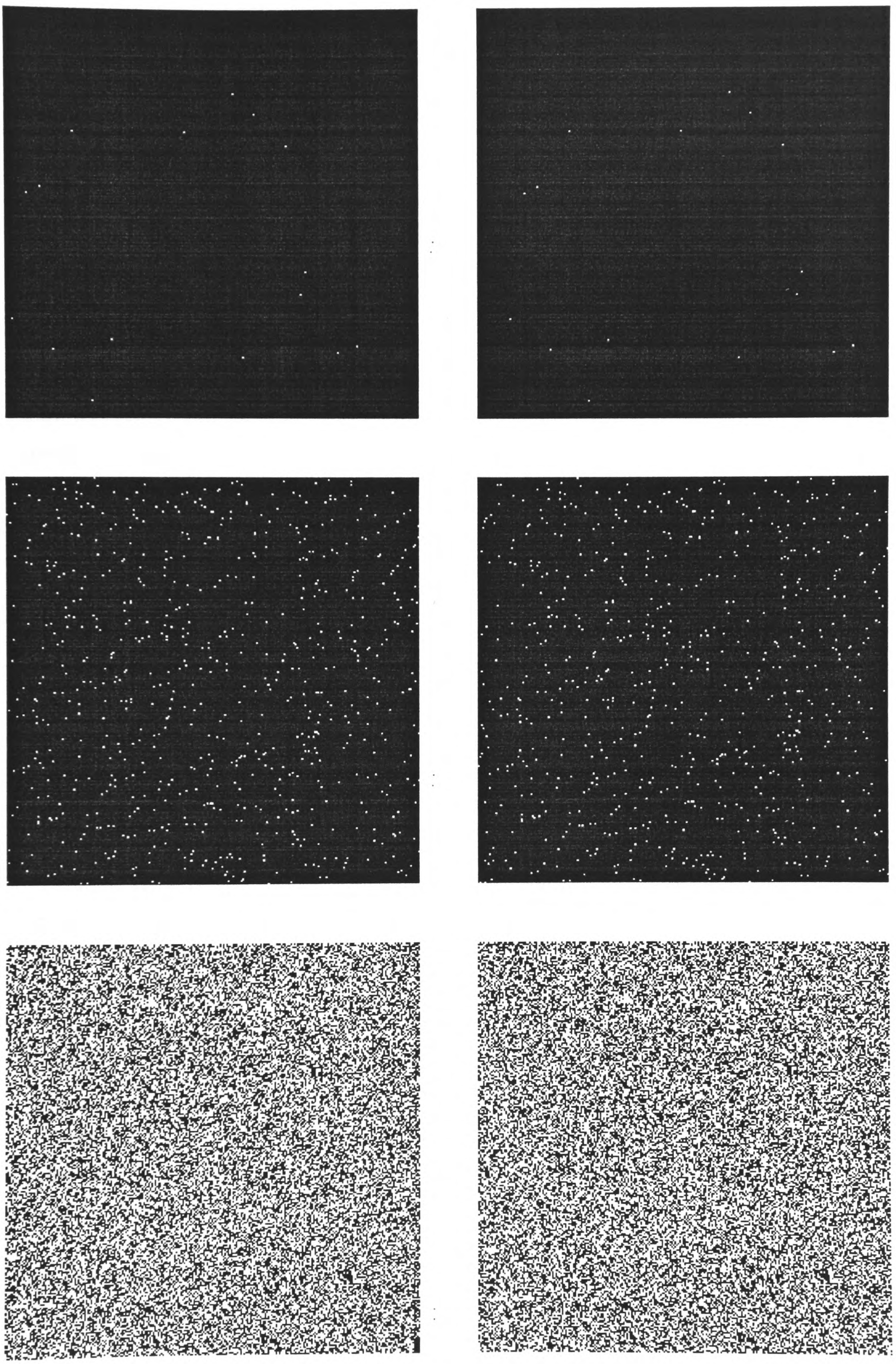


Fig 6.4(b) (legend on previous page)

unlimited time to view such patterns, the crossed stimuli appears less "solid" and the edge dots are seen as lustrous which is not the case for the uncrossed stimuli. For the brief exposures used in this experiment, however, subjects could not detect a difference in the appearance of crossed and uncrossed stimuli and there was no obvious bias in subjects' responses (the subject's responses were displayed at the end of each experimental run both as percent right button, from which any clear bias could be observed, and also as percent correct). The priority was to keep stimulus conditions as similar as possible to those in Eagle and Rogers' (1991) experiment. (The main differences were that, first, a square patch was used in their experiment, the largest of which was 25.4° by 25.4°. In the experiment described here a rectangular patch, 21° by 16° (21° in the direction of displacement) was used, which filled the whole screen. Second, motion was up or down in their experiment but left or right in the present experiment. And third, exposure duration in Eagle and Rogers' experiment (100 ms per frame) was slightly briefer than that used in the present experiment.)

After a trial the screen remained blank apart from a 15 by 15 arcmin fixation cross (which was present throughout). In experiment I the fixation cross was bright (32 cd/m^2) on a dark background (0.12 cd/m^2), in experiment II it was dark on a bright background. The subject responded by pressing one of two switches, and this triggered the next display.

6.4.4 Psychometric procedure

A 2-alternative forced choice (2AFC) procedure was used. The subject had to respond "left" or "right" for the motion task, "in front" or "behind" for the stereo task, by pressing one of two keys. A run consisted of 100 trials in which the stimuli were all of one density and all either motion or stereo. Five displacements (10 trials in each direction) were used in each run, presented in random order. The displacements used in any run were equally spaced (on a linear axis). Appropriate displacements (which would cover a range between a displacement at which a subject made 0% errors and one giving rise to 50% errors) were determined in a pilot run of 50 trials for each density. Larger displacements (and spacings) were used for the low density patterns.

Results from at least three runs were averaged (i.e. 60 trials per point). More data was gathered for low density patterns (at least one extra run for stimuli containing

less than 32 dots*). Errors increase monotonically with increasing displacement up to 50% errors (chance performance) at large displacements. D_{max}, either for motion or stereo was defined as the displacement that would give rise to 20% errors (e.g. Baker and Braddick, 1982). If none of the displacements gave exactly 20% errors then the two displacements giving rise to errors just below and just above 20% were used to estimate d_{max} (by assuming that error rates rose linearly between these two displacements, i.e. by "linear interpolation", (Baker and Braddick, 1982, p1257)).

6.5 Experiment I

In the first experiment all the patterns contained white dots on a black background. Data was gathered for the stereo and the motion conditions for two observers.

6.5.1 Results

Figure 6.5 shows results for two observers for both the stereo and the motion task. Results for the stereo task are shown at the top. There is a smooth transition between large stereo-d_{max} values (5 or 6°) at low densities to much smaller values (50-60 arcmin) at high densities. The slope of the best fitting power function for the mean of the data for the two subjects is -0.20. This is very similar to the slope found by Eagle and Rogers (1991) for motion-d_{max} for their largest patch size (see figure 6.2).

Fig 6.5 (overleaf)
At the top (a) results for the stereo experiment are shown for two observers. Stereo-d_{max} (defined in the text) is plotted against the density of dots in the stimulus, from 0.006% (i.e. two dots) to 50% density. In the centre (b), motion-d_{max} is plotted for the same two observers and the same range of densities. At the bottom (c), data for the two observers have been averaged so that results for the stereo and motion tasks can be compared directly.

---

* Performance was more variable for low density patterns, probably because the information in the stimulus is more variable when there are only a small number of dots. For instance, if a higher than average proportion of the dots falls in the uncorrelated region then the signal-to-noise ratio on that trial will be lower. The simulated values of dmax (section 6.7) were also more variable at low densities.

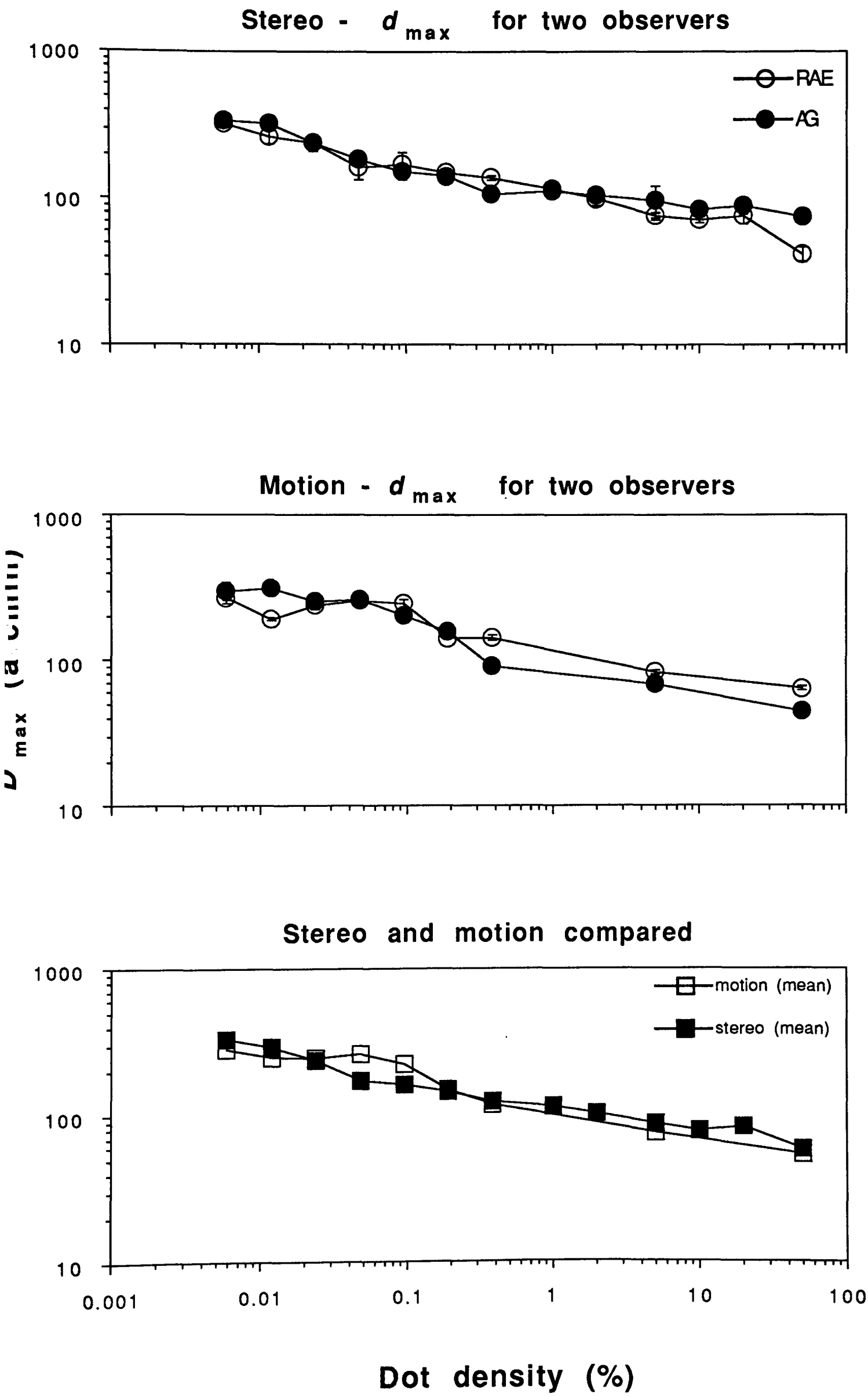


Fig 6.5 (legend on previous page)

Motion-d_{max} results for the current experiment are shown in figure 6.5 (centre). They are very similar to Eagle and Rogers' results and to the stereo data. The slope of the best fitting power function for the mean of the data for the two subjects is -0.18.

At the bottom the data for both the motion and stereo experiments are re-plotted (averaged across the two subjects) so that an explicit comparison can be made between performance in the two domains. The agreement between the data for stereo and motion tasks (both the slope and the absolute values of motion- or stereo-d_{max} at each density) is close.

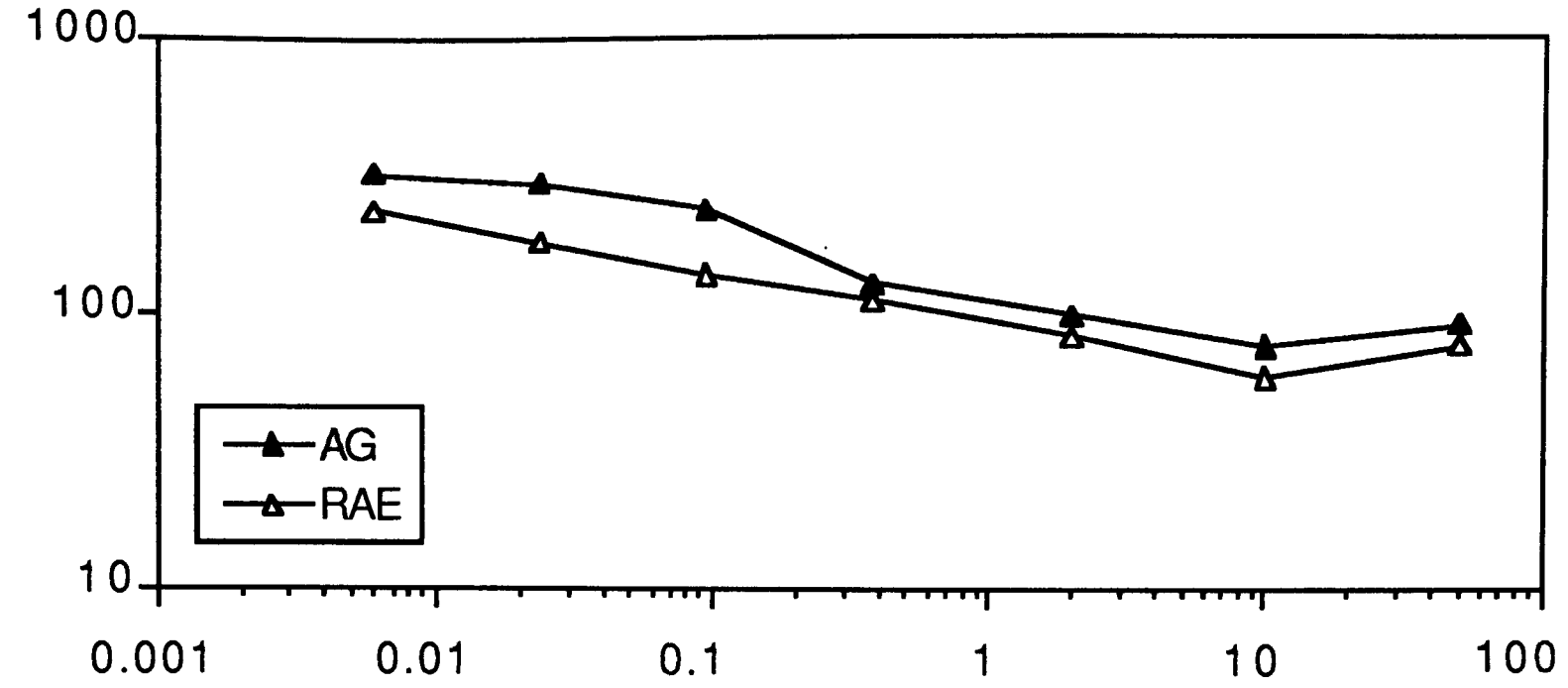
6.6 Experiment II

In the second experiment the patterns contained dark dots (0.12 cd/m²) on a bright background (32 cd/m²). Hereafter, this type of stimulus is referred to as black on white. Changes in dot density have the opposite effect on mean stimulus luminance in this experiment than in experiment I whereas the effect on dot spacing is unchanged. The effect of contrast-reversal has already been studied for motion-d_{max} (Eagle, 1992) and the results are discussed below. Hence, data was collected only for stereo-d_{max} in this experiment. Apart from the contrast reversal, the conditions and the range of densities used was the same as in experiment I.

6.6.1 Results

Figure 6.6 shows stereo-d_{max} for black-on-white patterns for two observers, for dot densities between 0.006 and 50% (top). The results are similar to those for experiment I. There is an important difference: the minimum stereo-d_{max} for both subjects is at a dot density of 10 rather than 50%. This pattern of results has also been reported by Eagle (1992) for motion-d_{max} using black-on-white patterns, and by Morgan and Fahle (1992) who found a rise in stereo-d_{max} for small black dots (4.5 arcmin) on a white background for densities between 5 and 50%. On the other hand, the results for experiments using white dots on a black background (Eagle and Rogers, 1991, shown in figure 6.2, and from experiment I, figure 6.5) show a drop in motion- or stereo-d_{max} over the same range of densities, provided patch size is sufficiently large.

The mean stereo-d_{max} results for experiments I and II are shown together in figure 6.6 (below). Despite the difference mentioned above, over the whole range of



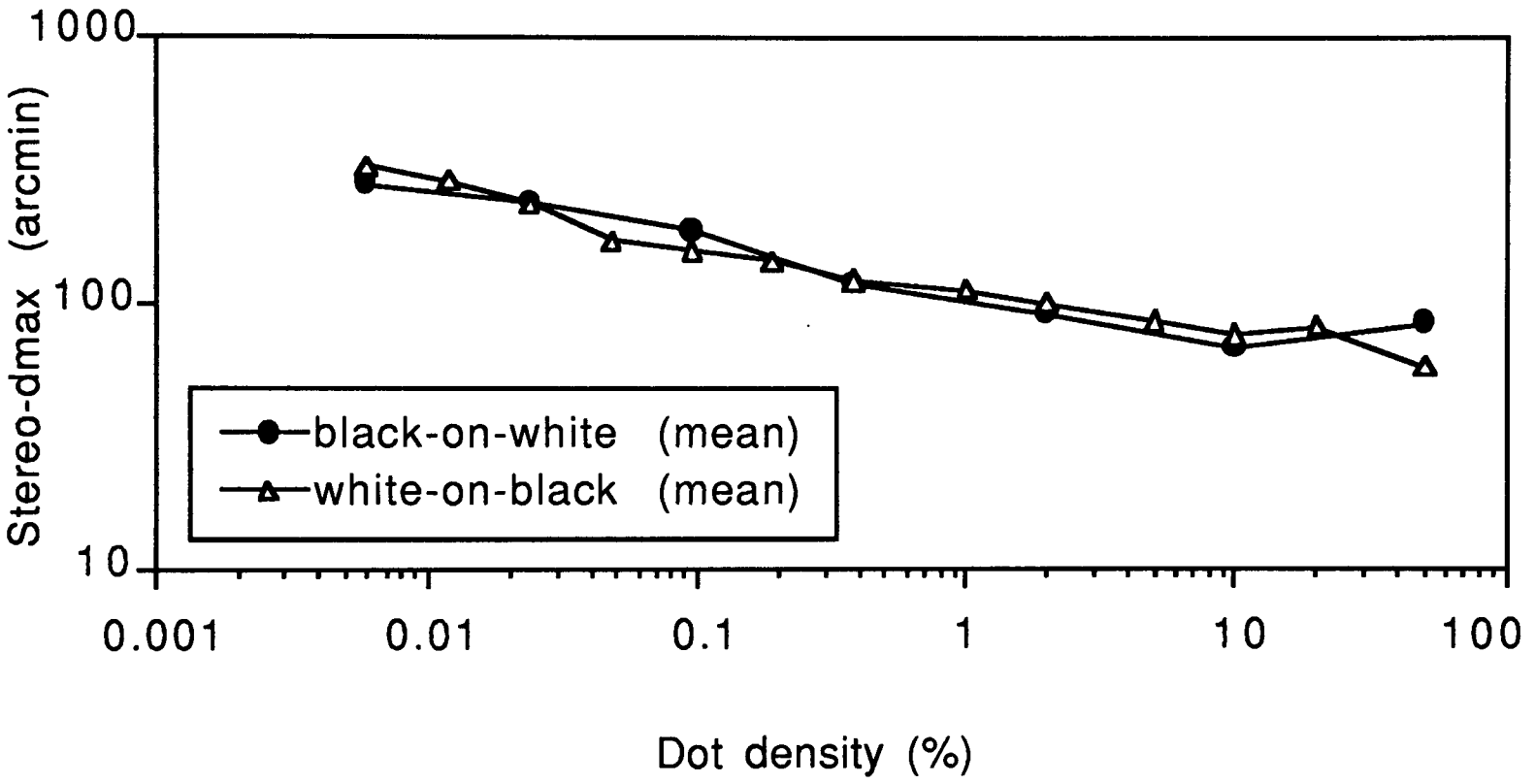


Fig 6.6

Results for experiment II. At the top, stereo-dmax is plotted against dot density (black dots on a white background) for two subjects. Below, stereo-dmax results for the opposite contrast condition (experiment I) can be compared directly: in each case the data shown are the the mean values of stereo-dmax for two subjects (white-on-black data are re-plotted from figure 6.5).

densities the results for the opposite contrast conditions are very similar. Interestingly, the main difference between the two conditions appears to be for 50% density patterns. The stimulus for a 50% pattern was of course the same whether the pattern was black-on-white or white-on-black but the luminance of inter-trial screen was very different in each case (0.12 cd/m².in experiment I and 32 cd/m².in experiment II). The 50% pattern appeared as a brief, bright flash in experiment I but this was not true for experiment II. Dawson and Di Lollo (1990) have shown that the level of adapting luminance can affect the magnitude of motion-d_{max} and this may help explain the difference (see discussion).

The mean luminance of the stimuli varies over a wide range as dot density is varied, but it does so in opposite directions for the stimuli used in experiment I and II. The fact that such similar results were obtained for the two experiments suggests that mean luminance is not the primary determinant of d_{max}.

6.7 Model

6.7.1 Similar limitations on the motion and stereo correspondence processes

The close similarity between the results for motion-d_{max} and stereo-d_{max} found in experiment I suggest that, at least for these stimuli, similar limitations apply to the motion and stereo correspondence processes. In less artificial circumstances, the limitations on the motion and stereo correspondence processes are probably quite different. For example, motion can be sampled at a range of temporal scales (or intervals) whereas there is a fixed interocular separation for stereo; objects can deform over time but this does not present a problem for binocular comparison; and motion can occur in any direction and so there is an aperture problem for mechanisms with a limited receptive field size (since they can only detect displacement perpendicular to an edge) whereas this problem is less severe for stereo since disparities are principally horizontal.

Marr and Poggio (1979) emphasised the importance of the 1-dimensional spacing of false targets in limiting the stereo correspondence process. In section 6.2 evidence was put forward that the 1-dimensional spacing of false targets may also limit the motion correspondence process in two-frame apparent motion sequences. If this is so, then the similarity between the results for motion and stereo found in

experiment I may imply that similar false targets, or spatial primitives, serve as input to the motion and stereo correspondence processes.

