From detection of complex motion to descriptions of moving surfaces in human vision

PHILIP ANTHONY ATKINSON

Doctor of Philosophy

ASTON UNIVERSITY

December 2004

Summary

Aston University
From detection of complex motion to descriptions of moving surfaces in human
Philip Anthony Atkinson
Doctor of Philosophy
2004

A preliminary study by Freeman *et al* (1996b) has suggested that when complex
of motion elicit impressions of 3-dimensionality, odd-item-out detection improv
targets can be differentiated on the basis of surface properties. Their results
accounted for, if it is supposed that observers are permitted efficient access to 3-
descriptions but access to 2-D motion descriptions is restricted. This is consist
previous research, which has demonstrated higher-level surface representatio
given priority access over early-level representations. To test the hypothesis, a
search technique was employed, in which targets could be discerned on the basis
sign. In one experiment, slant impressions were induced through the sum
deformation and translation components. In a second they were induced thr
summing of shear and translation components. Neither showed any evidence of
access. A third experiment explored the possibility that access to these representat
have been hindered by a lack of grouping between the stimuli. Attempts to
grouping failed to produce convincing evidence in support of this. An al
explanation is that complex patterns of motion are simply not processed simult
Psychophysical and physiological studies have, however, suggested that
mechanisms selective for complex motion do exist. Using a subthreshold su
technique I found evidence supporting the notion that complex motions are pro
parallel. Furthermore, in a spatial summation experiment, coherence thresho
measured for displays containing different numbers of complex motion patches. C
with the idea that complex motion processing proceeds in parallel, increases in th
of motion patches were seen to decrease thresholds, both for expansion and
Moreover, the rates of decrease were higher than those typically expected from p
summation, thus implying mechanisms are available, which can pool signals from
distinct complex motion flows.

Key Words: Complex motion; Surface description; Slant; Parallel processin
search

Dedicated to the memory of Joan Davis

Acknowledgements

My thanks to Dr Tim Meese, for academic supervision, and the four subjects who sat for the experiments of this thesis. Special thanks also go to Ellie Cartwright for her kindness and support.

List of Contents

List of Figures

List of Tables

Chapter One: General Introduction

1.1 Chapter Introduction

The human visual system faces a task of constructing meaningful representations of the world from 2-D projections. Its ability to do this is impressive, especially given the fact that sub-optimal photoreceptor densities and the optics of the eye limit the sampling of incident light at the eye. Furthermore, brain processes are inherently noisy. Despite the above, vision typically provides us with stable representations that are sufficiently faithful to afford complex behavioural interaction with the external world. For example, our visual capabilities can play key roles in activities such as navigation and the foraging of fruit.

How though, can the brain obtain veridical interpretation from these noisy and incomplete descriptions of the world? Under some circumstances, these ambiguities may not be resolved and need not be. For example, human colour vision provides a perceptual mapping for the spectral composition of visible light, but it is severely limited by its sampling capabilities; in the majority of humans, only three photoreceptor types exist, and these are used to encode the entire spectral range. This coarse coding permits an opportunity for vastly different power spectra to elicit identical combinations of photoreceptor responses and consequently the same colour perceptions (assuming the surround is constant). In our natural ecology, we would rarely, if ever be challenged by this ambiguity, and therefore, in evolutionary terms, there is little to be gained from solving it. However, other ambiguities are solved. One example of this comes from the fact that, even under monocular viewing, comprehensive impressions of 3-dimensionality can be achieved.

To avoid ambiguity, the visual system must make assumptions about the likely origins of the signals it receives. Their interpretation is not the product of a single step, but emerges from a multistage hierarchical processing, in which increasingly abstract and complex properties are represented at progressively latter stages (Treue, 2003). Psychophysics and physiology have been particularly successful in describing the early stages of this hierarchical processing. For example, using psychophysical techniques, Campbell & Robson (1968), and Blakemore & Campbell (1969) revealed the existence of multiple spatial-frequency tuned channels. Hubel & Wiesel (1968) also confirmed the existence of cells in the monkey, which preferentially respond to luminance contrasts of a certain orientation. In the time since these seminal findings, substantial evidence has accumulated in support of the notion that there is an initial image-based description, which is served by mechanisms selective for spatial frequency, orientation, and retinal position. The task of subsequent stages is to further add to this description. For example, edges that are not defined by luminance contrast at the initial stages, might be deduced by taking into account the local representations in the surrounding area.