In the next section, alternative explanations for a change in d_{max} across different dot densities are examined.

6.7.2 The effect of contrast and luminance

Three factors vary with dot density, each of which have been shown to affect d_{max} when varied independently: dot spacing, mean luminance and contrast. Is it possible to determine which of these, or what combination, is responsible for the change in d_{max} as dot density is varied?

Dawson and Di Lollo (1990) measured motion-d_{max} for a two-frame apparent motion task at a range of adapting luminances. They used neutral density filters which affect mean luminance while maintaining stimulus contrast. They used a small patch (2 by 2° in the fovea or 1.5 by 8° in the parafovea) and a density of 20 dots/degree (bright dots, size 0.25 arcmin). There is a difficulty in comparing this value with the densities described in this chapter because of the very different dot sizes used. Although the patterns used by Dawson and Di Lollo were technically of a low density (0.03%), they contained the same number of dots/degree as the 20% density pattern used in the experiments described in this chapter and in Eagle and Rogers (1991).

Dawson and Di Lollo found that reducing adapting luminance by two log units raised d_{max} by about 50% for both foveal and parafoveal presentations. Dawson and Di Lollo also varied the contrast of the patterns and found no effect for dot/background luminance ratios between 6 and 200. Morgan and Fahle (1992), on the other hand, found that for a 5% dot density pattern reductions in contrast (from a Michelson contrast near 100% to 5%) caused a significant reduction in d_{max}. Dawson and Di Lollo used bright dots on a dark background while Morgan and Fahle used dark dots on a bright background and it is possible that this accounts for the difference in the results.

Although there is a wide range of mean luminances used in the experiments described in this chapter, there are many differences between the experiment Dawson and Di Lollo carried are comparable with the high density patterns used in the experiments described in this chapter, then their results might help to explain why a greater value of stereo-d_{max} was found using the brighter inter-trial screen

luminance for 50% patterns (experiment II). However, there are many differences between the experiments and the comparison is not a simple one.

The mean luminance of the random dot patterns used in experiment I and II can be expressed as:

$$\text{mean luminance} = P_w.L_w + (1 - P_w).L_b$$

where P_w is the probability of a white dot at each point in the image, L_w is the luminance of a white dot (32 cd/m^2) and L_b is the luminance of a black dot (i.e. background, 0.12 cd/m^2). Thus, mean luminance in experiment I varied from about 16 cd/m^2 for the 50% density patterns to about 0.12 cd/m^2 for the lowest density patterns. In experiment II, mean luminance changed in the opposite direction with dot density, from 16 cd/m^2 for a 50% density pattern to about 32 cd/m^2 for the lowest density patterns. The fact that stereo-d_{max} varied in a similar way with dot density in experiments I and II suggests that mean luminance is not the main factor determining d_{max} in this situation.

Mean luminance, or some measure that depends on it, is used in most methods of calculating contrast (Peli and Goldstein, 1988) but this is not the only factor that affects contrast as dot density is varied. There are many definitions of contrast and the way in which contrast varies with density depends on which definition is chosen. Michelson contrast is defined as

$$\text{Michelson contrast} = \frac{\text{Lmax - Lmin}}{\text{Lmax + Lmin}}$$

but this is only appropriate for periodic patterns such as interference fringes for which Michelson originally proposed the measure (1891). A different measure, which again gives a single number to describe the contrast of a whole image, is root mean square contrast, defined as

$$\text{R.M.S. contrast} = \sqrt{\frac{1}{n-1}\sum_{n}^{i=i}(x_i - \bar{x})^2}$$

where $\bar{x}$ is the mean luminance, x_i is the luminance of the i^{th} pixel and n is the total number of pixels in the image. The r.m.s. contrast for the 50% density patterns used in experiment I and II is about 0.5, and for a 0.006% density pattern is 0.06.

Thus, by this definition, contrast, for both the white-on-black and black-on-white patterns, goes down as density is reduced.

An alternative (and more modern) approach to assessing contrast is to use a local measure. Maintaining the general notion of contrast as a measure of luminance difference divided by a measure of mean luminance, Peli and Goldstein (1988) suggest that a local measure of contrast could be derived by calculating the ratio of outputs of a band-pass operator and a low-pass operator at each point in the image

$$c(x,y) = \frac{a(x,y)}{l(x,y)}$$

where c is the contrast at point (x,y), a(x,y) is the output of a band-pass operator at that point (they suggest an ideal filter with a bandwidth of one octave) and l(x,y) is the output of a low-pass operator (they suggest one containing all frequencies below the band-pass of a). Using this definition the peak contrast in an image goes up as density is reduced for a pattern made up of white dots on black but down for a pattern of black dots on white. (The local amplitude of a(x,y) goes down in both cases with density but, for the white-on-black pattern the amplitude of l(x,y) goes down at an even greater rate. For the black-on-white pattern l(x,y) goes up as density is reduced.) Again, as for mean luminance, it is unlikely that contrast, at least according to this definition, can account for the change in d_{max} with dot density given the similarity between the results of experiment I and II.

The implementation of Peli and Goldstein's measure of contrast would involve local mechanisms with receptive fields perhaps similar to Laplacian operators whose output would be compared to that of a low-pass mechanism (e.g. a Gaussian operator with a larger receptive field. An alternative measure, which Peli and Goldstein do not consider, is one in which the blurring and differencing process take place within the same receptive field (which is one way to describe the operation of a Laplacian of Gaussian). If the the input to this operator were passed through a log mechanism before arriving at the LoG filter then the output would correspond quite closely to the measure c(x,y) that Peli and Goldstein propose. For instance, the response to a high frequency sine wave would reduce as the pedestal luminance was increased (fig 4, Peli and Goldstein, 1988). The output of photoreceptors is proportional to the log of the luminance input, at least in the mid-range of their response (e.g. Fain and Dowling, 1973) This type of non-linear process before filtering has been discussed in the past (e.g. Davidson, 1968).

Of the three main factors that could potentially account for the change in d_{max} across dot density (contrast, luminance and element spacing) only element spacing appears to be compatible with the results of experiment I and II taken together. The way in which element spacing might account for the results is considered in the next section.

6.7.3 The spacing of spatial primitives

If the spacing of false targets accounts for the change in d_{max} across density, then what *are* these spatial primitives or false targets? As mentioned in the introduction, the slope of the graph in figure 6.5 gives an important clue. For example, the dots themselves are not suitable candidates. Their mean spacing in the direction of displacement varies inversely with the square root of dot density. This predicts a slope of -0.5 when d_{max} is plotted against dot density on log-log axes. The data, for both stereo and motion, changes much more gradually (a slope of about -0.2).

Primitives derived from a single filter output are also unsuitable candidates. Figure 6.7 shows examples of stimuli at different densities filtered at a range of spatial scales. For a small filter and low density (shown in figure 6.7.(a)), there is one "blob" in the filtered output corresponding to each dot in the image. For high density patterns, on the other hand, many of the blobs coalesce and the number of blobs in the filtered output is determined by the filter size rather than dot density. For example, the output of larger filters (shown in figure 6.7 (b) and (c)) is very similar for a 1% and 50% density image. (The amplitude of the output changes but not the density of spatial primitives such as peaks, zero-crossings or centroids.)

Fig 6.7a

On the left are shown examples of the type of stimuli used in the experiment (i.e. one eye's image or one frame in the motion experiment). Three densities are shown: 0.024% or 16 dots (top), 1% and 50%. On the right the same images are shown after filtering with a Laplacian of Gaussian filter with a space constant of 2 pixels (i.e. 12 arcmin at the viewing distance used in the experiment). Each filtered image has been scaled to cover the same range of grey levels.

Fig 6.7b

The same images as shown (a) filtered with a LoG with a space constant of 4 pixels (left) and 8 pixels (right).

Fig 6.7c

The same images as shown (a) filtered with a LoG with a space constant of 16 pixels (left). On the right is shown the MIRAGE S+ response to the same images. The filter outputs contributing to MIRAGE response have space constants of 16, 8, 4 and 2 pixels, i.e. those shown in the previous examples.

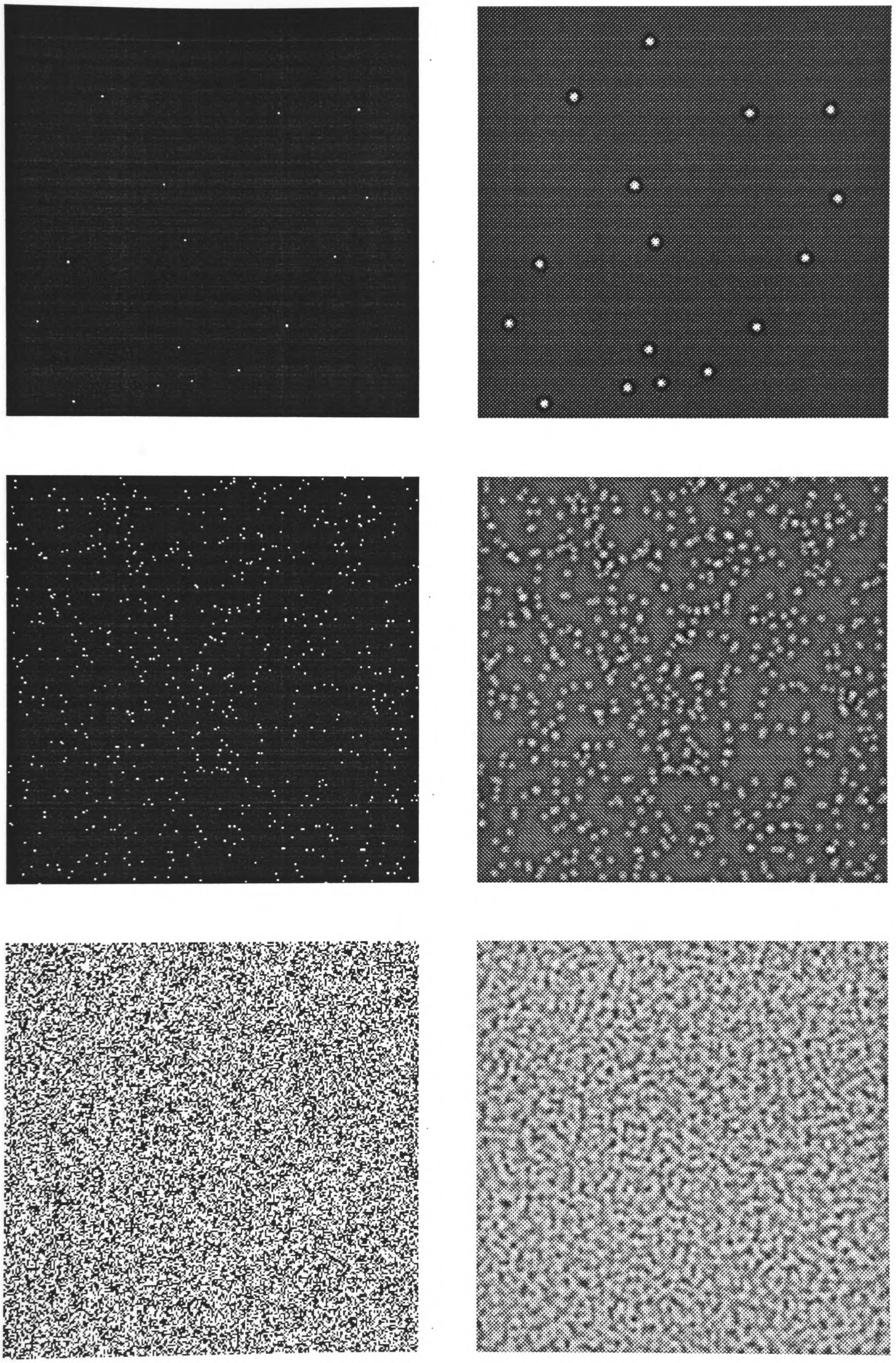


Fig 6.7(a) (legend on previous page)

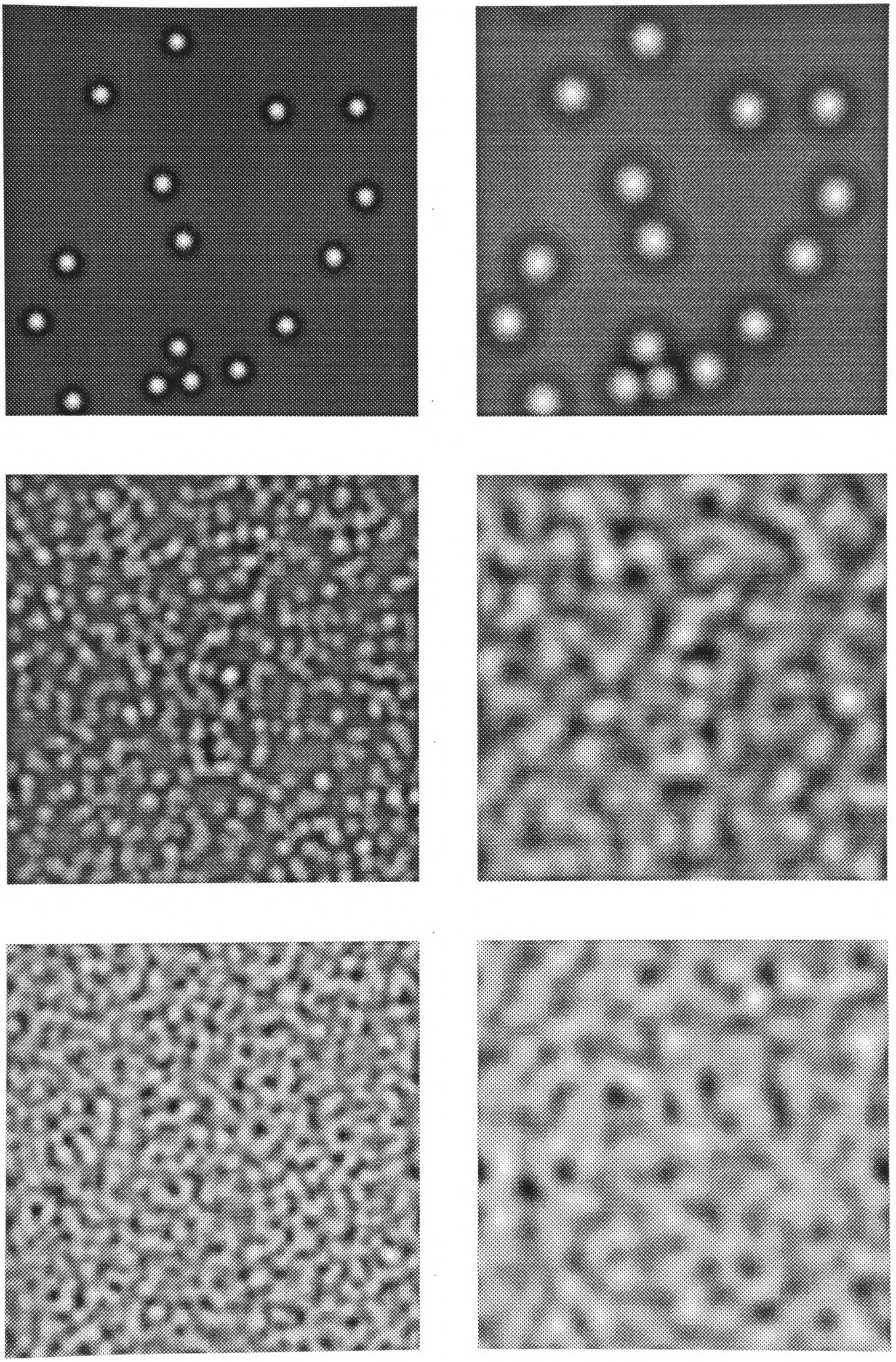


Fig 6.7(b) (legend on previous page of text)

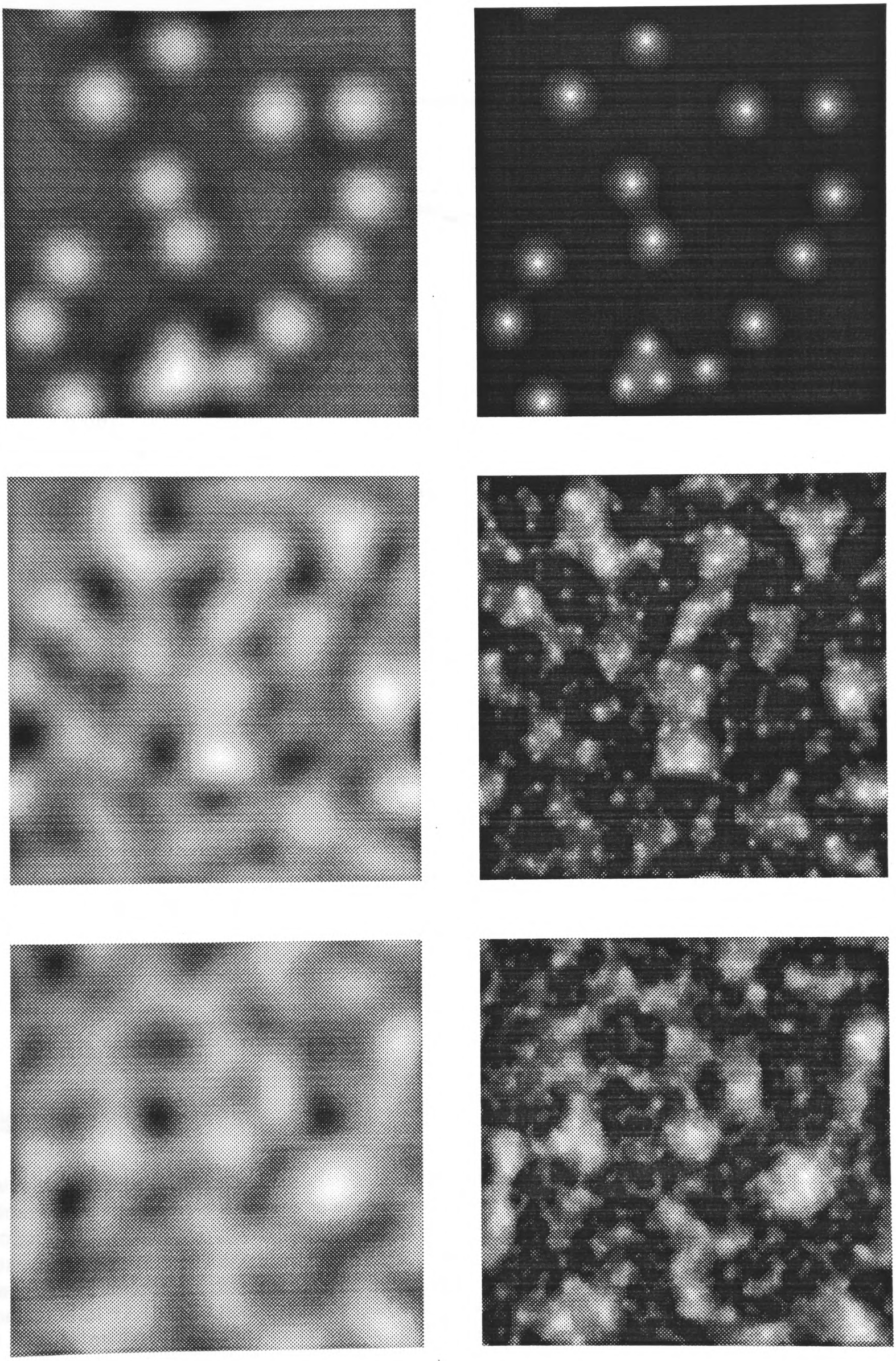


Fig 6.7(c) (legend on previous page of text)

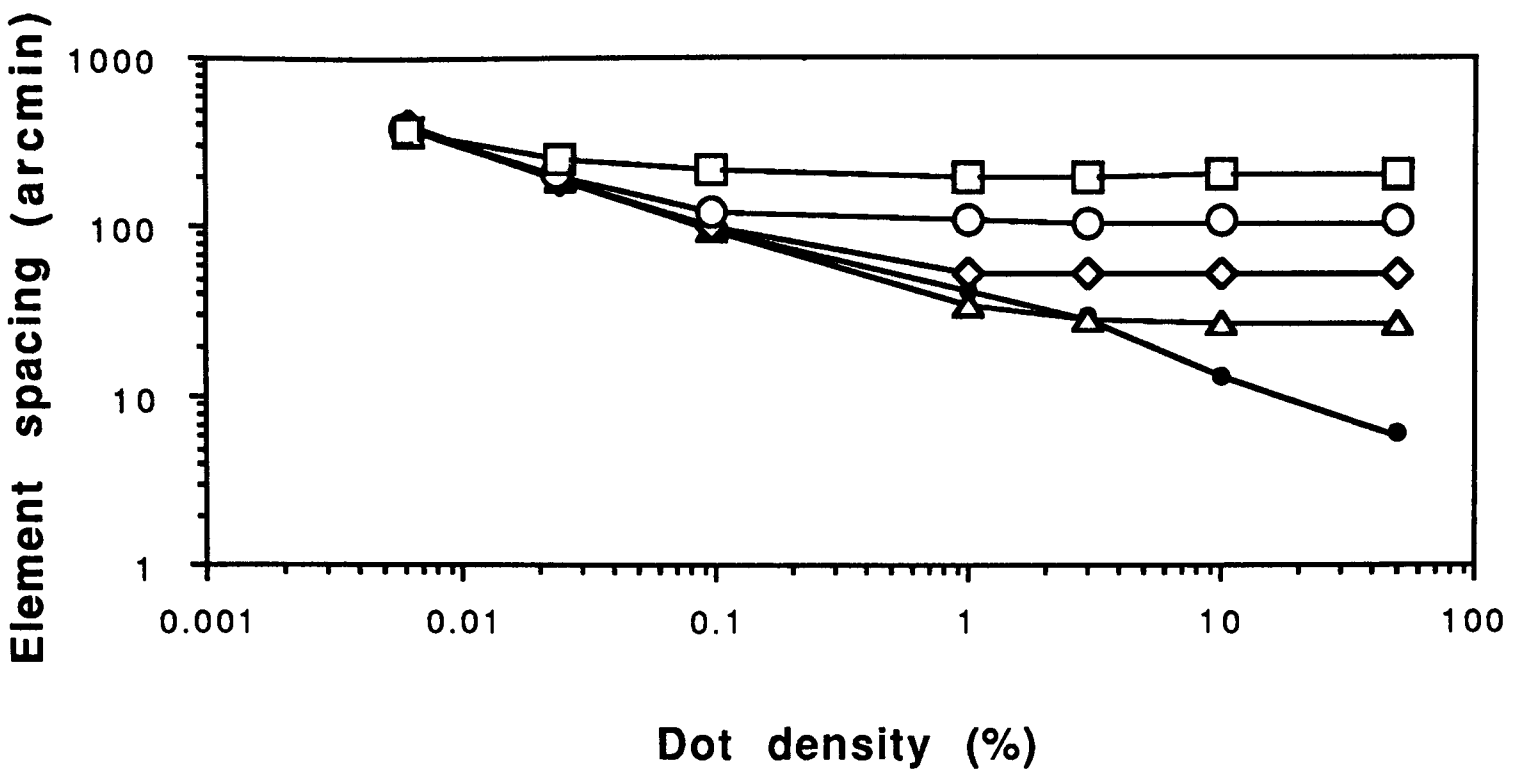


Fig 6.8

Results of a model in which the average spacing of elements in a filtered version of the stimulus (in this case the elements chosen are the 2-D peaks in the filtered output) is plotted against the number of dots in the original (unfiltered) stimulus. The average spacing of elements is defined as the stimulus area divided by the number of elements. Results are shown for patterns filtered with Laplacian of Gaussian filters with space constants of 17, 34, 68 and 138 arcmin (dot size was 6 arcmin). The average spacing of dots in an unfiltered pattern is also shown (filled circles).