There is also a degree of modularity in the hierarchical processing, to the extent that specific extrastriate areas have been linked to the processing of specific facets of vision like form, colour, and motion. In the early visual areas (e.g. V1 and V2), processing proceeds in a highly parallel manner. The modular organisation of the visual brain suggests that, in some way, this parallelism must persist to higher levels, as the different facets of vision are served by distinct neural substrates. It is of interest, however, as to whether these specialised areas can derive complex representations simultaneously. If processing resources tend to be more limited at these latter stages, with respect to processing demand, multiple representations may need to be derived in a sequential manner. The primary purpose of this thesis is to examine the nature of processing for higher-level representations of motion.

1.2 Optic flow and the extraction of 3-dimensionality from motion

Because retinal images are 2-dimensional, the flow fields that are induced by the relative motions of the observer to the real world are inherently ambiguous with respect to their origin. For example, a shearing retinal flow might be caused by a translating surface, which is slanted relative to the observer or the motion of a non-rigid deforming entity (such as a fluid in motion). Of course, by utilising other facets of vision, like form perception, the occurrence of these ambiguities might be avoided. Rogers and Graham (1979), however, showed that motion cues alone are sufficient in inducing stable and compelling impressions of 3-D structure. The following discussion explores a method of analysis that the visual system might use to derive convenient descriptors of 3-D structure and self-motion.

1.2.1 Optic flow analysis

Observer motion can be fully described as a combination of rotation and translation. Four components of motion are therefore needed to describe observer motion in terms of a 3-D coordinate system: two orthogonal axes of rotation and two orthogonal axes of translation. The convention taken here is to have one of these orthogonal axes, from both the translation and rotation components, coincide with the line of sight; these are denoted the radial components (t_R,r_R). The two other components are denoted as the transverse components (t_T,r_T). Fig.1.3 shows a situation in which a mobile observer views a planar surface from a single instantaneous vantage point. Both the observer's rotational and translational components are considered in relation to a single infinitesimally small patch of this surface.

1.2.2 Defining 'specific transverse velocity' (a_T) and 'specific radial velocity' (a_R)

In schematic A of fig.1.1, the observer is shown to have a translation, *t*, while the surface is stationary. In schematic B, the surface is shown to have the same translation, *t*, while the observer is stationary. In both cases the instantaneous transverse component of this motion (t_T) is given by $t \cdot \sin(\theta)$. This measure, however, is insufficient for fully describing the relative motion between the surface and the observer. To provide a more accurate description, t_T must also take into account viewing distance (*vd*). This is done by scaling t_T, by *vd*.

$$a_T = \frac{t \cdot \sin(\theta)}{vd} \qquad (1.1)$$

alternatively

$$a_T = \frac{t_T}{vd} \qquad (1.2)$$

This resultant measure, a_T,, is of 'specific transverse velocity', and has the units, $\sec^{-1}$. As the 'specific transverse velocity' has the same units as the 'transverse rotational velocity' it can be interpreted as a measure of "apparent rotation" (see Koenderink, 1986). The radial component of translation (t_R) can be expressed in a similar way, through the scaling of t_R by *vd*, giving a measure of 'specific radial velocity' (a_R). This can be considered a measure of 'nearness in time' (see Koenderink, 1986).

From eqn.1.2 it is clear that the extraction of 'transverse rotational velocity' (r_T) and 'specific transverse velocity' (a_T) could be useful in determining viewing distance (*vd*). In fig.1.1A, where the observer is moving and the surface is stationary, both a_T and r_T might, for example, be estimated from the proprioceptive feedback that is derived from ego-motion and tracking eye movements.