These points are illustrated in the graph shown in figure 6.8. One spatial primitive has been chosen in this example, 2-D peaks in the filter output, but any candidate primitive based on a single filter output would show a similar pattern. For low densities, all the filter outputs contain a similar number of 2-D peaks so the curves for each filter co-incide. For high densities the curves divide, each reaching a plateau at a different level. All the curves have the same property: they fall with a slope of -0.5 and then flatten off. The only difference between them is the point at which they reach their plateau.

The experimental data (figure 6.5) do not fit any of these single filter predictions. D_{max} changes only slowly at low densities, which is a property of the larger filters shown in figure 6.8 (because they are near their plateau level). D_{max} continues to change over most of the range of densities (the *whole* range for white-on-black patterns) which seems to imply that quite fine filters are affecting d_{max} too.

The properties of the MIRAGE response fit this description well. Figure 6.7 (c) shows the MIRAGE (S+) response* for a range of densities. The filters contributing to the MIRAGE response are the four shown in figure 6.7(a) and (b). The method used for combining the filters is described in appendix B. At high densities the response contains many fine scale blobs, some grouped within larger scale blobs, others isolated. At the lowest density the image looks rather like the finest filtered version but, in fact, it is the zero bounded regions that are important for the MIRAGE analysis and these correspond much more closely to the outline of the coarsest filter blobs.

This point is made more clearly by figure 6.9 which shows the 1-D centroids of the MIRAGE blobs. (The calculation of centroids is described below and in appendix B.) On the left of this figure are shown the centroids for the positive

---

Fig 6.9

On the left is shown the positive response of a LoG filter (σ = 4 pixels) to the three patterns shown in figure 6.7a with, overlaid, the 1-D centroids (in a horizontal direction) of the blobs. On the right are shown the MIRAGE S+ responses to the same images and again the centroids of each blob are marked.

---

* The MIRAGE S- response for a 50% density pattern is the same as the S+ response (statistically). For low density patterns of white dots on a black background the S- response is a large "sea" with a few "holes" corresponding to each dot. Examples of this type of response are illustrated in chapter 2. For patterns consisting of black dots on a white background the characteristics of the S+ and S- signals are reversed.

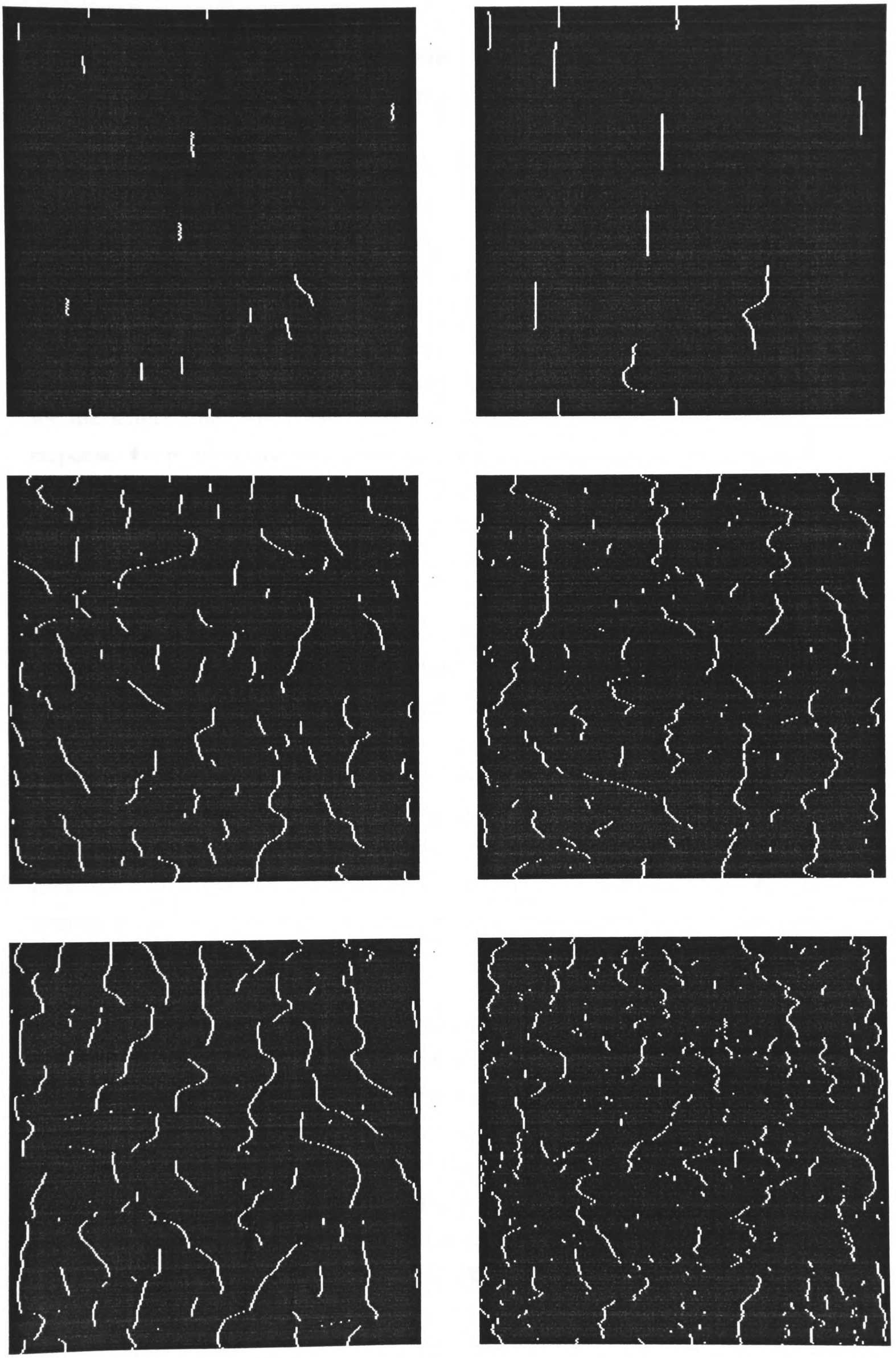


Fig 6.9 (legend on previous page)

response of a single filter (σ = 48 arcmin). On the right, the centroids for the MIRAGE S+ responses to the same patterns are shown.

For the lowest density pattern, the single filter (shown on the left) has resolved most of the dots while the MIRAGE response (shown on the right) has grouped them together into a small number of blobs, i.e its behaviour is like a coarser filter. In the centre the responses to a 1% density pattern are shown. There are a similar number of centroids in the single filter and MIRAGE responses, although the MIRAGE blobs are less regular. The responses to the 50% pattern are shown at the bottom. There is little change in the spacing of centroids between 1% and 50% for the single filter (left), but there is a noticeable change in the MIRAGE response. There are many more centroids, mostly corresponding to small, isolated blobs. That is, the MIRAGE response, or at least the *change* in the MIRAGE response, between 1% and 50% densities reflects the contribution of a finer filter than the single filter shown on the left.

In other words, MIRAGE centroids display the crucial properties that are required to model the dot density results successfully: they mimic a coarse filter output at low densities and a fine filter output at high densities.

Predictions based on MIRAGE centroids can be tested more quantitatively. For instance, their mean separation at different densities can be calculated. It is not appropriate simply to count them and calculate their average separation as was done for 2-D luminance peaks (figure 6.8), since MIRAGE centroids are 1-D primitives, but it is reasonable to calculate their mean separation in a horizontal direction (the direction of displacement).

The results of a model of this type are shown in figure 6.10. For each centroid in an image the horizontal distance to the nearest centroid along the same raster line was calculated and the mean of these distances plotted against the dot density in the original image. If no neighbouring centroid was found, the separation for that centroid was set as equal to the width of the pattern (256 pixels, i.e. 256 times 6 arcmin). Since this assumption skews the results, the model was only used for patterns in which over 90% of the centroids had neighbouring centroids along the same raster line. This applied to patterns of 1% density and above.

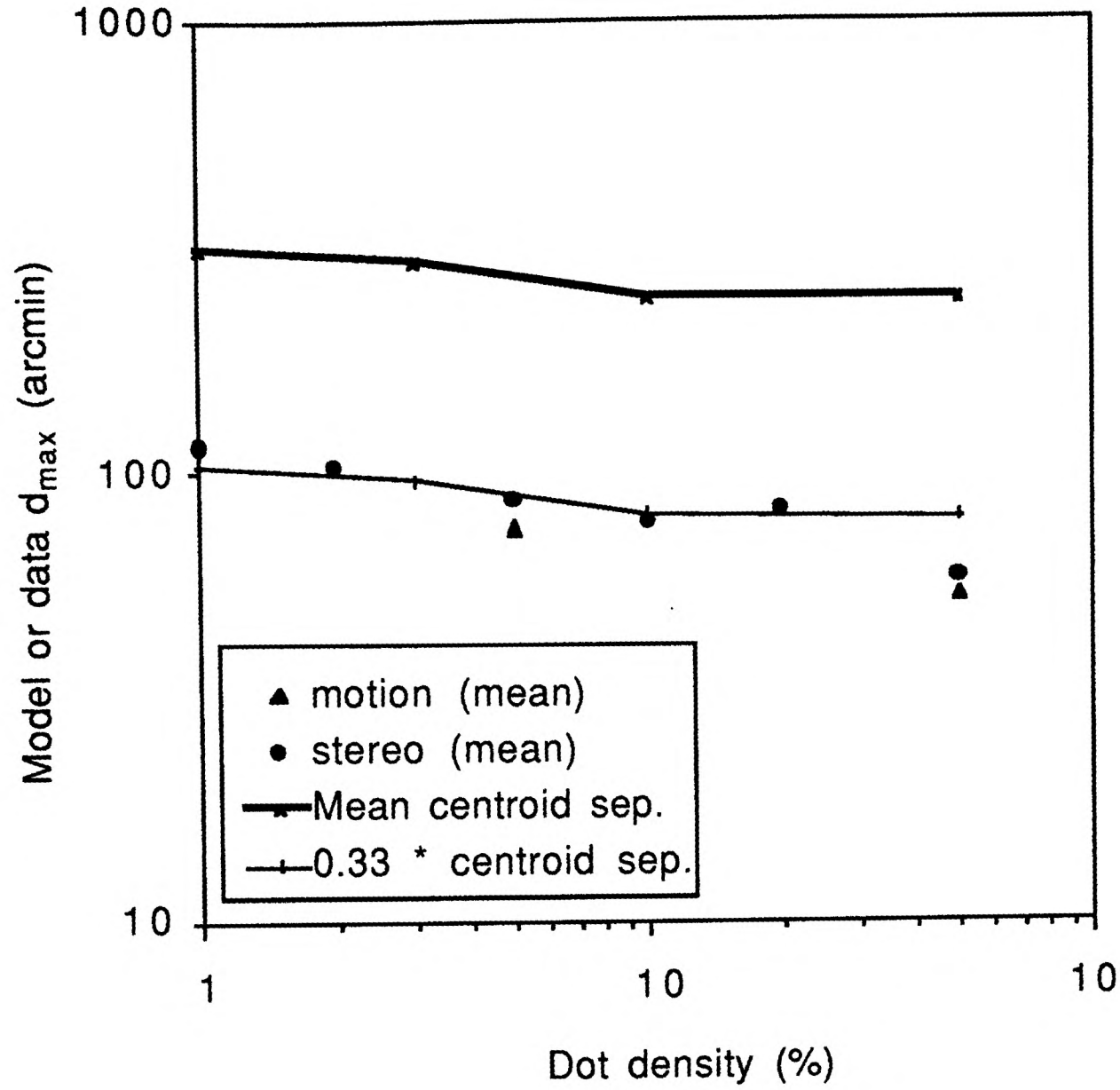


Fig 6.10

This figure shows how the mean horizontal separation of MIRAGE centroids varies with dot density (bold line). Values are shown only for densities above 1% because the model was not considered to be valid for lower densities (see text). The data points indicate the mean values of dmax for two subjects for motion and stereo (taken from figure 6.5). The psychophysical data are approximately one third of the mean centroid separation, although the fit is not close.

An extra assumption is required to predict the magnitude of d_{max} at any density. That is, a false-targets theory predicts that d_{max} should be proportional to primitive spacing but does not determine *what* proportion. D_{max} for a repetitive pattern such as a sine wave is somewhere between a quarter and a half the spatial period. Figure 6.10 shows the mean spacing of MIRAGE centroids and also the predicted results if d_{max} were one third of centroid separation. The experimental data lie within this range, but the fit of the model is not close. (In a similar model of d_{max}, Morgan and Fahle (1992) used a constant of a quarter).

This type of model is not ideal and is especially poor at low densities, because of the problem discussed above. A slightly more sophisticated version of the same idea is to use a simple matching algorithm. It is important to note that the purpose of this exercise is to obtain a measure of element spacing and the point at which a false targets scheme might break down, it is not an attempt to model the details of a matching algorithm that might be used by the visual system.

Two random dot patterns were created, one a shifted version of the other and the MIRAGE output derived for each (1-D centroids). Then, for each centroid in one image, the nearest neighbour centroid in the other image was found. For small displacements, all the correct matches should be made. For large displacements, and certainly by the time the displacement is equal to the mean centroid separation, many false matches should be made and direction discrimination should fail. Two typical centroid images, used as input to the matching algorithm are illustrated in figure 6.11. Details of the model are given below.

A pair of random dot patterns, each 256 by 256 pixels, was created by adding a given displacement to the dots in one image. Dots "wrapped round" so that dots shifted out of the image were re-plotted on the opposite side of the image, as for the experimental stimuli. For low densities, as in the experiment, dots that wrapped round were given a new (random) vertical position. Each image was filtered (with Laplacian of Gaussian filters with space constants 2, 4, 8 and 16 pixels, i.e. modelling filters of 12, 24, 48 and 96 arcmin). The equation for a Laplacian of Gaussian filter in the spatial domain is

$$\nabla^2 G\,(r,f,\theta) = \left(1 - \frac{r^2}{2f^2} \right) e^{-r^2/2f^2}$$

where r is the distance and θ the direction from the centre, f is the space constant (standard deviation of the Gaussian). The peak-to-trough amplitude of each filter in

the spatial domain was equal, as in the original MIRAGE model. The output of each filter was half-wave rectified and the positive responses summed to give an S+ signal. The filters are much larger than those described in chapter 2 but the patch size (and hence eccentricity) is greater and the exposure duration shorter. The centroid of each blob was calculated in 1-D, i.e. along horizontal raster lines. The centroid, P_i, is the position within a zero-bounded distribution about which the first order moment is zero:

$$P_{i,} = \frac{\int_{Z_{c_i}}^{Z_{c_{i+1}}} x \,.\, R(x) \,.\, dx}{\int_{Z_{c_i}}^{Z_{c_{i+1}}} R(x) \,.\, dx}$$

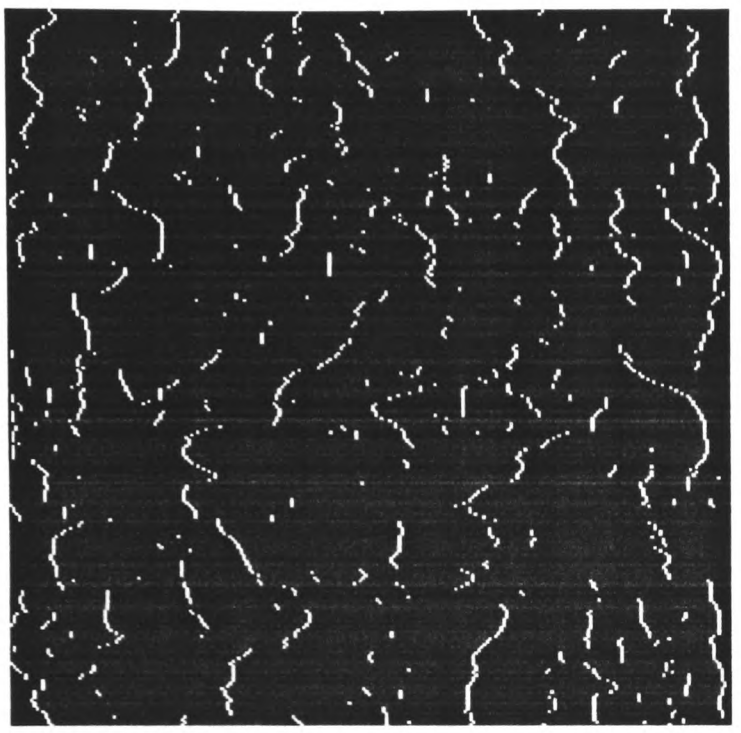
where Z_{c_i} and $Z_{c_{i+1}}$ are the positions of adjacent zero-crossings and R(x) is the response at point x. All these steps are given in more detail in appendix B. The input to the matching algorithm was a difference picture like that illustrated in figure 6.11. (In fact, one image was multiplied by 2 before subtraction so that when two centroids overlapped the output would not be zero.) For each centroid derived from the left eye's image (white pixel) the nearest centroid from the right eye's image (black pixel) was found (along a raster line). The proportion of matches made in the "correct" direction was recorded.

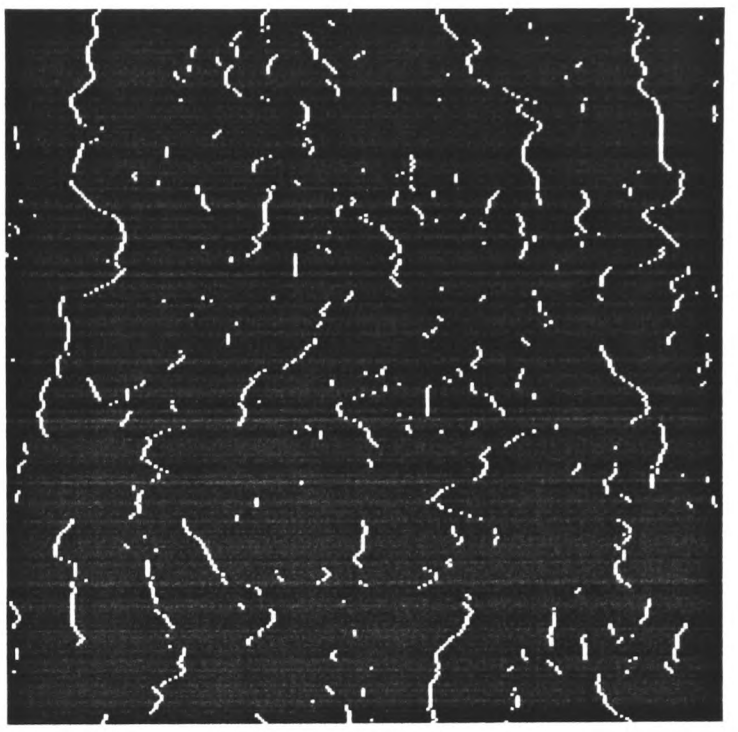
This was repeated for a range of dot displacements. The proportion of matches in the correct direction falls as the displacement increases and asymptotes at 50% errors, in the same way as the proportion of correct responses falls with increasing displacement in the experiment. Ten different random dot patterns were used. For each pattern, displacements of increasing magnitude were tested, covering the range from a displacement giving rise to no errors to one giving rise to 50% errors. The mean proportion of centroid matches made in the correct direction was

---

Fig 6.11

The pair of images at the top shows the centroids of the blobs derived from the MIRAGE S+ response to a 50% pattern (σ = 16,8,4 and 2 pixels). The original pattern has been shifted by 10 pixels and this is reflected in the positions of the centroids. The image below (a subtraction of the pair of images at the top) demonstrates the shift: white centroids are from one image, black from the other. An image similar to this would form the input to the matching algorithm (see text for details).