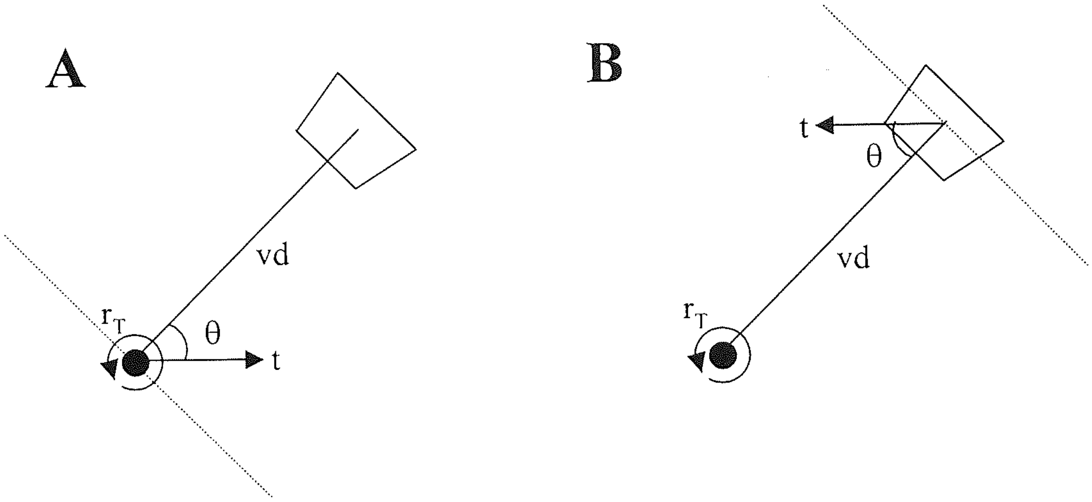


Fig. 1.1. Analysis of both observer motion relative to a stationary surface, and surface motion relative to a stationary observer. Both **A** and **B** show top-down views in which a local patch of surface (trapezium shape) and observer (black dot) are separated by a distance of *vd*. In **A**, the observer is rightwardly translating at speed, t, at an angle θ relative to the direction of a local patch of surface. In **B**, the same translation is occurring, but this time from the allocentric perspective of the surface. The observer's rotation is around a vertical axis, so that r_T indicates the magnitude of yawing motion, induced by horizontal eye movement (obviously head and body movements could also contribute to this yaw). (Modified version of fig.11.1 of Bruce *et al*, 2004.)

1.2.3 The extraction of differential invariant components

Koenderink (1975; 1986) has proposed an analysis of optic flow based upon local sampling and an assumption of object rigidity. This mathematical analysis can fully describe the local flow field in terms of a 1-D translation vector and two, 2-D differential invariant vectors (one describing motions belonging to the generic spiral group and the other describing motions belonging to the generic deformation group). In his exposition, these latter two vectors are considered as pairs of scalars with orientations orthogonal to each other. Typically, the following scalars are used. One component of the spiral vector has its speed gradient in the same orientation as its direction of motion and is referred to as *Div* (see fig.1.4. for schematic illustrations, which provide approximations to these four components). The second component has its speed gradient at right angles to the direction of motion and is known as *Curl.* Deformations contain two orthogonal axes of pure compression of an opposite polarity. Therefore, unlike the spirals with their radially symmetric RF organisation, the orientation of the speed gradient relative to the direction of motion varies depending upon position. For this reason, the orientation of deformation is instead defined in terms of one of its compression axes. Within this thesis the axis is taken to be that of positive compression (i.e. a 1-D expansion) with its angle measured clockwise from vertical. The first component of deformation, *DefI*, has an orientation of -45° and the second component, *DefII*, an orientation of 90°.

The choice of these scalar components is not entirely arbitrary; the two spiral group scalars correspond to the pure flow field structures that would be imposed by the radial components of rotations and translations derived from mobile observers viewing frontoparallel surfaces. A radial translation would generate a field of pure *Div*, while radial rotations would generate a field of pure *Curl.* An argument can also be constructed from a biological perspective for the favouring of *Div* and *Curl* (section 1.3.3). Burr *et al* (2001) have provided psychophysical evidence that is consistent with the notion that a set of *Div* and *Curl* mechanisms are used to encode spiral group motions. However, other research has suggested that intermediate mechanisms might also exist (Graziano *et al,* 1994; Snowden & Milne, 1996; Meese & Harris, 2001a; Meese & Anderson, 2002).

The choice of the deformation components bears less of a relationship to the effects of self motion in the physical world, and more to do with convenience. In fig.1.2 slanted surfaces are shown translating transversely (either vertically or horizontally). Those depicted on the right side (horizontal tilts of 90°) and translating horizontally are shown to elicit flows containing *DefI* but no *DefII.* Conversely, those depicted on the left side (vertical tilts of 0°) and translations in the same directions are shown to elicit flows containing *DefII* but no *DefI.* If these same surfaces translate in an orthogonal direction, however, the opposite is true (i.e. there is no direct relationship between *DefI-DefII* ratios and surface tilt - tilt is also dependent upon the direction of translation).

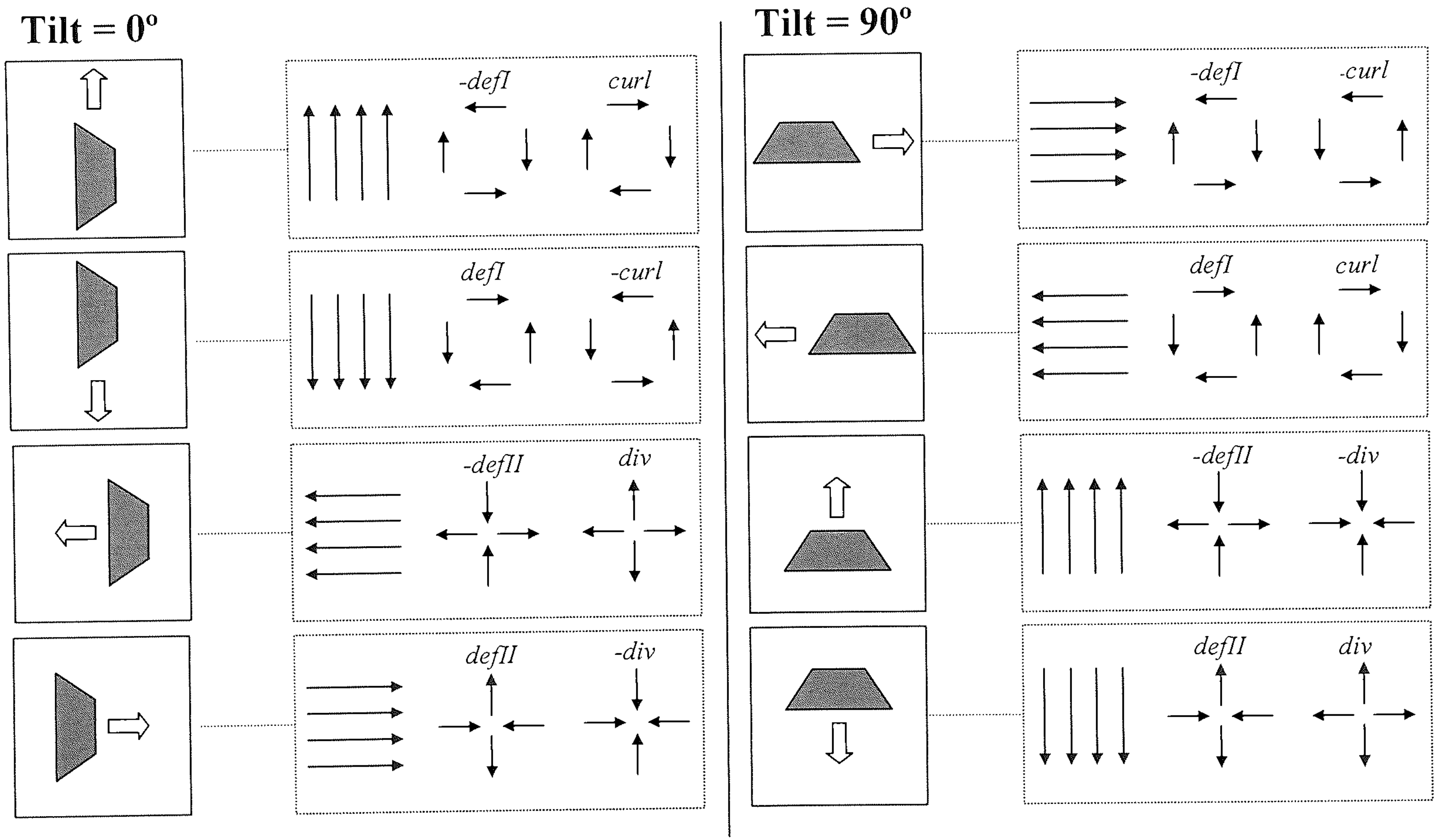


Fig. 1.2. Differential invariants associated with translating planar surfaces. Within the solid squares are slanted surfaces travelling up, down, left or right relative to the observer. Trapezium shapes illustrate a square planar surface slanted in depth, with its direction of motion indicated by the pictorial arrows. The other three arrowed configurations within the accompanying dashed boxes illustrate which of the four optic flow components (*defI, defII, Curl* and *Div*) are stimulated by the corresponding surface motion given a 2-D analysis occurs. Note that orthogonal pairs of 2-D mechanisms alone would be incapable of providing sufficient information to disambiguate surface tilt; comparing conditions across the two rows shows that identical components of deformation (*defI* and *defII*) are elicited from different tilts.