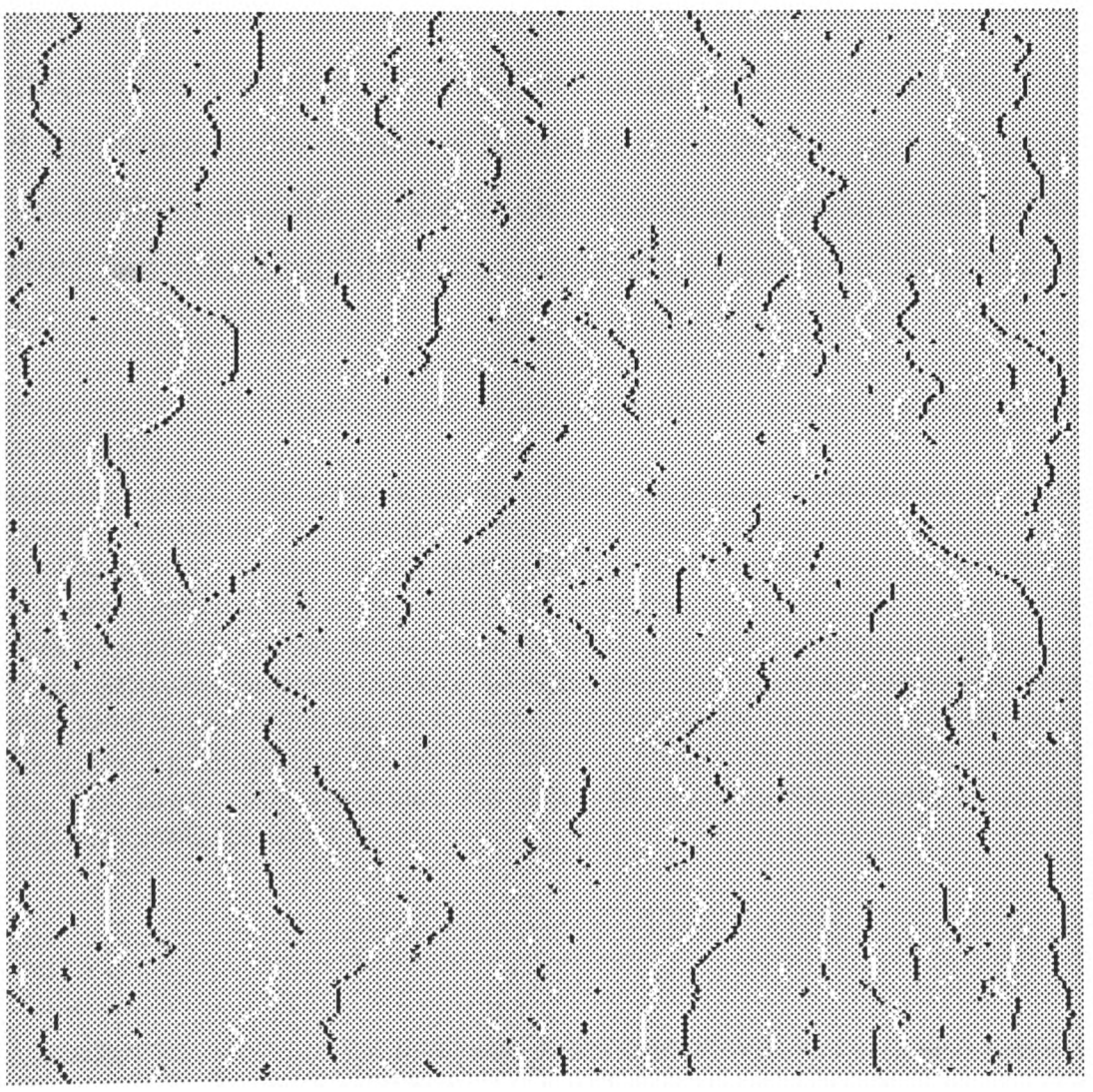


Fig 6.11 (legend on previous page)

calculated for each displacement (averaging across the ten patterns). The variance of these values was small for the high density patterns and large for the low density patterns. The same is true for the number of correct responses made at different displacements in experiment I and II: a larger number of trials was needed to arrive at a reliable estimate of motion- or stereo-d_{max} for the low density patterns.

A similar criterion was used to define the theoretical d_{max} as for the experimental data. The displacement for which 20% of centroid matches were made in the wrong direction was defined as a theoretical d_{max}*. This is plotted for a range of dot densities in figure 6.12 (top).

Also shown in figure 6.12 (top) are the mean results for motion- and stereo-d_{max} from experiment I (dotted lines). The model fits the data well across the whole range of densities. The model shows a decrease in the value of theoretical d_{max} between 10 and 50% densities, so it would not fit the data from experiment II so well over this range.

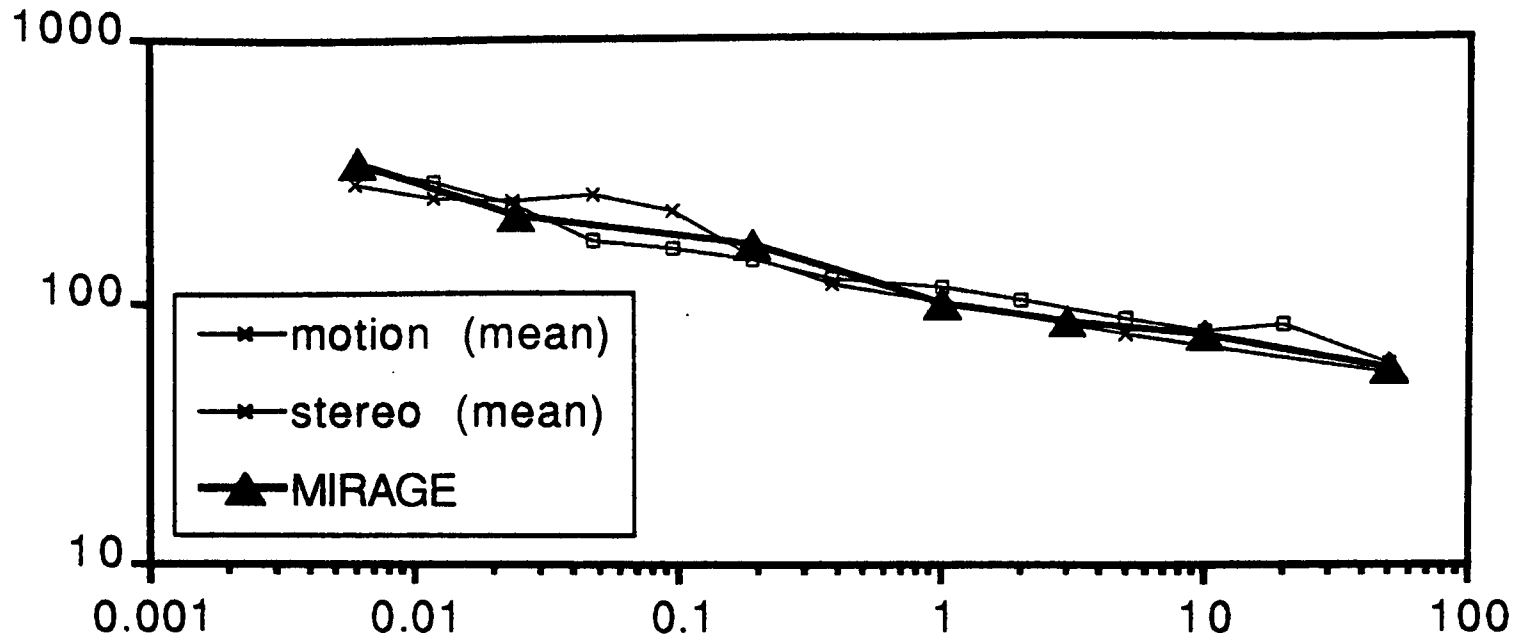
It may be a co-incidence that the absolute values of theoretical d_{max} are so close to the experimental results. One free parameter in the model is the minimum mass of a zero-bounded distribution for which a centroid is recorded, and changing this might affect the vertical position of the model. Another parameter which was not used in the model, but which might be expected to play a part in a more realistic

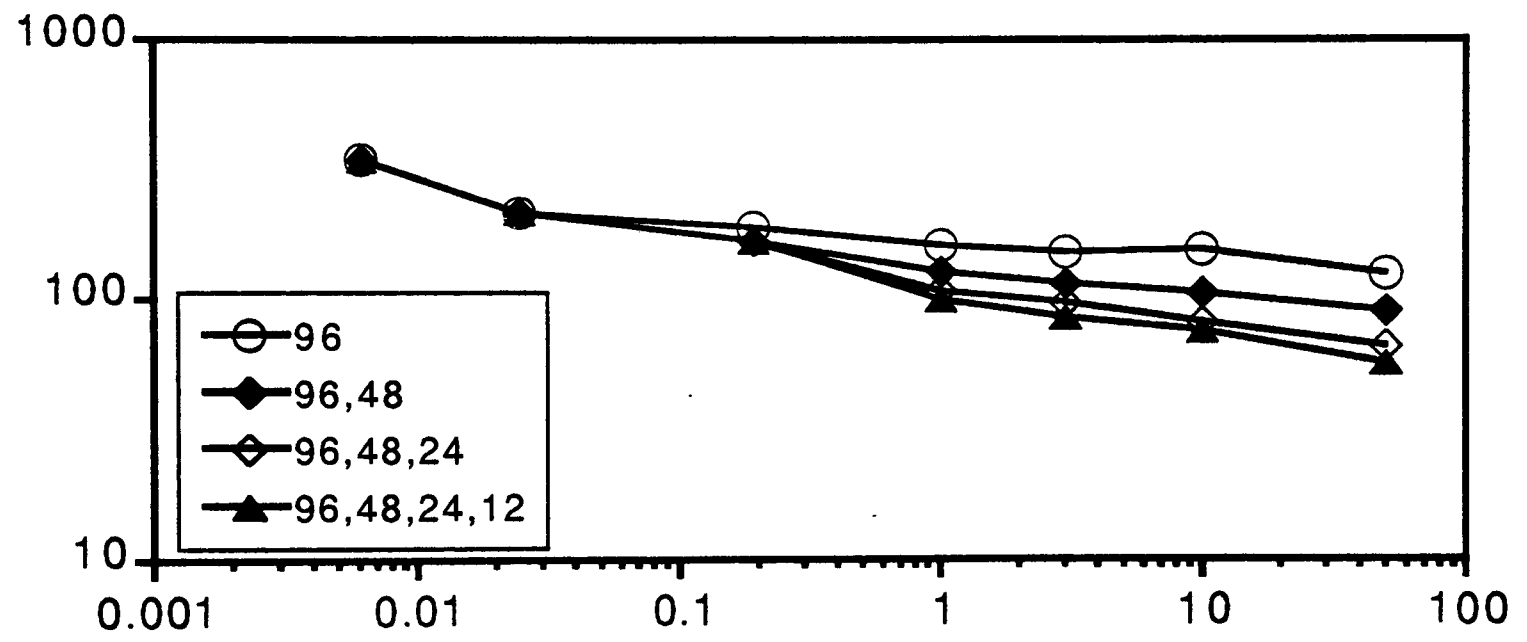
---

Fig 6.12 (overleaf)
Results of the matching model described in the text. Theoretical d_{max} is plotted on the ordinate in arcmin (pixel size is six arcmin), dot density is plotted on the abscissa from 0.006% (4 dots) to 50%. At the top the results of the model using MIRAGE centroids are shown. The space constants of the filters contributing to the MIRAGE signal were σ = 96, 48, 24 and 12 arcmin. Also shown in this plot are the mean motion and stereo-d_{max} data from experiment I (re-plotted from figure 6.5) for densities between 0.006% (2 dots) and 50%. In the centre plot, results are shown for the matching model when the filters contributing to the MIRAGE signal are σ = 96, 48, 24 and 12 (as in the top plot); σ = 96, 48 and 24; σ = 96 and 48; and σ = 96, i.e. the largest filter on its own. The lower plot illustrates the effect of omitting the largest filter from the MIRAGE signal (i.e. σ = 48, 24 and 12).

---

* The mean and standard deviation of the distances to the nearest matches was also recorded. For small dot displacements, most matches are correct so the mean distance to the nearest match equals the dot displacement and the standard deviation of estimates is small. For very large displacements (or uncorrelated patterns) the mean is zero (there are an equal number of matches made in either direction) and the standard deviation is large. D_{max} lies between these extremes and could sensibly be defined as the displacement for which the standard deviation is equal to the mean of the distribution. In fact, this definition gives very similar values to those obtained using 20% errors to define d_{max} .





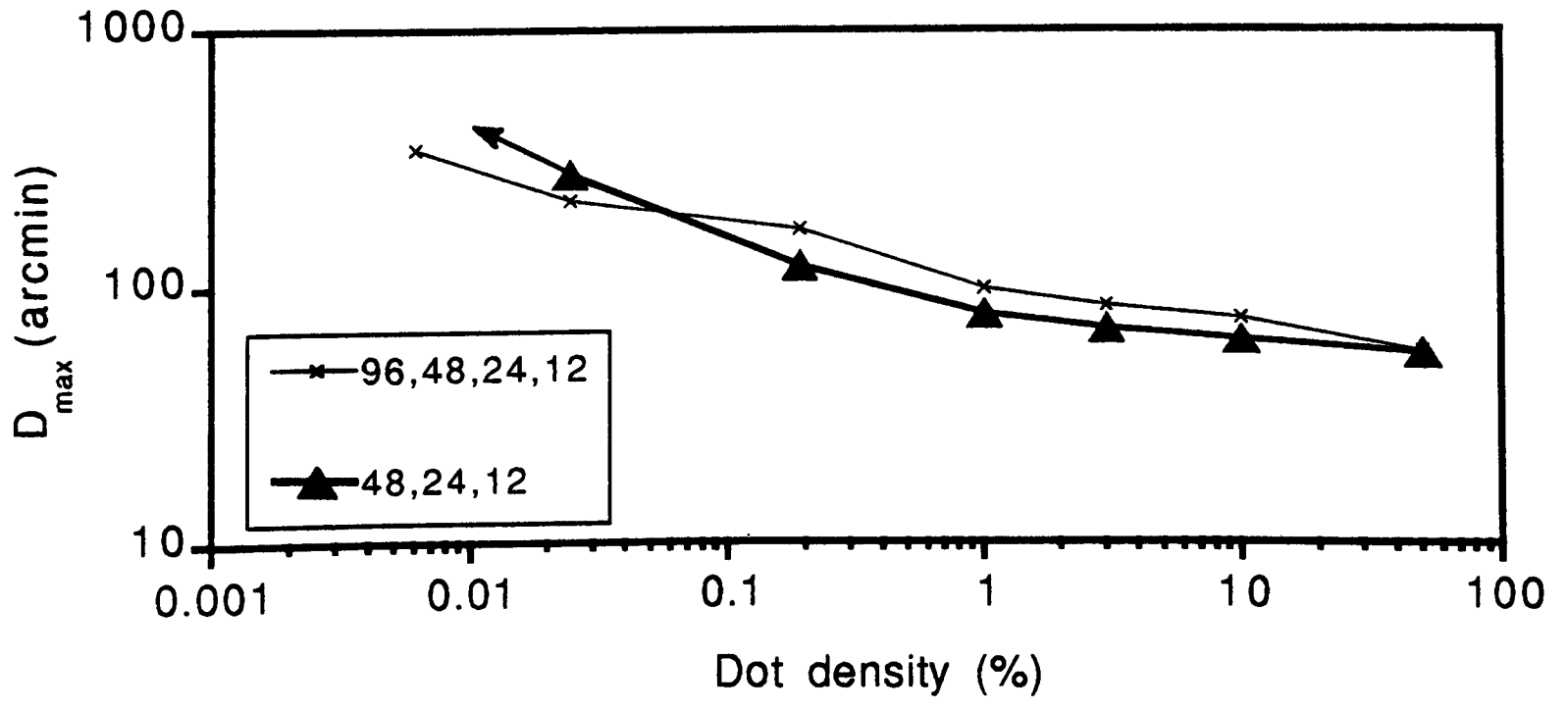


Fig 6.12 (legend on previous page)

matching model, is the degree of co-operativity between centroids on adjacent raster lines in the search for a match. The model described here is only one dimensional but it would seem more likely that 2-dimensional blobs are matched than unconnected 1-dimensional centroids. The most important point about the theoretical d_{max} values shown in figure 6.12 is that the slope is the same as for the data: the model can explain the gradual change in d_{max} at low densities which simple single filter models fail to predict. It can also account for the fact that d_{max} changes, at a similar gradual rate, at high densities (which is true for both black-on-white and white-on-black patterns up to 10% and for white-on-black patterns up to 50%).

The centre plot in figure 6.12 shows the effect running the matching model using a different range of filters to make up the MIRAGE output. The triangles show the same data as discussed above, i.e. MIRAGE with four filters contributing ($\sigma = 96$, 48, 24 and 12 arcmin). The other three lines show the effect of progressively removing the smallest filter (i.e. $\sigma = 96$, 48 and 24; $\sigma = 96$ and 48; and $\sigma = 96$ alone). The effect for a 50% pattern is to increase d_{max} so that it is approximately proportional to the size to the smallest filter remaining. There is no effect of removing the fine filters for the lowest density patterns.

This behaviour can be understood by considering the spatial relationship of the positive responses of fine and coarse filter outputs, as discussed in section 3.3. For very sparse patterns the positive responses of the coarse and fine filters will co-incide spatially: they will be centred on the dots. Thus, removing the fine filter contribution to MIRAGE will have little or no effect on the outline of the blobs or the pattern of centroids. On the other hand, for a high density pattern, the positive responses of the fine and coarse filters do not co-incide spatially. The positive responses of the coarse filter are widely separated but the positive responses of the fine filters *fill in the gaps* between the coarse filter blobs. Thus, for a 50% pattern, progressively removing the smallest filter contribution to MIRAGE increases the spacing between blobs and hence the predicted value of d_{max}.

There is some evidence from experiments using filtered random dot patterns to support this model, but only for 50% density patterns. Cleary and Braddick (1990b) showed that motion-d_{max} for ideal low-pass filtered, 50% random dot patterns rises in inverse proportion to the cut-off frequency of the filter (below some critical frequency, which depends on patch size). This is similar to the behaviour of MIRAGE for 50% density patterns. The strong prediction of the

MIRAGE model is that low-pass filtering would have little or no effect on d_{max} for very sparse random dot patterns.

There is some support for this model in the literature. Cleary and Braddick (1990b) showed that low-pass filtering a 50% random dot kinematogram raised d_{max} and did so in inverse proportion to F_h, the high spatial frequency cut-off of the filter. They concluded from their results that d_{max} was limited by the highest spatial frequency in the pattern. This fits exactly with the behaviour of the MIRAGE model for 50% patterns, shown in figure 6.12. (That is, removing the finest filter output raises d_{max}). But the prediction from the three MIRAGE models shown in figure 6.12 is that low-pass filtering the lowest density patterns would not affect d_{max} at all. In fact, there is a quantitative prediction which could be tested. For any given density, low-pass filtering (at progressively coarser scales) should not affect d_{max} until the size of filter that was limiting d_{max} at that density in the unfiltered case was reached. Such an experiment, if carried out, would provide an important test of the model.

The lower plot in figure 6.12 illustrates the effect of removing the largest filter contributing to the MIRAGE response (σ = 48, 24 and 12 is compared to σ = 96, 48, 24 and 12 arcmin). The effect is to make the shape of the curve slightly more like that of a single smaller filter, i.e. flatter at high densities and rising sharply (theoretically with a slope of -1) at low densities. D_{max} for the lowest density was higher than the maximum displacement tested (60 pixels, equivalent to 360 arcmin). At the highest density, d_{max} for the two filter combinations (σ = 48,24,12 and σ = 96,48,24,12) is the same, i.e. it appears to be limited by the smallest filter contributing to the signal. Taken together the variations on the model shown in figure 6.12 indicate that the spacing of MIRAGE centroids depends primarily on the size of the largest filter for sparse patterns and on size of the finest filter for high density random dot patterns.

6.7.4 Summary of the model

In summary, if d_{max} (either for motion or stereo) is assumed to reflect the spacing of false targets in the image (as discussed in section 6.2) then the spatial primitives derived from MIRAGE predict the data very well. The spacing of primitives changes at about the same rate with changing dot density, whatever the density. This behaviour arises because, as more dots are added to the image each filter reaches it own "plateau" (see figure 6.8) and, as it does so, the next finest filter begins to add new blobs to the image.

In order to model a gradual change in d_{max} from 0.006% to 50% density patterns, as was found experimentally, it must be assumed that the range of filters contributing to MIRAGE includes a coarse filter large enough to blur dots together at the lowest densities and a fine filter small enough to distinguish the highest density patterns.

The model as it stands does not take account of the possible effects of pattern mean luminance or contrast but it could be modified to take account of both. Dawson and Di Lollo (1990) suggest that the effects of adapting luminance levels may be modelled by assuming larger filters are used at low luminances. Low contrast levels could be modelled by adding noise to the signal before the calculation of centroids. The model would remain one dependent on the density of false-targets. In both manipulations the final determinant of modelled d_{max} would be the maximum shift for which the correct direction of displacement could be detected reliably.

6.8 Discussion

The experiment described in this chapter has verified the result obtained by Eagle and Rogers (1991), i.e. that d_{max} for motion rises smoothly as dot density is reduced, and has shown that the same result holds for the equivalent stereo task. It was concluded that, for these stimuli, similar factors limit the correspondence process in both domains. If the assumption is made that d_{max} reflects the density of false targets in the input to the correspondence process, then the results imply that similar spatial primitives are derived from each image before matching takes place, either in stereo or motion. (Not only the slope of the function, but the absolute magnitude of d_{max} is similar for motion and stereo across a wide range of densities. This suggests that the same type of false targets, with the same mean spacing, are limiting the correspondence process in each case.) In the model section an attempt was made to discover a spatial primitive whose density changes as dot density is reduced at the same rate at which d_{max} was found to vary in these experiments. It was shown that 1-D MIRAGE centroids meet this criterion.

6.8.1 Alternative models

It is important to consider alternative explanations for the experimental data. Marr and Poggio's (1979) theory is not a candidate, unless it is substantially modified. There are low spatial frequencies in all the stimuli and, if low spatial frequency

mechanisms solve the correspondence problem successfully for low density patterns, then they should be able to do so for equally large displacements at high densities. Indeed, the amplitude of the output of band-pass mechanisms increases with increasing dot density. Marr and Poggio do not consider "inhibition" or "masking" by high spatial frequency channels and yet something of this kind must be happening, given the results reported in this chapter.

The proposal that high spatial frequencies mask the output of low spatial frequency channels was put forward by Cleary and Braddick (1990b). Their stimuli were low-pass filtered 50% random dot patterns. They studied the effect of varying the spatial frequency of the high frequency cut off (F_h). For any given patch size, low-pass filtering had no effect up to some point (as if the stimulus had already been low-pass filtered, perhaps by larger filters at greater eccentricities) after which d_{max} rose steadily as the stimulus was more and more severely low-pass filtered. In fact, d_{max} was approximately equal to the d_{max} of the highest frequency (or band of frequencies) present, (which they determined separately in Cleary and Braddick (1990a)). As discussed in the previous section, this result is also predicted by a MIRAGE model (see figure 6.12).

The masking proposed by Cleary and Braddick does not constitute a quantitative model and as it stands (the highest spatial frequency channel masks all the others) it would not predict the dot density data presented in this chapter - the highest spatial frequency does not change with dot density. Cleary and Braddick considered that there was a separate motion signal from each spatial frequency tuned "channel" and that interactions would take place after each channel had detected (or failed to detect) the stimulus motion. It is not obvious how the model could be modified to account for the dot density data because the relative amplitude of high and low spatial frequencies is the same whatever the density (the spectra are flat*). Note that, according to a Cleary and Braddick model, at d_{max} for low densities, most of the channels will signal noise (it is a large displacement) and yet the coarse filter signal can "win through". On the other hand, at d_{max} for the high densities, most of the channels give the correct motion signal (it is a small displacement) and yet the finest filter signal "masks" the signal from other channels.

---

* Random dot patterns are a form of white noise pattern (there is no correlation between the luminance values of different pixels) and so they have a flat spectrum (e.g. Voss, 1985). For frequencies higher than (1/pixel size) this is no longer true (see Morgan and Fahle, 1992).

It is worth emphasising the point about the fine filter "masking". It applies not only to Cleary and Braddick's model but to any model that ignores the spatial aspects of the stimulus and considers the stimulus only in the Fourier domain (or, equivalently, considers only activity within "channels"). The "masking" effect of high frequencies in the MIRAGE model depends crucially on the spatial position of the fine scale filter output in relation to the coarse filter output. When positive response of the fine filters lies entirely within the region of positive response from the coarse filter, as is the case for low density patterns, the fine filter outputs have no effect on d_{max}: they do not alter the spatial structure of the MIRAGE response. However, for high density patterns there are positive responses from the fine filters right across the image and these have the effect of "filling in" the gaps between the coarse filter blobs. Hence, in high density patterns, fine filter outputs *do* affect d_{max}. A "channel model" must rely on a completely different type of explanation, unless the channels are very local.

A recent paper by Morgan and Fahle (1992) attempts to explain the effect of dot density on d_{max} using a single filter model. (They investigated the effects of dot density over a comparatively small range of densities, from 5 to 50%.) Unlike Cleary and Braddick, Morgan and Fahle suppose that d_{max} for the "channel" is limited by the density of false targets. Their results do not follow the predictions of a single filter (like those shown in figure 6.8), and they suggest the explanation is that, for small dot sizes and a pattern of black dots on a white background, a single channel will receive only a low contrast signal for low density patterns. They describe the interaction of density and pattern contrast as follows:

> *"Reducing the density of pattern elements reduces the size and thus the contrast energy of aggregated elements in the pattern. Because performance with broad-band patterns is contrast limited over a wide range, density causes the decline in performance because of its effects upon pattern energy, and this effect counteracts its effect upon element separation."*
>
> (Morgan and Fahle, 1992, p197)

Morgan and Fahle's model, although it is compatible with the small fall in d_{max} over the range of densities from 50 to 5%, would not account for the rise in motion-d_{max} found by Eagle (1992) for dot densities between 5 and 0.0125% (black dots on white). The slope of the data in this range is very similar to that found for white dots on black (Eagle and Rogers, 1992 and motion data in experiment I of this chapter). Also, it would not account for the very similar results obtained in experiment I and II (white dots on black versus black dots on white) since, as

discussed above, dot density has the opposite effect on stimulus contrast (as defined by Peli and Goldstein, 1988) in the two conditions.

Morgan and Fahle are not explicit in describing how reduced contrast might "counteract" the effect of large element separation but one possibility, as discussed above, is that at low contrasts (low signal-to-noise ratios) spurious false targets may be introduced. Thus, it may be possible to combine the effects of contrast and dot density in one false-targets model considering the spacing of elements arising both from the stimulus and from noise.

Finally, the predictions of a co-operative algorithm should be considered, that is, what is the predicted maximum detectable disparity and how would it depend on dot density? The answer is not obvious. The main limitation on large disparities is the disparity gradient limit. The disparity gradient in the stimuli used in this experiment is infinitely large at the edges of the stereogram (a step edge) and zero across the patch. Provided the search covers a sufficiently wide range of disparities (the "PMF" algorithm searches over a range of 20° (Frisby and Pollard, 1991) which is certainly sufficient), the stereogram should be matched successfully. Higher dot densities might be expected to help rather than hinder the matching process since each dot provides extra support for the correct disparity match. In other words, the results of the experiment described in this chapter do not appear to fit the predictions of a co-operative ("local-to-global") algorithm.

6.8.2 Other measures of an upper disparity limit

How does "stereo-d_{max}" relate to other measures of the limits of stereo matching at large disparities? There is no general agreement on a measure for the largest disparity that can be detected by the visual system. The predictions that Marr and Poggio (1979) made illustrate this confusion. They refer to the disparity at which diplopia occurs, the disparity above which depth discriminations cannot be made and the maximum disparity that can initiate a vergence eye movement and seem to imply at certain points (see sections 1.5.5 and 1.5.6) that the same limits apply to all of these measurements.

In fact, as discussed in chapter 1, these criteria are very different and yield markedly different estimates of an "upper disparity limit". For example, the results of Schor, Wood and Ogawa (1984) suggest that diplopia reflects the highest spatial frequency in the stimulus whereas the maximum detectable disparity is, in general, likely to reflect the low spatial frequency content of the stimulus. This is true of the

"upper depth limit" (yet another criterion) which Schor et al. (1984) define as the maximum disparity for which the stimulus is seen to have any depth (with respect to the fixation plane). Both these criteria are subjective and may display hysteresis (as described for diplopia by Fender and Julesz, 1967)

A more objective measure of the upper disparity limit is the maximum disparity that will drive a vergence eye movement directly to the correct depth (i.e. without "hunting" vergence movements). This was done indirectly by Frisby and Mayhew (1980) and directly by Mowforth, Mayhew and Frisby (1981). These experiments have been discussed in section 1.5 in connection with Marr and Poggio's theory and in section 6.1.2. They are the stereo equivalent of the two-frame apparent motion experiments using filtered 50% random dot patterns (Chang and Julesz, 1983; Cleary and Braddick, 1990a; Bischof and Di Lollo, 1990; Eagle and Rogers, 1992). The results show that larger disparities can be matched (i.e. initiate a vergence eye movement) for low spatial frequency filtered patterns than for high. In this respect the results are qualitatively in agreement with the apparent motion experiments. However, there is a large discrepancy in the quantitative results. Mowforth et al. (1981) claim that vergence movements can be initiated by disparities of just over three cycles of the stimulus centre frequency. For an equivalent task in motion, d_{max} is, in all the above experiments, about 1 cycle of the stimulus centre frequency. Given the close similarity between results for motion- and stereo-d_{max} found in the experiment described in this chapter, the discrepancy is surprising. One possibility is that vergence eye movements can be initiated by stimuli for much larger disparities than stereo-d_{max}, but this seems unlikely. It would be interesting to carry out an experiment in which the maximum disparity that will drive a vergence eye movement to the correct depth could be compared directly with stereo-d_{max}.

6.9 Summary

The aim of the experiment described in this chapter was to explore the nature of spatial primitives used as input to the stereo correspondence process. The experiment measured the upper disparity limit for reliable detection of a random dot pattern as in front of or behind the screen ("stereo-d_{max}") given only a brief exposure duration (150 ms). This measurement was assumed to reflect the spacing of false targets (i.e. spatial primitives) in the stimulus. The evidence for this

assumption comes from a variety of experiments using 2-frame apparent motion sequences (discussed in section 6.2).

Stereo-d_{max} was measured for random dot patterns at a range of densities from 50% down to only a few dots both for patterns of white dots on a black background and for black dots on white. D_{max} values increased as dot density was reduced in both conditions and this was assumed to reflect the wider spacing of spatial primitives at lower dot densities. The pattern of results suggests that coarse spatial filter outputs dominate the response (i.e. the input to the correspondence process) at low densities and that fine filter outputs determine primitive spacing at high densities. The characteristics of the MIRAGE response fit this description. A model based on the horizontal separation of MIRAGE centroids was shown to match the experimental data well.

In addition, motion-d_{max} was measured for a 2-frame apparent motion task using the same stimuli as for the stereo task. (The left and right eyes' images were displayed as a 2-frame apparent motion sequence.) The results were almost identical to those for the stereo experiment. It was concluded that the same primitives may be used in both the stereo and motion correspondence processes.

CHAPTER 7

7.1 Review

The central idea in this thesis is that the position of a feature in the image might be defined in relation to other features in the image (Watt, 1988) and that binocular disparity might also be encoded in this way. The justification for a relative, "hierarchical" representation of position was considered in detail in chapter 2; some of the consequences for a theory of stereopsis were explored in chapter 3; and some evidence compatible with this as a model of stereo matching in the human visual system was presented in chapters 4, 5 and 6.

7.1.1 Time course

In the first experiment (shape discrimination in random dot stereograms), a stimulus was used in which, at a coarse scale, the left and right eyes' images were very similar, i.e. at this scale there was only very poor information about the target shape. At a fine scale, however, there was precise information about the target shape. The results from this experiment fit less well with a local-to-global theory, or of *any* stereo algorithm that starts from fine scale information (e.g. Marr and Poggio, 1976; Pollard et al.., 1985). On the other hand, subjects' responses at short exposure durations can be successfully modelled by considering the (limited) information present in coarse-filtered versions of the stimuli.

7.1.2 A hierarchical database

In the second experiment (comparison of the 2-D and 3-D Muller-Lyer illusions), a stimulus was used in which left and right eyes' images *differed* at a coarse scale but which, with respect to the task subjects were asked to make (the length or slant of the shaft), were the same at a fine scale. The results of this experiment suggest that there is no "direct access" to the fine scale position of features and this fits with the idea that fine scale position is available only in relation to the position of coarse scale features. In addition, the results of this experiment suggest that large scale judgements, in general, are made on the basis of coarse scale information, and small scale judgements are made using fine scale information. In other words it is suggested that without extensive practice, judgements that require a combination of information measured at several scales are not made accurately (they are prone to bias) even with unlimited exposure duration. Finally, the evidence from this experiment supports the idea that the same system for representing position is used for making 2-D and 3-D judgements.

7.1.3 Spatial primitives

The third experiment (a comparison of stereo- and motion-d_{max}), explored the limits of the correspondence process when the stimuli contained large disparities or displacements. Random dot patterns of different densities were used, from 50% down to just two dots. Three factors change with dot density, all of which might have an effect on d_{max}: the mean luminance of the pattern, stimulus contrast and element spacing. The results suggested that element spacing was the principal determinant of d_{max}, for both stereo and motion. This fits with the conclusion derived from other studies using two-frame apparent motion stimuli.

The results do not fit well with the predictions of a single filter model and suggest that the primitives used in the stereo correspondence process are derived from the outputs of a range of different sized filters. A version of the spatial primitives proposed by Watt and Morgan (1985), which are formed from the combined outputs of several filters, model the data well.

The same pattern of data as was found for the "large disparity" stereo task was also found for an equivalent "large displacement" motion task. The results suggest that the same spatial primitives are used in the motion correspondence process and the same factors govern the establishment of correspondences.

7.1.4 MIRAGE

All three experiments are relevant to the predictions of the MIRAGE algorithm (Watt and Morgan, 1985; Watt, 1988). The results of the first experiment were modelled by assuming stereo matching was a coarse-to-fine process. In the MIRAGE algorithm the spatial structure of the image is analysed from coarse to fine spatial scales (for reasons discussed in chapter 2). The data from this experiment were insufficient to plot a precise time course of the coarse-to-fine process, but the results were compatible with the time course deduced by Watt (1987) in a different context. The results from the second experiment were consistent with a model in which information about each eye's image is stored in a hierarchical database as proposed in the MIRAGE algorithm. The results from the third experiment were modelled by assuming that the spatial primitives used for matching in the correspondence process were MIRAGE centroids.

Even if alternative explanations are found for all these results, an important theoretical point remains: A hierarchical image description, which is the principal objective of the MIRAGE algorithm, is an efficient basis on which to encode disparity.

7.1.5 In defence of a hierarchical model

The theory put forward in chapter 3 was a hierarchical scale based model, that is, it proposed that the position and hence disparity of fine scale features is determined with respect to the position and slant of the coarse scale group of which they form a part. Much of the evidence presented in this thesis concerns related issues, such as the time course of stereo processing of the combination of filter outputs, and does not directly address the hypothesis that the matching and perceived depth of fine scale features is determined relative to coarse scale features. Part of the reason for this is that many of the relevant experiments have already been carried out and have reported results consistent with a hierarchical model, although they have not been described in that way before. In chapter 5 the results of Mitchison and McKee (1987a and b) and Mitchison and Westheimer (1984) were discussed in detail to illustrate this point. The experiment described in chapter 5 can be seen as an extension of Mitchison and McKee's findings which relates their results to a filter-based, hierarchical model rather than one based on ambiguous and unambiguous matches.

The experiments in chapter 4 and 6 are also relevant to a hierarchical model, not just to any coarse-to-fine theory. For instance, the stimuli in chapter 4 in which the target was defined by a "+1,-1" disparity were reported by subjects to appear

different (e.g. "less solid") than other patterns for long exposures, although thresholds for the height-to-width task were as low as for other stimuli at an exposure duration of 1 second. Why should this be so? In a model such as that of Nishihara (1984) or Quam (1984) where the coarse signal is used to guide the matching of fine scale features there is no reason to regard the matches close to zero disparity as any less stable than those at other disparities: the fine scale representation can "stand alone" since it is not dependent on the coarse scale signal. In a hierarchical model this is not the case. The role of fine scale disparities is to provide detail to fill in the outline discovered at a coarse scale. The "+1,-1" target is special in consisting of a relatively large shape which must be "built up" from fine scale information. The suggestion put forward in chapter 4 was that the visual system may be poor at doing this. If this is correct, then the subjective appearance of the "+1,-1" patterns fits better with a hierarchical model than with other coarse-then-fine models.

The experimental data in chapter 6 also fits better with a hierarchical model than with other coarse-to-fine models, such as Marr and Poggio (1979). All the random dot patterns contain low spatial frequencies and yet d_{max} is very different for different dot densities. The data appear to suggest that high spatial frequencies have a greater masking effect (reducing d_{max}) for high density random dot patterns than for low densities. This is exactly the prediction of a MIRAGE model but not of a coarse-to-fine model that assumes independent access to the output of each filter (in particular the coarse filter).

The way in which filter outputs are combined in MIRAGE is relevant to a hierarchical model because it provides a way of grouping features. This issue was discussed in detail in chapter 2. It has often been noted (e.g. Prazdny, 1987) that an important consideration in any coarse-to-fine model is the way in which fine scale features are related to those at a coarse scale. In principle, features such as zero-crossings can be tracked from coarse to fine scales (e.g. Witkin, 1988) but, as Blake and Zisserman (1987) point out in the context of robotics applications, in practice sampling of scale space at a relatively small number of spatial scales makes this a very much more difficult task (e.g. their figure 4.11). The problem is avoided in MIRAGE since the "parenthood" of each fine scale blob is established in the initial, coarsest scale, representation but there is a potential cost. It means that for some stimuli, of which a 50% random dot pattern is a good example, the coarse scale representation can be "crowded" with only small gaps between spatial primitives. The results of the experiment described in chapter 6 suggest that the visual system may suffer in just the way predicted by the MIRAGE algorithm.

7.2 Related work

In this final section the results of other experiments on stereopsis are discussed in the context of a hierarchical scale-space model.

7.2.1 Physiological evidence on the correspondence process

A paper by Poggio, Motter, Squatrito and Trotter (1985) provides important evidence that neurons in striate (V1) and prestriate (V2) areas of the cerebral cortex can respond selectively to disparity-defined targets in random dot stereograms. They recorded from neurons in alert macaque monkeys who were trained to fixate while random dot stereograms were displayed in the region of the neuron's receptive field. The stimuli consisted of dynamic random dot patterns (with a typical density of 10% bright dots). A new stereogram, but with the same pattern of disparities, was displayed 100 times a second. The disparate area (i.e. the test figure) within the stereogram was defined in one of two ways: either it consisted of dots of the same density as the background but with an added disparity (a "cyclopean stimulus") or it consisted of dots which were all bright or all dark (a "solid figure stimulus"), again with an added disparity.

The two types of pattern differ in several respects. In the solid figure stimuli the shape of the test figure is visible monocularly and it does not change with time whereas the background is changing 100 times a second. The test figure in the cyclopean stimulus is invisible monocularly and the whole pattern has the same temporal frequency. It is possible to measure the response of a neuron to all disparities for the solid test figure but presumably it is impossible to test the sensitivity of a neuron to zero disparity for a cyclopean stimulus, since both the target and the background are made up of correlated dots (see their figure 1).

Poggio et al.. analysed the response of 244 neurons, all of which responded to the solid figure stimuli. A subset of these (50) also responded to the cyclopean stimuli. There were a number of differences between the responses of cells to cyclopean stimulation and their response to the solid figures. Some of the properties are interesting in relation to the results reported in this thesis. In particular, the analysis in chapter 5 of Papert's cyclopean demonstration of the Müller-Lyer illusion is relevant. There it was shown that an alternative interpretation of a "cyclopean effect" could be given if matching were assumed to take place between coarse monocular features. Poggio et al. report that the response of neurons often varied

systematically with the orientation of the solid figure (they showed orientation tuning) but for the cyclopean figures orientation tuning was reduced or absent. This pattern of response is what would be expected if the neurons responded selectively to the orientation of monocular features in each eye's image (and, in addition, were selective for a particular disparity). This possibility is supported by the finding of Poggio et al. that, above a certain minimum size, some neurons were equally responsive to a cyclopean stimulus whatever its size while the same neuron might respond very selectively to a narrow range of sizes for the solid figure stimulus (their figure 5 and 6). Poggio et al. do not seem to consider the possibility that it may be the size or width of features in the monocular (filtered) image that is important in determining the neuron's response.

Poggio et al. describe the detection of the cyclopean figures as the outcome of a "global" stereoscopic process and the detection of the solid figures as "local", following the terminology used by Julesz (e.g. Julesz, 1971). This distinction was not adopted by Marr and Poggio (1979) because, as they pointed out, random dot stereograms could be solved using a coarse-to-fine algorithm without the need for a co-operative process. A similar assumption has been adopted in the modelling described in this thesis. The stimuli used in the experiment described in chapter 6 ranged from very low densities (suitable stimuli for a "local" mechanism) to 50% density random dot patterns (which Julesz considered must be solved by a "global" stereoscopic process) and yet a single mechanism, based on a nearest neighbour matching rule, was proposed to model the whole range of densities. A prediction arising from this experiment is that the cyclopean stimuli that Poggio et al. used (10% density patterns) would be detectable (by human subjects or monkeys in a 2-alternative forced choice task, or by disparity selective neurons) over a smaller range of disparities than the solid figure stimuli (for which it could be argued the matching problem is similar to a very low density pattern). Poggio et al.. provide some information about the disparity selectivity of the different types of neurons but do not give data that would address this prediction directly.

7.2.2 Before or after correspondence?

In chapter 1, binocular stereopsis was described as being divided into two stages: a correspondence process and a stage of interpreting disparities. The borderline between these two stages is not as clear as it might at first appear. In chapter 5, for instance, it was pointed out that the perceived slant of the Müller-Lyer shaft could be interpreted either in terms of matching of coarse scale (monocular) images (i.e. before correspondence) or in terms of blurring in the disparity domain, after correspondence had been achieved. The same ambiguity was evident in

discussions of disparity averaging (chapter 4). In the following examples, several phenomena that have traditionally been considered in terms of interactions taking place *after* the measurement of disparity, are discussed within the framework of a hierarchical model. The purpose is not to provide solutions to any of the problems but rather to illustrate a possible approach.

7.2.3 Anisotropy

An intriguing and unexplained aspect of stereopsis is the anisotropy of slant perception (e.g. Wallach and Bacon, 1976; Rogers and Graham, 1983; Graham and Rogers, 1982; Mitchison and McKee, 1990). Surfaces slanting or curved about a vertical axis ("wall" surfaces) are much more difficult to see as slanted (thresholds are higher and suprathreshold matching reduced) when compared to surfaces oriented about a horizontal axis ("sky/ground" surfaces). Rogers and Graham (1983) showed that the cause of the effect is related to the type of transformation (expansion versus shear) rather than to the orientation of the surface per se (vertical versus horizontal). (They demonstrated that the same anisotropy occurred for surfaces defined by motion parallax and, when head movements were vertical rather than horizontal, the anisotropy reversed.) The question, then, is why an expansion/compression pattern of disparities is more difficult to detect than a shear.

In chapter 3 it was pointed out that the curvature of a set of points in a horizontal direction (i.e. the rate of change of expansion or compression) is obscured at a coarse scale while curvature in a vertical direction (rate of change of shear) can be detected at any scale*. If curvature disparity is an important signal to surface shape (Rogers and Cagenello, 1989) then this may help explain some of the effect, but only in experiments on curved surfaces (e.g. Rogers and Graham, 1983). (For planar surfaces, this explanation will not suffice: in a hierarchical model width disparity should signal the slant of a surface about a vertical axis just as orientation disparity should signal the slant about a horizontal axis. Rogers and Graham put forward an *ad hoc* hypothesis that orientation disparity is used but not width disparity.)

---

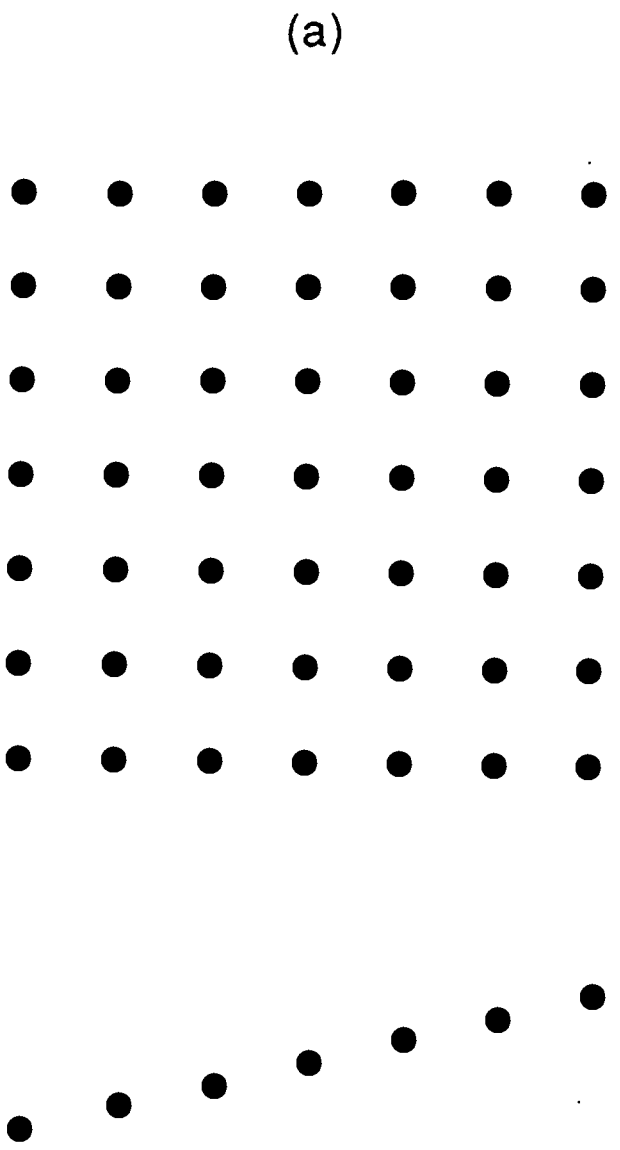
* Note that it is not just that curvature in a horizontal direction cannot be signalled at a coarse scale, it cannot be signalled explicitly at *any* scale. Of course, the curvature can be deduced from the orientation of smaller scale blobs but it is possible that the visual system is as bad at "building up" curvature as it is at "building up" the shape of an object from fine scale blobs (as discussed in section 4.9).