Koenderink (1985; 1986) described the mathematical relationships between observer motions relative to local patches of surface (here concern is directed only toward those first-order characteristics of slant and tilt) and the magnitudes of the 2-D differential invariants obtained from the resulting flow fields. In fig.1.3 the vector, F, describes the slant and tilt of the viewed surface relative to the orientation of the observer. The arctan function of $/F/$ gives slant magnitude (S_{Φ}) relative to the line of sight. The angle of F relative to the vertical gives the surfaces tilt angle, (T_{Φ}). Here, the angular difference between, T_{Φ} and the transverse component of translation (t_T) is denoted θ. Formulae for the magnitudes of the differential invariants and translation components are as follows:

$$trans = -t_r - r_r \tag{1.3}$$

$$defI = \sin(\theta) F a_T \tag{1.4}$$

$$defII = \cos(\theta) F a_T \tag{1.5}$$

$$div = -\cos(\theta) F a_T + 2a_R \tag{1.6}$$

$$curl = -\sin(\theta) F a_T - 2r_R \tag{1.7}$$

From the formulae, it is clear that first-order surface structure and the 'specific transverse velocity' influence all four components (based upon the formulae given in Harris, 1994). As mentioned previously however, *Div* is also dependent upon the 'specific radial velocity' and *Curl* on the 'radial rotation velocity'. Note also that, based upon this local analysis, deformation magnitude is influenced only by the transverse translation and the slant magnitude. This also implies that neither the rotation components nor the specific radial velocity component is able to yield information about surface structure.