One result, from an experiment using *planar* surfaces, is worth mentioning in the context of a hierarchical matching scheme. Mitchison and Westheimer (1984) showed that the threshold for detecting a disparity between two vertical columns of dots was influenced to a great extent by the presence or absence of several flanking columns of dots, i.e. whether they were on their own or part of a grid of dots to which a slant had been applied (see figure 7.1). As the results in figure 7.1 show, a much larger slant had to be added to the grid before it could be seen as slanted than for the two columns alone, even when the slant is expressed as a disparity difference between adjacent columns. Mitchison and Westheimer went on to show that the sensitivity for detecting deviations from the plane of the grid was just as high as when the grid was fronto-parallel despite the marked insensitivity to the slant of the grid itself. It is as if, in the terms used by Mitchison and McKee (1987a), an interpolated plane had been drawn through the surface and the disparity of features relative to this plane known with great precision while the slant of the interpolated plane itself was not known. In terms of a hierarchical scheme this is equivalent to saying that the position of fine scale features is recorded in terms of the width of the coarse scale blob (implying that this is measured at some stage) and that as a result any difference in the recorded position of a blob in the two eye's images are automatically disparities *with respect to the slant of the coarse scale blob*. This was illustrated in figure 3.2. What this does not explain, and it remains a mystery, is why, under some circumstances, the width difference of the coarse scale blobs is apparently so underestimated. (The anisotropy shows marked individual differences as documented by Rogers and Cagenello, 1990; Mitchison and McKee, 1990.)

7.2.4 Simultaneous depth contrast

Simultaneous contrast effects occur in the depth domain as they do in other modalities such as luminance (Graham and Rogers, 1982; Westheimer, 1986; Westheimer and Levi, 1987). Two examples are shown in figure 7.2 taken from Graham and Rogers (1982).

If the example shown in figure 7.2 (a) is presented in the depth domain then, as Graham and Rogers have illustrated in figure 7.2, the two bars with the same



G.W	83	14.4
M.G.	140	13.5
R.Y	160	6.9

Fig 7.1

Slant thresholds for a square grid or pair of columns, taken from an experiment by Mitchison and Westheimer (1984). The stimuli used were a 7 x 7 lattice of dots spaced 8 arcmin apart horizontally and vertically (a) and a pair of columns from this lattice (b). In each case the stereoscopically viewed stimulus was slanted about a vertical axis (one eye's view was horizontally expanded, the other compressed). The thresholds for reliable discrimination of the direction of slant are shown below for three observers (in arcsec). These are given in terms of the disparity between adjacent columns at threshold, i.e. the total disparity between the left and right hand edges of the lattice (a) was six times the values shown here.

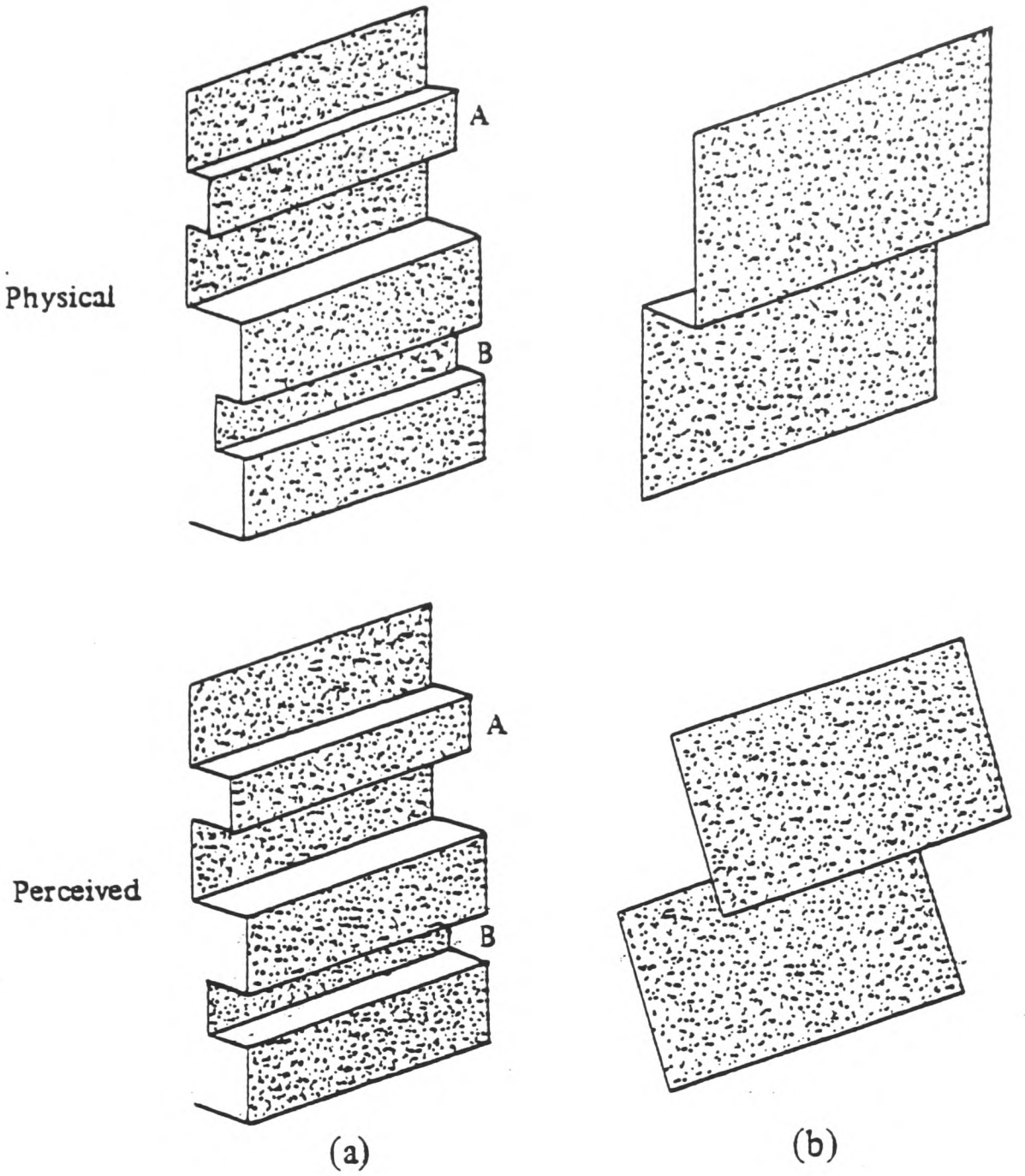


Fig 7.2

This illustration is taken from Graham and Rogers (1982). It shows two examples of simultaneous depth contrast in stereoscopically defined surfaces. In (a) two horizontal bars, A and B, had the same disparity but A appeared in front of B. In (b) two fronto-parallel surfaces were shown with a step change in disparity between them (the upper surface was closer to the observer). As illustrated below, both surfaces appeared to observers to slope away in depth.

disparity (A and B) appear at different depths. The same simultaneous contrast effect can be observed for an analogous stimulus presented in the luminance domain (i.e. a grey bar on a dark background appears lighter than the same grey on a light background). The same is not true for the step edge (figure 7.2 (b)). The perception of this stimulus displays a simultaneous contrast effect in the depth domain, as illustrated in figure 7.2 (b), but is seen veridically in the luminance domain. Graham and Rogers (1982) link simultaneous contrast effects in both domains to the output of centre-surround mechanisms (either for luminance or disparity) but the comparison between processing in the two domains is not a straight forward one.

Supposing that centre-surround mechanisms do exist in the disparity domain, an assumption must be made about how their output is interpreted. The output of a centre-surround mechanism is a twice differentiated signal, i.e. these cells would indicate the disparity *curvature* (Rogers and Cagenello, 1989). If the output of these cells is interpreted as such then the shape and contrast of an edge should be seen veridically. (The same argument applies to the interpretation of filter outputs in the luminance domain.).

The alternative approach, taken, for example, by Mitchison and Westheimer (1984 and 1990) is to assume that the output of the centre-surround mechanisms is *not* interpreted as disparity curvature but as a signal of the depth of the object (albeit distorted), in their terms its "salience". This does account for simultaneous contrast, although it raises other questions (e.g. why are Mach bands not seen at every edge in stereo?).

A hierarchical model of contrast would be rather different (since there is no disparity "map" which can subsequently be convolved with a centre-surround mechanism). A simple example to consider is the stimulus shown in figure 7.2 (b) from Graham and Rogers, 1982. The stimulus is a step edge in disparity in which the top half of the display has a crossed disparity. The planes on either side of the edge are in fact fronto-parallel but, as illustrated in the figure, subjects tend to see them as slanted ("ground planes").

In figure 7.3 the left and right eye's views of a coarse blob that crosses the step edge are shown. The blobs have an orientation disparity and this signals explicitly the coarse scale slant. Most of the fine scale blobs lie on one of the two fronto-parallel planes (one crosses the step edge). Those lying entirely on one plane have no orientation disparity when considered in retinal co-ordinates. But if in a

hierarchical scheme the orientation of the fine scale blobs is signalled with respect to the orientation of the coarse scale blob (as shown in figure 7.3 (b)), then they *would* have a (relative) orientation disparity.

One further step is required to explain the perceived slant. The orientation disparity of the fine scale blobs is equal and opposite to that of the coarse scale blobs. By adding these two disparities together the true (retinal) orientation disparity of the fine scale blobs can be recovered (i.e. zero). However, as discussed in chapter 5, information recorded at different scales may not be combined very efficiently in the visual system. If this is the case, then a fine scale estimate of the slant of the surfaces would follow the pattern found by Graham and Rogers (1982). It would also suggest that simultaneous contrast should be weak at short exposures and increase for longer exposures. (There is, at present, no evidence on this prediction.)

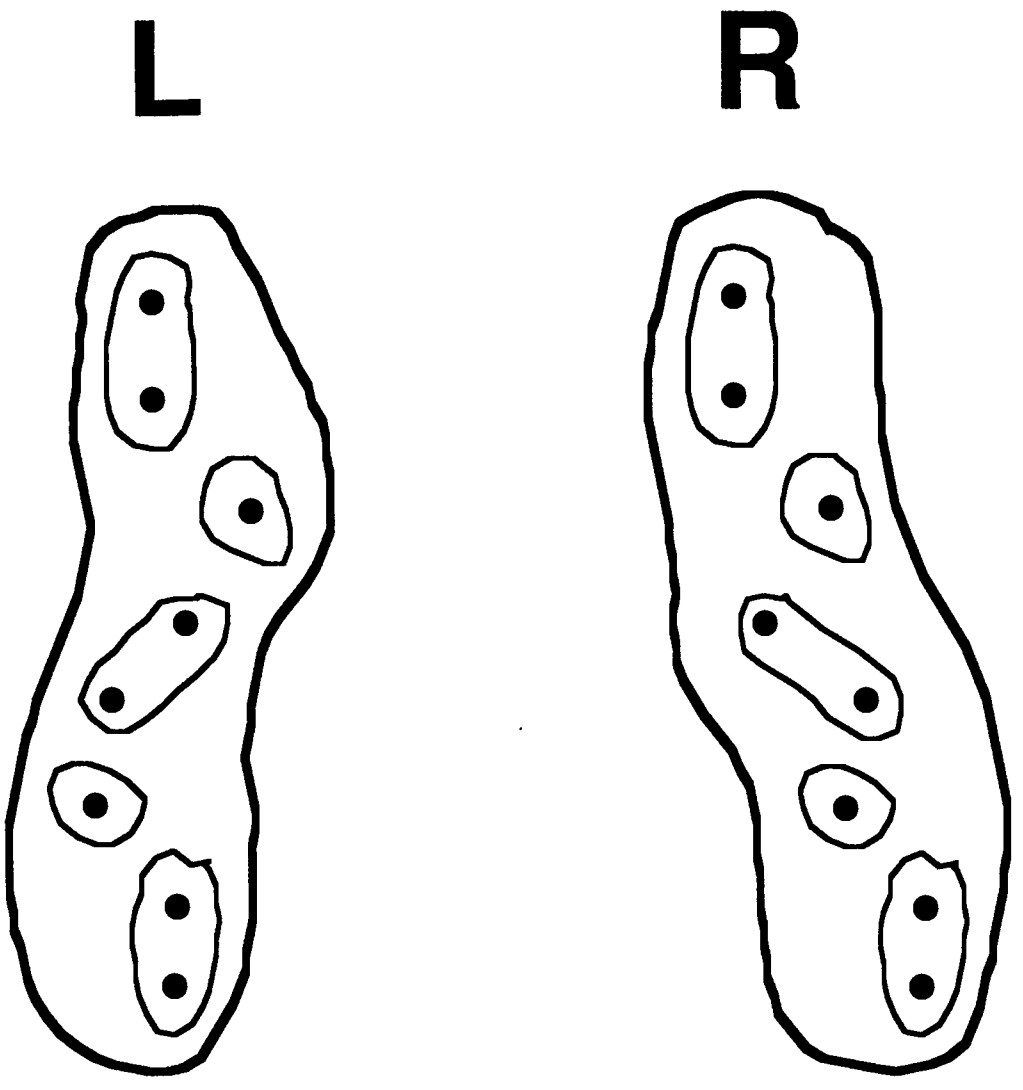
7.2.6 Transparency

As has been pointed out previously (section 6.1), transparent stimuli pose a real problem for coarse-to-fine algorithms and a hierarchical model is no exception. This is because coarse scale matches can only guide fine scale matching if coarse and fine scale disparities are similar. As Prazdny (1987) says, the assumption behind a coarse-to-fine strategy is that coarse scale information is a reasonable summary of fine scale information and for some transparent stimuli this is not the case. The stimuli used in chapter 6 are one example. For high density patterns used in that experiment stereo matching failed at relatively small disparities. (These stimuli are transparent in the sense that Prazdny described, of looking between fence posts at a patch of grass in the background.)

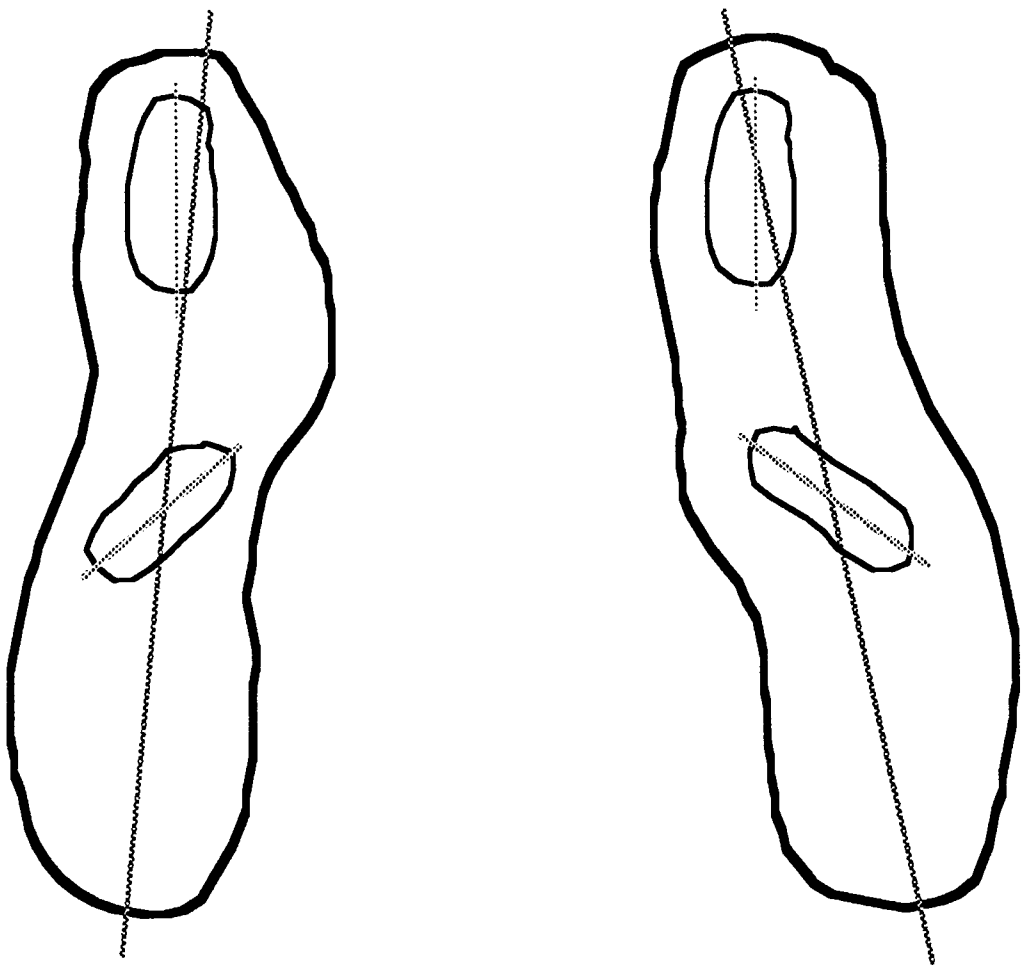
A more common form of transparent stimulus is one containing two planes, each defined by a texture of low density random dots, at different disparities. One

---

Fig 7.3
The left and right eye's images of a coarse scale blob are shown in (a). The blob crosses a horizontal step edge like that illustrated in figure 7.2 (b). (The upper four dots lie on a fronto-parallel surface nearer to the observer than the lower four dots.) Fine scale groupings of the dots are also shown (one fine scale blob crosses the step edge). In (b) the principle axis of the coarse scale blobs and two of the fine scale blobs are shown. In a hierarchical scheme the orientation of fine scale blobs is likely to be recorded with respect to the coarse scale orientation.



(a)



(b)

Fig 7.3 (legend on previous page)

surface can be seen through the other. A hierarchical model has no convincing explanation of this phenomenon. As described in section 2.3, the grouping principle used in a hierarchical model is proximity which is not appropriate for transparent surfaces. A better grouping principle in this case might be disparity or even blur. It is possible that the *default* assumption is that surfaces are opaque. When this fails to yield a coherent surface a different grouping principle might be used.

7.2.7 Vertical disparities

A hierarchical model of stereopsis emphasises explicit information about each blob: its disparity relative to other blobs at the same scale; the slant, and perhaps the curvature of the blob. In chapter 3 it was claimed that the slant of a blob about a vertical axis could be derived from the different width of the blob in the two eyes' images. But there is a complication in calculating the slant of blob in this way. The width of a blob can be affected not only by the surface slant but also by the differential size of the blob in the two eye's images. This can be different if one eye is closer to the feature than the other (which occurs in its most extreme form when an object is close to the observer and eccentric with respect to the straight ahead (i.e. off the median plane)).

Is it possible to disentangle these two effects and derive the slant of the blob (with respect to the fronto-parallel)? Mayhew and Longuet-Higgins (1982) have shown that it is possible to derive the slant of a surface by using vertical size differences in the two eyes' images. Below, a rather different approach is described which uses information only about the particular blob whose slant is to be determined.

The ratio of sizes of the feature in the two eyes is simply related to the ratio of the distances from that feature to each eye, at least for vertical size. If the feature were a sphere, which projects a circular image from all vantage points, then horizontal size differences would give exactly the same information. However, for a planar feature, such as a disc, the slant of the surface (about a vertical axis) affects horizontal size differences as well. The two sources of horizontal size difference can be separated.

Figure 7.4 (a) illustrates the Vieth-Muller circle. A blob lying on a this circle will subtend an equal angle at the left and right eyes (by definition) i.e. its horizontal size will be equal. In other words, if the difference in horizontal extent in the two eyes is used to measure slant, then slant is defined relative to the Vieth-Muller circle. That is, in order to interpret the slant of the blob as a slant with respect to

the fronto-parallel the fixation distance and the eccentricity of the blob (with respect to the head) must be known.

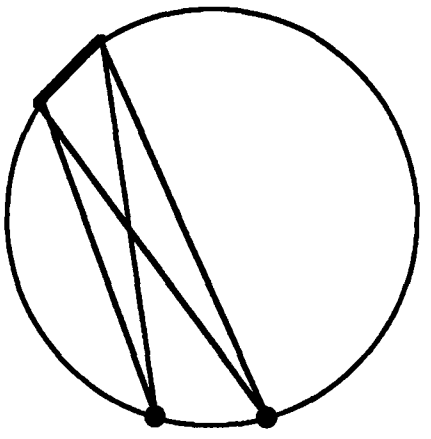
Figure 7.4 (b) illustrates a cross section through the "cyclopean sphere" in the plane of regard, i.e. a circle centred on the cyclopean point. A blob lying on this surface is equally oblique with respect to either eye. Therefore, any size difference signals the relative distance of the blob from each eye, i.e. horizontal and vertical sizes are equally affected. This means that if an isotropic expansion or

---

Fig 7.4

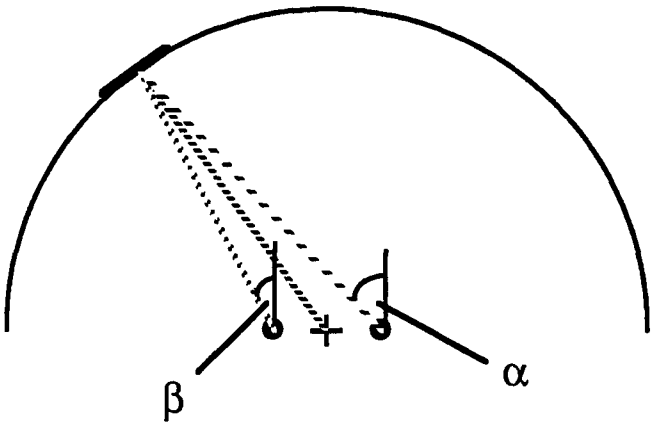
This figure illustrates possible ways to define the slant of a blob.
A plan view is shown in (a) of the two eyes and a small blob (shown in bold) which lies at a tangent to the Vieth-Muller circle. In this case, by definition, the horizontal width of a blob in the left eye (H_L) is equal to the width in the right eye (H_R). In other words, measuring the difference in width of a blob in the left and right eye's images gives the slant of the blob with respect to the Vieth-Muller circle.
The width of a blob in the two eye's images might be compared after each had been "adjusted" to take account of the expansion of the blob in one eye's image. The plan view in (b) shows the arc of the circle centred on the cyclopean point mid-way between the eyes (indicated by a cross). The distance from the left eye to the blob is proportional to $1/(\cos \beta)$ (where β is the angle made with the straight ahead) so the horizontal width of the blob is proportional to $\cos \beta$. The same argument applies to the vertical height of the blob. The distance of the blob from the two eyes is the only factor affecting its horizontal size since its slant with respect to the line of sight from the left and right eye is equal (and opposite). In other words, measuring the difference in width of a blob in the left and right eye's images after an adjustment has been made for the expansion of the blob in one eye's image (i.e. using the ratio V_L/V_R) gives the slant of the blob with respect to the "cyclopean circle".
The slant of the blob with respect to the fronto-parallel could also be calculated if a more complex adjustment was made. In (c) a plan view of a blob in the fronto-parallel plane is shown. The distance to the blob from the two eyes affects the horizontal and vertical size ratios as described in (b). In addition, the horizontal size ratio is determined by the relative obliqueness of the surface. The width of the blob, perpendicular to the line of sight from the left eye, is proportional to $\cos \beta$ (shown in inset). The angle that this perpendicular ("virtual blob") subtends at the eye is, because of its distance from the eye, itself proportional to $\cos \beta$. These two factors together mean that the width of the blob in the left eye is proportional to $\cos^2 \beta$. The same argument applies to the right eye. The vertical height ratio is, as in (b), $\cos \beta / \cos \alpha$. In other words, measuring the difference in width of a blob in the left and right eye's images after an adjustment has been made for the expansion and extra obliqueness of the blob in one eye's image (using the ratio $(V_L/V_R)^2$) gives the slant of the blob with respect to the fronto-parallel.

$$\frac{H_L}{H_R} = 1$$



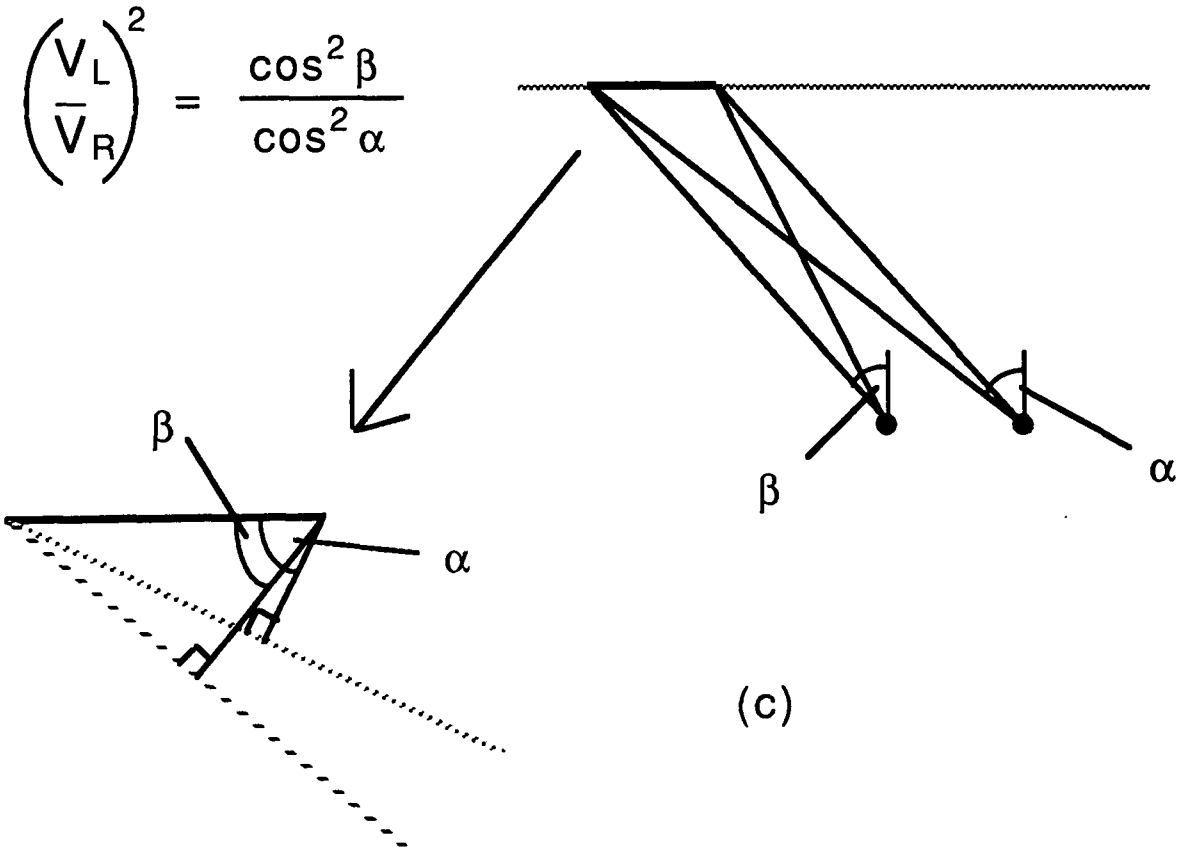
(a)

$$\frac{H_L}{H_R} = \frac{V_L}{V_R} = \frac{\cos\beta}{\cos\alpha}$$



(b)

$$\frac{H_L}{H_R} = \left(\frac{V_L}{V_R}\right)^2 = \frac{\cos^2\beta}{\cos^2\alpha}$$



(c)

Fig 7.4 (legend on previous page)

compression were applied to the blob in one eye's image sufficient to make the vertical sizes of the blob equal in both eye's, then any "residual" horizontal size difference would signal the slant with respect to the cyclopean sphere. In this case, in order to interpret the slant of blob as a slant with respect to the fronto-parallel only the (cyclopean) eccentricity of the blob must be known.

Figure 7.4 (c) illustrates a fronto-parallel surface. A blob lying on this surface is further from one eye than the other *and* more oblique in one eye than the other. Both these effects can be accounted for. If the horizontal width of a blob were expanded or compressed according to the *square* of the vertical height ratio, then any "residual" horizontal size difference would signal the slant with respect to the fronto-parallel (see figure legend). This latter relationship only applies close to the horizontal meridian. The surface for which the relationship remains true is a cylinder whose axis passes through the two eyes. For objects near the mid-line or far (e.g. greater than 1 metre) from the observer, these three estimates of slant coincide very closely.

The example given here is only one issue within an area that is currently a topic of much debate (Cumming, Johnston and Parker, 1991; Sobel and Collett, 1991; Rogers and Bradshaw, 1992). The example is not intended to predict psychophysical findings (although it would relate, for instance, to the curvature of the apparent fronto-parallel plane (Ogle, 1950)) but rather to indicate how issues can be considered within the framework of a hierarchical scheme. The important aspect of this approach is that the manipulations are specific to a blob, use simple measurements (e.g. blob widths) and do not necessarily involve explicit calculations of "viewing system parameters" such as viewing distance or even the eccentricity of a feature.

7.3 Summary

In the first part of this chapter the main experimental findings (from chapters 4, 5 and 6) have been reviewed. The common theme to these experiments is a hierarchical scale-space model of stereo-matching (described in chapter 3) in which the left and right eyes' images are analysed initially at a coarse scale and, over the first second after the onset of a stimulus, analysed at progressively finer scales (chapter 4). The purpose of this process would be to organise information about the image (the position, separation, orientation, and curvature of features) into a scale-based hierarchy in which only local spatial relations are recorded explicitly

and, via coarse scale groupings, related to features in other parts of the image (chapter 5). According to the model, filter outputs of all sizes contribute to the coarse scale representation, which has the advantage that fine scale information can be represented statistically at short exposures and in peripheral vision, but it has the disadvantage that for dense, fine scale textured patterns only small disparities can be detected (chapter 6).

In the second part of this chapter other experiments were discussed in relation to hierarchical model of stereopsis. The most important of these is the work by Mitchison and McKee (1987a and b). Their model of the matching process is similar in many respects to that put forward in this thesis. Several phenomena such as the anisotropy of slant perception and simultaneous depth contrast were reviewed briefly in the context of a hierarchical scheme to illustrate how the model might be applied to other areas of stereopsis.

The goal of binocular stereopsis is to compare two images, each containing a vast amount of information, and which could be combined in many different ways. The hypothesis put forward in this thesis is that a system based on hierarchical encoding may help to make that problem tractable.

References

Anderson, C.H. and Van Essen, D.C., (1987) Shifter circuits: a computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Sciences, USA*, **84**, 6297-6301

Andrews, D.P. (1964) Error-correcting perceptual mechanisms. *Quarterly Journal of Experimental Psychology,* **16**, 104-115

Baker, C.L. and Braddick, O.J. (1982) The basis of area and dot number effects in random dot motion perception. *Vision Research,* **22**, 1253-1259

Badcock, D.R. and Schor, C.M. (1985) Depth-increment detection function for individual spatial channels. *Journal of the Optical Society of America A*, **2**, 1211-1215

Barlow, H.B. (1978) The efficiency of detecting changes of density in random dot patterns. *Vision Research*, **18**, 637-650

Bischof, W.F. and Di Lollo, V. (1990) Perception of directional sampled motion in relation to displacement and spatial frequency: evidence for a unitary motion system. *Vision Research*, **30**, 1341-1362.

Blake, A. and Zisserman, A. (1987) *Visual Reconstruction*. MIT Press, Cambridge, Mass.

Blakemore, C. (1970a) A new kind of stereoscopic vision. *Vision Research,* **10**, 1181-1199

Blakemore, C. (1970b) The range and scope of binocular depth discrimination in man. *Journal of Physiology, London*, **211**, 599-622

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

Braddick, O.J. (1974) A short-range process in apparent motion. *Vision Research*, **14**, 519-527.

Burbeck, C.A. (1986) Exposure duration effects in localisation judgements. *Journal of the Optical Society of America A*, **3**, 1983-1988

Burt, P. and Julesz, B. (1980) A disparity gradient limit for binocular fusion. *Science*, **208**, 615-617

Cagenello, R.B. (1990) *Perception and representation of stereoscopic slant and curvature*. D.Phil. thesis, University of Oxford.

Cagenello, R.B. and Rogers, B.J. (1988) Local orientation differences affect the perceived slant of stereoscopic surfaces. *Investigative Ophthalmology and Visual Science,* **29**, 339

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

Canny, J. (1986) A computational approach to edge detection. *IEEE transactions on pattern analysis and machine intelligence.* **8**, 679-698

Carlson, C.R., Moeller, J.R. and Anderson, C.H. (1984) Visual illusions without low spatial frequencies *Vision Research,* **24**, 1407-1413

Cavanagh, P., Boeglin, J. and Favreau, O.E. (1985) Perception of motion in equiluminous kinematograms. *Perception*, **14**, 151-162.

Chang, J.J. and Julesz, B. (1983) Displacement limits for spatial frequency filtered random-dot cinematograms in apparent motion. *Vision Research*, **23**, 1379-1385

Chubb, C. and Sperling, G. (1988) Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception *Journal of the Optical Society of America, A.* **5**, 1986-2007

Cleary, R. and Braddick, O.J. (1990a) Direction discrimination for band-pass filtered random dot kinematograms. *Vision Research*, **30**, 303-316.

Cleary, R. and Braddick, O.J. (1990b) Masking of low frequency information in short-range apparent motion. *Vision Research*, **30**, 317-327.

Collett, T.S. (1985) Extrapolating and interpolating surfaces in depth. *Proceedings of the Royal Society of London (B),* **224**, 43-56

Coren, S. (1986) An efferent component in the visual perception of direction and extent. *Psychological Review*, **93**, 391-410

Coren, S. and Hoenig, P. (1972) Effect of non-target stimuli upon the length of voluntary saccades. *Perception and Motor Skills,* **34**, 499-508

Cornell, C.O. (1978) *Quantitative predictions of length in the Müller-Lyer illusion as perceived by the human visual system* MS thesis GE/EE/78M-9, Wright Patterson Air Force base, Ohio

Cumming, B.G., Johnston, E.B. and Parker, A.J. (1991) Vertical disparities and the perception of three-dimensional surfaces. *Nature*, **349**, 411-413

Davidson, M.L. (1968) Perturbation approach to spatial brightness interaction in human vision. *Journal of the Optical Society of America*, **58**, 1300-1309

Dawson, M. and Di Lollo, V. (1990) Effects of adapting luminance and stimulus contrast on the temporal and spatial limits of short-range motion. *Vision Research,* **30**, 415-429

Dev, P. (1975) Perception of depth surfaces in random-dot stereograms: A neural model. *International Journal of Man-Machine Studies*, **7**, 511-528

DeValois, R. and DeValois, K. (1988) *Spatial Vision.* Oxford University Press.

DeAngelis, G.C., Ohzawa, I. and Freeman, R.D. (1991) Depth is encoded in the visual system by a specialised receptive field structure. Nature, 352, 156-158

Dreher, B. and Sanderson, K.J. (1973) Receptive field analysis: Responses to moving visual contours by single lateral geniculate neurons in the cat. *Journal of Physiology (London)*, **234**, 95-118

Eagle, R.A. (1992) *Spatial aspects of human visual motion detection.* D.Phil thesis, University of Oxford.

Eagle, R.A. and Rogers, B.J. (1991) Maximum displacement (d_{max}) as a function of density, patch size, and spatial filtering in random-dot kinematograms. *Investigative Ophthalmology & Visual Science*, **32**, 893

Eagle, R.A. and Rogers, B.J. (1992) D_{max} for motion detection is a dependent on element density not spatial frequency. *Investigative Ophthalmology & Visual Science*, **33**, 1139

Enroth-Cugell, C. and Robson, J.G. (1966) The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology (London)*, **187**, 517-552

Erkelens, C.J. and Collewijn, H. (1985) Motion perception during dichoptic viewing of moving random dot stereograms. *Vision Research*, **25**, 583-588

Fain, G. and Dowling, J.E. (1973) Intracellular recordings from single rods and cones in the mudpuppy retina. *Science*, **180**, 1178-1181

Field, D. J. (1987) Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, **4**, 2379-2394

Findlay, J.M. (1980) The visual stimulus for saccadic eye movements in human observers. *Perception*, **9**, 7-21

Finney, D.J. (1971) *Probit Analysis.* 3rd ed. Cambridge University Press

Felton, T.B., Richards, W. and Smith, R.A. (1972) Disparity processing of spatial frequencies in man. *Journal of Physiology, London,* **225**, 349-362

Fender, D. and Julesz, B. (1967) Extension of Panum's fusional area in binocularly stabilised vision. *Journal of the Optical Society of America* , **57**, 819-830

Frisby, J.P. and Mayhew, J.E.W. (1978) Contrast sensitivity function for stereopsis. *Perception*, **7**, 423-429

Frisby, J.P. and Mayhew, J.E.W. (1980) The role of spatial frequency tuned channels in vergence control. *Vision Research*, **20**, 727-732

Frisby, J.P. and Pollard, S.B. (1991) Computational issues in solving the correspondence problem. In *Computational Models of Visual Processing*, Ed. Landy, M.S. and Movshon, J.A. MIT Press, Cambridge, M.A.

García-Pérez, M.A., (1991) Visual phenomena without low spatial frequencies: a closer look. *Vision Research*, **31**, 1647-1653

Ginsburg, A.P. (1978) *Visual information processing based on spatial filters constrained by biological data.* PhD thesis, Aerospace medical research laboratory, report number TR-78-129

Graham, M and Rogers, B.J. (1982) Simultaneous and successive contrast effects in the perception of depth from stereoscopic surfaces. *Perception*, **11**, 247-262

Grimson, W.E.L. (1981) *From images to surfaces: A computational study of the human early visual system.* Cambridge, MA: MIT Press.

Gu, Y. and Legge, G.E. (1991) Spatial localisation accuracy of centroids, peaks and zero-crossings. *Investigative Ophthalmology and Visual Science,* **32** 1268

Henning, G.B., Hertz, P.F. and Broadbent, D.E. (1975) Some experiments bearing on the hypothesis that the visual system analyses patterns in independent bands of spatial frequency. *Vision Research*, **15**, 887-899

Howard, I.P. and Ohmi, M. (1992) A new interpretation of the role of dichoptic occlusion in stereopsis. *Investigative Ophthalmology and Visual Science,* **33**, 1370

Julesz, B. (1971) *Foundations of Cyclopean Perception.* Univ. of Chicago Press, Ill.

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

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

Julesz, B. and Spivack, G.J. (1967) Stereopsis based on vernier acuity cues alone *Science*, **157**, 563-565

Koenderink, J. and van Doorn, A.J. (1976) Geometry of binocular vision and a model of stereopsis. *Biological Cybernetics*, **21**, 29-35

Lau, E. (1922) Versuche uber das stereoskopische sehen. *Psychologische Forschung,* **2**, 1-4

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

Lehky, S.R. and Sejnowski, T.J. (1990) Neural models of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *Journal of Neuroscience*, **10**, 2281-2299

Levi, D.M. and Klein, S.A. (1989) Both separation and eccentricity can limit precise position judgements: a reply to Morgan and Watt. *Vision Research*, **29**, 1463-1469

Lewis, E.O. (1909) Confluxion and contrast effects in the Müller-Lyer illusion. *British Journal of Psychology*, **2**, 19-41

Lindeberg, T.P. (1991) *Discrete scale-space theory and the scale-space primal sketch.* PhD. thesis, Stockholm University, Sweden.

Longuet-Higgins, H.C. (1962) The distribution of intervals between zeros of a stationary random function. *Philosophical Transactions of the Royal Society of London A*, **254**, 557-599

Marr, D (1976) Early processing of visual images. *Philosophical Transactions of the Royal Society of London, B,* **275**, 483-524

Marr, D. (1982) *Vision. A computational investigation into the human representation and processing of visual information.* W.H. Freeman, New York

Marr, D. and Hildreth, E. (1980) Theory of edge detection. *Proceedings of the Royal Society of London (B),* **207**, 187-217

Marr, D. and Nishihara, H.K. (1978) Representation and recognition of the spatial organisation of three-dimensional shapes *Proceedings of the Royal Society of London (B),* **200**, 269-294

Marr, D. and Poggio, T. (1976) Co-operative computational theory of human stereo vision. *Science*, **194**, 283-287

Marr, D. and Poggio, T. (1979) A computational theory of human stereo vision *Proceedings of the Royal Society of London (B),* **204**, 301-328

Mayhew, J. (1983) Stereopsis. In *Physical and Biological Processing of Images*: ed. Braddick, O.J. and Sleigh, A.C., Springer-Verlag, Berlin.

Mayhew, J.E.W.and Frisby, J.P.(1978) Stereopsis masking in humans is not orientationally tuned. *Perception*, **7**, 431-436

Mayhew, J.E.W. and Frisby, J.P. (1979) Convergent disparity discriminations in narrow-band-filtered random-dot stereograms.*Vision Research* ,**19**, 63-71

Mayhew, J.E.W. and Longuet-Higgins, H.C. (1982) A computational model of binocular depth perception. *Nature*, **297**, 376-378

Michelson, A.A. (1891) On the application of interference methods to spectroscopic measurements. I. *Phil. Mag. Ser. V* **31**, 338-348

Mitchison, G.J. (1988) Planarity and segmentation in stereoscopic matching. *Perception*, **17**, 753-782

Mitchison, G.J. (1992) Learning a stable representation of stereoscopic depth. (In press)

Mitchison, G.J.and McKee, S.P. (1985) Interpolation in stereoscopic matching. *Nature*, **315**, 402-404

Mitchison, G.J.and McKee, S.P. (1987a) The resolution of ambiguous stereoscopic matches by interpolation. *Vision Research,* **27**, 285-294

Mitchison, G.J.and McKee, S.P. (1990) Mechanisms underlying the anisotropy of stereoscopic tilt perception. *Vision Research,* **30**, 1781-1791

Mitchison, G.J. and Westheimer, G. (1984) The perception of depth in simple figures. *Vision Research*, **24**, 1063-1073

Mitchison, G.J. and Westheimer, G. (1990) Viewing geometry and gradients of horizontal disparity. In *Vision: coding and efficiency*. ed. Blakemore, C., Cambridge University Press, p302

Morgan, M.J. (1989) Vision of solid objects. *Nature*, **339**, 101-102

Morgan, M.J. (1992) Spatial filtering precedes motion detection. *Nature*, **335**, 344-346

Morgan, M.J. and Fahle, M. (1992) Effects of pattern element density upon displacement limits for motion detection in random binary luminance patterns. *Proceedings of the Royal Society of London (B),* **248**, 189-198.

Morgan, M.J. and Glennerster, A. (1991) Efficiency of locating centres of dot-clusters by human observers. *Vision Research*, **31**, 2075-2083

Morgan, M.J., Hole, G.J. and Glennerster, A. (1990) Biases and sensitivities in geometric illusions *Vision Research,* **30**, 1793-1810

Morgan, M.J. and Watt, R.J. (1989) The Weber relation for position is not an artefact of eccentricity. *Vision Research,* **29**, 1457-1462

Motter, B.C. and Poggio, G.F. (1984) Binocular fixation in the rhesus monkey: spatial and temporal characteristics. *Experimental Brain Research,* **54**, 304-314

Motter, B.C. and Poggio, G.F. (1990) Dynamic stabilisation of receptive fields of cortical neurons (V1) during fixation of gaze in the macaque. *Experimental Brain Research,* **83**, 37-43

Mowforth,P., Mayhew, J.E.W.and Frisby, J.P.(1981) Vergence eye movements made in response to spatial-frequency-filtered random-dot stereograms. *Perception*, **10**, 299-304

Nachmias, J. and Rogowitz, B.E. (1983) Masking by spatially modulated gratings. *Vision Research*, **23**, 1621-1629.

Nakayama, K. and Silverman, G. (1984) Temporal and spatial characteristics of the upper displacement limit for motion in random dots *Vision Research*, **24**, 293-299

Nishihara, H.K. (1984) Practical real-time imaging stereo matcher. In *Readings in Computer Vision* ed.Fischler, M.A. and Firschein, O., Kauffman, Los Altos, California, p 63-72

Ogle, K.N. (1950) *Researches in binocular vision*, W.B. Saunders, Philadelphia and London

Ogle, K.N. (1952) On the limits of stereoscopic vision. *Journal of Experimental Psychology,* **44**, 253-259

Ogle, K.N. (1953) Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America,* **43**, 906-913

Ogle, K.N. (1962) The optical space sense. *The eye*, ed. H. Davson, New York: Academic Press, **4**, 211-432

Ohwaki, S. (1960) On the destruction of geometrical illusions in stereoscopic observation. *Tohoku Psychologica Folia*, **29**, 24-36

Ohzawa, I., DeAngelis, G.C. and Freeman, R.D. (1990) Stereoscopic depth perception in the visual cortex: neurons ideally suited as disparity detectors. *Science,* **249**, 1037-1041

Papert, S. (1961) Centrally produced geometrical illusions. *Nature*, **191**, 733

Parker, A.J., Johnston, E.B., Mansfield, J.S. and Yang, Y. (1991) Stereo, surfaces and shape. In *Computational Models of Visual Processing*, Ed. Landy, M.S. and Movshon, J.A. MIT Press, Cambridge, M.A.

Parker, A.J. and Yang, Y. (1989) Spatial properties of disparity pooling in human stereo vision. *Vision Research*, **29**, 1525-1538

Peli, E. and Goldstein, R.B. (1988) Contrast in images. *SPIE* **1001** *Visual Communications and Image Processing*, 521-528

Pelli, D.G. and Zhang, L. Accurate control of contrast on microcomputer displays. *Vision Research*, **31**, 1337-1350

Pollard, S.B., Mayhew, J.E.W. and Frisby, J.P. (1985) PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception,* **14**, 449-470

Poggio, G.F., Motter, B.C., Squatrito, S. and Trotter, Y. (1985) Responses of neurons in the visual cortex (V1 and V2) of the alert macaque monkey to dynamic random-dot stereograms. *Vision Research*, **25**, 397-406

Poggio, G.F. and Poggio, T (1984) The analysis of stereopsis. *Annual Review of Neuroscience,* **7**, 379-412

Prazdny, K. (1985) Detection of binocular disparities. *Biological Cybernetics*, **52**, 93-99

Prazdny, K (1987) On the coarse-to-fine strategy in stereomatching. *Bulletin of the Psychonomic Society,* **25**, 92-94

Quam, L.H. (1984) Hierarchical warp stereo. In *Readings in Computer Vision* ed.Fischler, M.A. and Firschein, O., Kauffman, Los Altos, California, p80-86

Rashbass, C. and Westheimer, G. (1961) Disjunctive eye movements. *Journal of Physiology*, **159**, 339-360

Rentschler, I. and Treutwein, B. (1985) Loss of spatial phase relationships in extrafoveal vision. *Nature*, **313**, 308-310

Robson, J.G. (1966) Spatial and temporal contrast sensitivity functions of the visual system. *Journal of the Optical Society of America*, **56**, 1141-1142

Rodieck, R.W. and Stone, J. (1965) Analysis of receptive fields of the cat retinal ganglion cells. *Journal of Neurophysiology*, **28**, 833-849

Rogers, B.J. and Bradshaw, M.F. (1993) Vertical disparities, differential perspective and binocular stereopsis. *Nature*, **361**,253-255 .

Rogers, B.J. and Cagenello, R. (1989) Disparity curvature and the perception of three dimensional surfaces. *Nature*, **339**, 135-137

Rogers, B.J. and Cagenello, R. (1990) The role of line intersection in binocular stereopsis. *Investigative Ophthalmology and Visual Science*, **31**, p303

Rogers, B.J. and Graham, M. (1983) Anisotropies in the perception of three-dimensional surfaces. *Science*, **221**, 1409-1411

Rogers, B.J. and Koenderink, J. (1986) Monocular aniseikonia: a motion parallax analogue of the disparity induced effect. *Nature*, **322**, 62-63

Sato, T. (1990) Effects of dot size and dot density on motion perception with random-dot kinematograms. *Perception*, **19**, 329.

Saye, A. and Frisby, J.P. (1975) The role of monocularly conspicuous features in facilitating stereopsis from random-dot stereograms. *Perception*, **4**, 159-171

Schor, C.M. and Wood, I. (1983) Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research,* **23**, 1649-1654

Schor, C.M., Wood, I. and Ogawa, J. (1984) Binocular sensory fusion is limited by spatial resolution. *Vision Research,* **24**, 661-665

Schumer, R.A. (1979) *Mechanisms in human stereopsis*. PhD. Thesis, Stanford University, Palo Alto, California

Schumer, R.A. and Julesz, B. (1984) Binocular disparity modulation sensitivity to disparities offset from the plane of fixation. *Vision Research*, **24**, 533-542

Shimojo, S. and Nakayama, K. Real world occlusion constraints and binocular rivalry. *Vision Research,* **30**, 69

Smallman, H.S. and MacLeod, D.I.A. (1992) Fine-to-coarse scale disambiguation in human stereopsis. *Investigative Ophthalmology and Visual Science,* **33**, 1369

Smallman, H.S. and MacLeod, D.I.A. A size-disparity correlation in stereopsis at contrast threshold. (Manuscript in preparation.)

Sobel, E.C. and Collett, T.S. (1991) Does vertical disparity scale the perception of stereoscopic depth? *Proceedings of the Royal Society, B*, **244**, 87-90

Springbett, B.N. (1961) Some stereoscopic phenomena and their implications. *British Journal of Psychology*, 52, 105-109

St-Cyr, G.J. and Fender, D.H. (1969) The interplay of drifts and flicks in binocular fixation. *Vision Research*, **9**, 245-265

Steinman, R.M., Cushman, W.B. and Martins, A.J. (1982) The precision of gaze. *Human Neurobiology*, **1**, 97-109

Stevenson, S.B., Cormack, L.K. and Schor, C.M. (1989) Hyperacuity, superresolution and gap resolution in human stereopsis. *Vision Research*, **29**, 1597-1605

Toet, A. and Koenderink, J.J. (1988) Differential spatial displacement discrimination thresholds for gabor patches. *Vision Research*, **28**, 133-143

Tyler, C.W. and Barghout, L. (1992) Multiple channels in stereoscopic masking: a computational analysis. *Investigative Ophthalmology and Visual Science*, **33**,1334

Tyler, C. and Sutter (1979) Depth from spatial frequency difference: an old kind of stereopsis? *Vision Research* ,**19**, 859-865.

Voss, R.F. (1985) "Random fractal forgeries" in *Fundamental Algorithms for Computer Graphics*, R.A. Earnshaw, ed. (Springer-Verlag, Berlin) p805-829

Wallach, H. and Bacon, J. (1976) Two forms of retinal disparity. *Perception and Psychophysics*, **19**, 375-382

Watt, R.J. (1985) Structured representation in low-level vision. *Nature*, **313**, 266-267

Watt, R.J. (1987) Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A*, **4**, 2006-2021.

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

Watt, R.J. (1991) *Understanding Vision*. Academic Press Ltd, London.

Watt, R.J. and Morgan, M.J. (1983) The recognition and representation of edge blur: Evidence for spatial primitives in human vision. *Vision Research*, **23**, 1457-1477.

Watt, R.J. and Morgan, M.J. (1984) Spatial filters and the localisation of luminance changes in human vision. *Vision Research*, **24**, 1387-1397

Watt, R.J. and Morgan, M.J. (1985) A theory of the primitive spatial code in human vision. *Vision Research*, **25**, 1661-1678.

Westheimer, G. (1979) Cooperative neural processes involved in stereoscopic acuity. *Experimental Brain Research*, **36**, 585-597

Westheimer, G. (1986) Spatial interaction in the domain of disparity signals in human stereoscopic vision. *Journal of Physiology*, **370**, 619-629

Westheimer, G. and Levi, D.M. (1987) Depth attraction and repulsion of disparate foveal stimuli. *Vision Research*, **27**, 1361-1368

Westheimer, G. and McKee, S.P. (1977) Spatial configurations for visual hyperacuity. *Vision Research*, 17, 941-947

Westheimer, G. and Tanzman, I.J. (1956) Qualitative depth localisation with diplopic images. *Journal of the Optical Society of America*, **46**, 116-117

Wilson, H.R. and Bergen, J.R. (1979) A four mechanism model for spatial vision. *Vision Research*, **19**, 19-32

Wilson, H.R. and Giese, S.C. (1977) Threshold visibility of frequency gradient patterns. *Vision Research*, **17**, 1177-1190

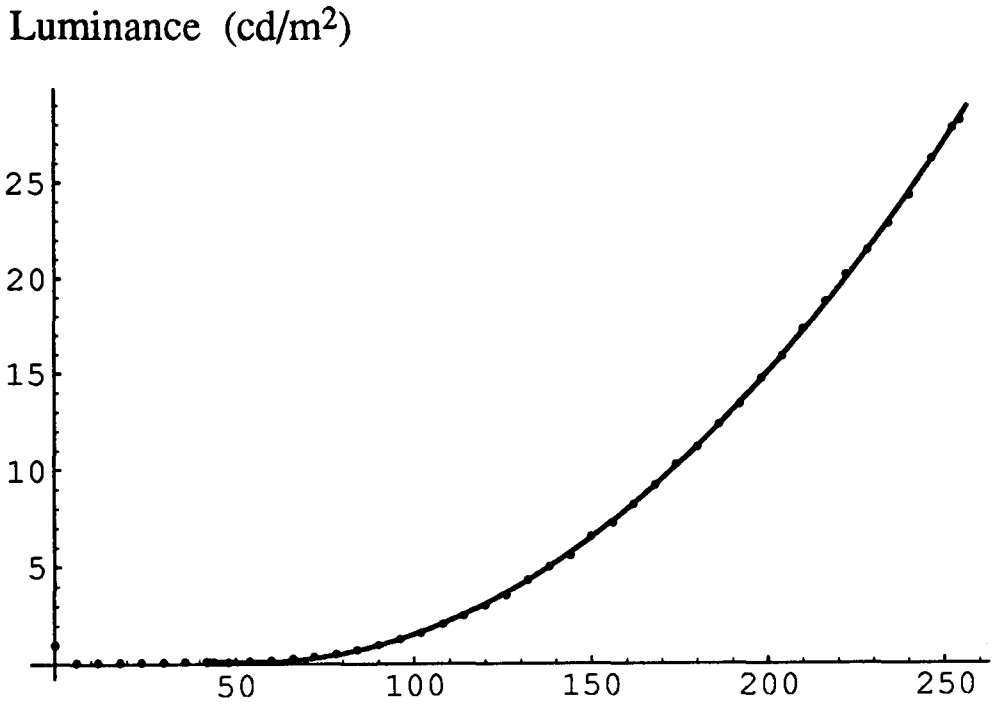
Witkin, A.P. (1988) Scale-space filtering. In *Readings in Computer Vision* ed.Fischler, M.A. and Firschein, O., Kauffman, Los Altos, California, p329-332

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

Appendix A

This appendix describes how the screens were linearised for the experiment using grey level stimuli (experiment IV) in chapter 6. The method is based on that described by Pelli and Zhang (1991). The monitors were Apple High-Resolution Monochrome monitors.

The screen luminance was measured using a Minolta Luminace Meter LS-110, (Minolta Camera Co. Ltd, Japan). The monitor was set up with the contrast and brightness controls at the levels used in the experiment. A square, 50 by 50 pixels, was displayed in the centre of the screen. The grey-level (input value) for this square could be varied. Apart from this square, the stimulus display was as close as possible to that during the experiment, i.e. a filtered image of 128 by 128 pixels was displayed in the centre of the screen and the rest of the screen was black (0.12 cd/m^2). The luminace meter was positioned 57 cm from the screen and aimed at the centre of the variable grey-level square.



This graph shows the measured luminance (L) before the screen was linearised as a function of the nominal voltage v = n / 255 resulting from loading the same number n (n = 0, 1, 2, ... 255) into all three DACs (red, green and blue) in the Apple Colour Video Card. The luminace was measured for every sixth grey level between 0 and 255 (plot shows mean of three measurements). The function has been fitted with a second order polynomial between the values of n = 42 and n = 254. (Pelli and Zhang recommend using one function to fit the plateau in the low luminace range and another for the rest of the curve.)

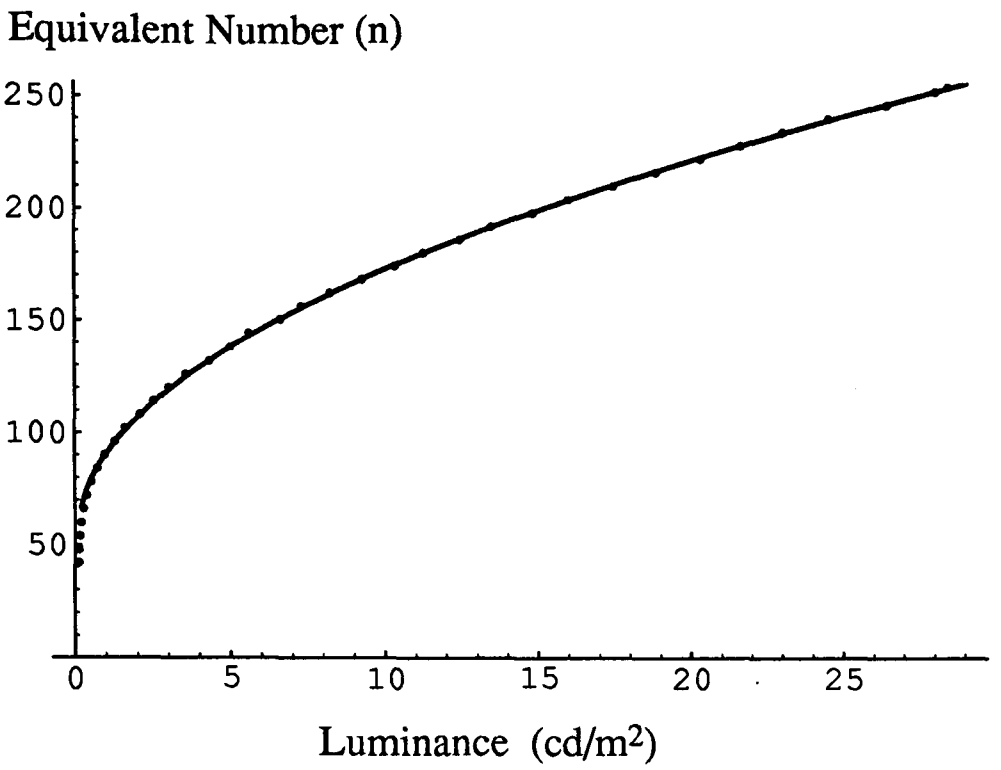
The equation of the curve is:

$$L = 2.34 - 0.0802n + 0.000722\, n^2 . \qquad \text{(i)}$$

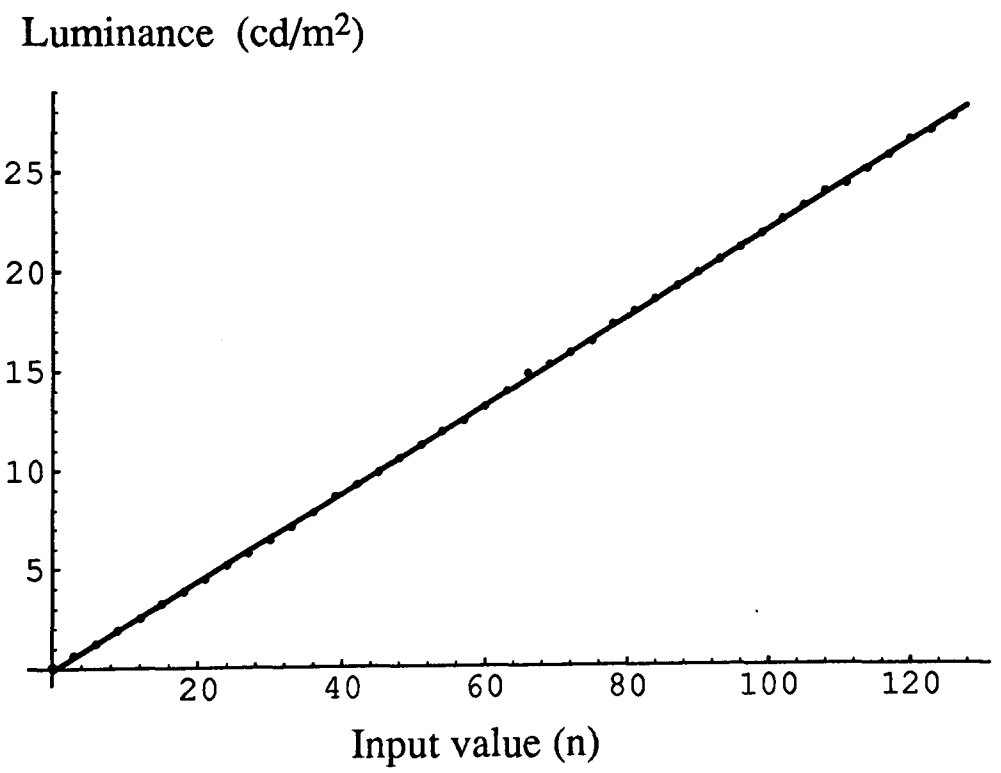
Solving for n (using Mathematica, positive root):

$$n = 1621 * (0.0343 + 0.02296 * \text{Sqrt}[-0.1116 + L]) . \qquad \text{(ii)}$$

This inverse function is shown in the graph below and was used to create a linear look-up table. Equation (ii) does not provide a good fit for the very lowest luminance values and the lowest two entries of the table were set by hand.



The stimuli (input arrays) used in the experiment contained up to 128 grey levels (input values) ranging from 0 to 127. The screen luminance was measured after linearisation for 43 of the 128 possible input values (i.e. every third input value). The results are shown in the graph below.



Appendix B

This appendix describes the method used for making Laplacian and Gaussian filters, calculating MIRAGE responses and the position of 1-D centroids.

The equation that was used in this thesis for a Gaussian is

$$G\,(r,f,\theta) = e^{-r2/2f2} \qquad \text{(i)}$$

where r is the distance and θ the direction from the centre, f is the space constant (standard deviation of the Gaussian). The equation that was used for a Laplacian of Gaussian (LoG) is

$$\nabla^2 G\,(r,f,\theta) = \left(1 - \frac{r^2}{2f^2} \right) e^{-r2/2f2} \qquad \text{(ii).}$$

The Laplacian and Gaussian filters are circularly symmetric (theta does not appear on the left hand side of the equation).In equation (i) and (ii) the maximum amplitude (at r=0) is unity for all filters. The LoG filters used by Watt and Morgan (1985) in their MIRAGE model all had an equal gain (peak-to-trough amplitude) in the spatial domain and the convention has been adopted in this thesis even though much larger filters have been used.

The code used to calculate LoG filters, using the HIPS image processing package (Turing Institute, Strathclyde, UK) was:

```
fcalcpix -g 256 'i1=256; /*frame size*/ d5 = 8.0; /*s.d. of gauss*/ d1=( (r-(i1/2))*(r-
(i1/2)) + (c-(i1/2))*(c-(i1/2)) ); d2 = d1 / (2 * d5 * d5); opix = (1 - d2) *
pow(2.71828,-d2);' | fourtr | phase_mag -m > l8
```

for a filter with a space constant of 8 pixels.

There are three stages to calculating the MIRAGE S+ and S- response: filtering, rectification and summation. In most of the examples given in this thesis (and in Watt and Morgan, 1985) four filters were used with space constants in the ratio of 1:2:4:8.

The filtering stage can be defined as:

$$R_i = F_i * I$$

i.e the image function, I, is convolved with a set of filters F_i (where, unless otherwise specified in the thesis, i = 1 to 4). The images used in this thesis were obtained by multiplication by a filter in the Fourier domain which is equivalent to convolution in the spatial domain.

Rectification: $R_i^+ = R_i^+$ if $R_i > 0$
$R_i^+ = 0$ otherwise.
$R_i^- = -R_i^-$ if $-R_i > 0$
$R_i^- = 0$ otherwise.

where R_i^+ and R_i^- are the positive and negative responses of each filter respectively;

and Summation: $S_i^+ = R_1^+ + R_2^+ + R_3^+ ..... + R_n^+$
$S_i^- = R_1^- + R_2^- + R_3^- ..... + R_n^-$

The code used to carry out these steps, again using the HIPS package, was:

```
mulseq l4 < FT | inv.fourtr -f > filt/l4
mulseq l8 < FT | inv.fourtr -f > filt/l8
mulseq l16 < FT | inv.fourtr -f > filt/l16
mulseq l32 < FT | inv.fourtr -f > filt/l32

pos < filt/l4 > Spos/l4
pos < filt/l8 > Spos/l8
pos < filt/l16 > Spos/l16
pos < filt/l32 > Spos/l32

addseq Spos/l8 < Spos/l4 | addseq Spos/l16 | addseq Spos/l32 > S4
```

where FT is a fourier transform of the image, I, and pos is defined as

```
fcalcpix 'if(ipix < 0) opix = 0;'.
```

All of the examples shown in this thesis are S+ responses.

The position of a 1-D centroid, P_i, is the position within a zero-bounded distribution about which the first order moment is zero:

$$P_i = \frac{\int_{Z_{c_i}}^{Z_{c_{i+1}}} x \,.\, R(x) \,.\, dx}{\int_{Z_{c_i}}^{Z_{c_{i+1}}} R(x) \,.\, dx}$$

where Z_{c_i} and $Z_{c_{i+1}}$ are the positions of adjacent zero-crossings and R(x) is the response at point x. In the thesis, centroids were always calculated along horizontal raster lines of pixelated images. In practice, Z_{c_i} and $Z_{c_{i+1}}$ were defined as pixels whose grey level exceeded a minimum threshold (0.0000001, which is less than 0.1% of the peak amplitude of the smallest filter, l4, for a 50% density pattern). Z_{c_i}, $Z_{c_{i+1}}$ and P_i were all rounded down to the nearest integer value. The left and right hand edges of the pattern (x = 0 and x = 255) were treated as zero-crossings even though the fourier transform "wraps round".

For the centroid-matching model described in chapter 6, only centroids above a certain threshold mass were used as input to the matching program. A fragment of the code used to calculate centroids is included below.

```
# define THRESH_MASS 0.000100
# define THRESH_GREY 0.0000001
for(j=0;j<VSIZE;j++)
  for(i=0;i<HSIZE;i++) {
     if((array[ (HSIZE * j) + i ] > THRESH_GREY ) && (i < (HSIZE-1))) {
     /*Within a blob or zero-bounded distribution*/
       zbd = TRUE;
       mass += array[ (HSIZE * j) + i ];
       moment += i * array[ (HSIZE * j) + i ];
     }
     else {
       if(zbd == TRUE) { /*i.e last pixel was in a zbd, now coming out of one*/
         centroid = moment / mass;
```

```
        if(mass > THRESH_MASS)
          array2[ (HSIZE * j) + centroid ] =  MARKER;
        zbd = FALSE;
        mass = 0.0;
        moment = 0.0;
      }
     else
       { } /* Region of Null Response*/
   }
}
```