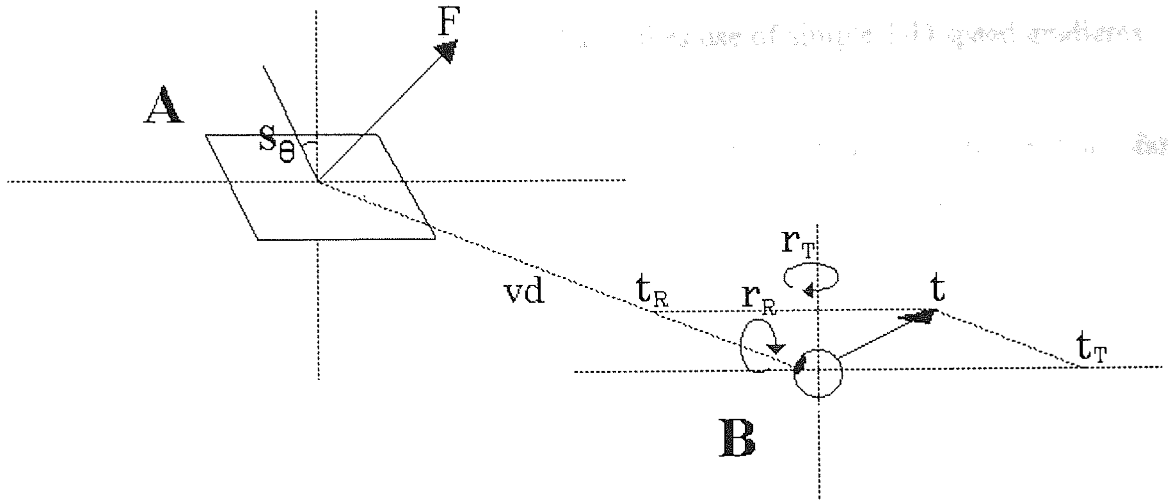


Fig. 1.3. Analysis of observer motion, relative to a stationary surface. **A**, a local patch of planar surface at a tilt of 90°, with the top slanted away by S_θ (floor-like slant) and, **B**, a point of observation with an instantaneous velocity of translation, t, and angular velocity of rotation, r. Both velocity vectors are shown, decomposed into two orthogonal components: the radial components (t_R, r_R) and the transverse components (t_T, r_T). The translation and rotation components denoted by the subscript, R (radial), are oriented upon an axis coincidental with the line of sight (illustrated here with a broken line and having a distance of, vd). The two axes that the transverse components (r_T and t_T) are shown to be orientated upon are orthogonal to each other. This is of course not compulsory: it has been made this way in this particular example to reiterate a point made in fig. 1.1 that the transverse component of translation could be made to cancel out locally by the addition of a transverse component of rotation (derived potentially from eye movements). F, denotes a vector, perpendicular to the surface. Its length represents the magnitude of slant and it orientation represents its tilt. (Modified version of figure 4 from Harris, 1994)

1.2.4 Slant and tilt estimation: derived from 2-D deformation gradients?

While optic flow fields contain potentially useful information for the control of observer locomotion (e.g. the direction of heading and time to contact) as well as object/surface structure, this thesis will primarily address the latter. Moreover, only the first-order properties, of slant and tilt, are considered.

If the filtering of optic flow in human vision involves an analysis of the differential invariant components, then it seems apt that the outputs of deformation mechanisms such as *DefI* and *DefII* would be directly involved in the estimation of tilt and slant. If the combined outputs of the deformation mechanisms could be subtracted from the outputs of the *Div* and *Curl* mechanisms it would permit estimation for two of the scene-based motions: translation velocity along the line-of-sight (t_R) and angular velocity around the line-of sight (r_R).

It therefore seems likely that, under such a system, surface slant and tilt would be derived from the two orthogonal deformational components (employing mechanisms with RFs similar to those of *DefI* and *DefII* would lead to intuitive vertical and horizontal cardinal axes for tilt). Of course, an alternative model for slant and tilt estimation, and possibly the one most favoured by introspection, is one that makes use of simple 1-D speed gradients.

Meese *et al* (1995) obtained empirical evidence supporting the hypothesis that slant magnitude estimates are derived from an analysis of deformation components as opposed to 1-D speed gradients (acting alone) or the differential invariants, *Div* and *Curl.* Subjects matched representations of pictorial slant (a perspective projection of a wire-frame cube) to ones derived from various types of motion stimuli. Eight stimuli types were tested in total; all were vector summed with a component of horizontal translation. They contained either 1-D speed gradients (horizontal shear, vertical shear, horizontal compression, vertical compression; see fig.1.4) or vector summed combinations of these 1-D gradients. These combinations formed 2-D speed gradients, generating global approximations of pure *Div*, *Curl*, *DefI* and *DefII.* Because the method of stimulus generation meant the actual stimuli were just approximations of these differential invariants, these approximations will be referred to hereafter by the notation *Rad, Rot, Def1* and *Def2*, respectively. In two of the experiments, slant estimations were made about a horizontal axis (90° tilt) and in another, slant estimations were made about a vertical axis (0° tilt). In the experiments investigating slant impressions about a 90° tilt, four permutations of shear gradient were tested: horizontal shear, vertical shear, horizontal-vertical shear (*Def1*) and horizontal+vertical shear (*Rot*). The four stimulus types are illustrated in the upper set of diagrams of fig.1.4. Perhaps unsurprisingly, the horizontal shear alone and *Def1* stimuli yielded strong slant impressions. However, the *Rot* stimuli did not, despite containing a horizontal speed gradient like *Def1*. In addition to this, the condition containing no horizontal shear – vertical shear alone – did induce robust impressions of slant at a tilt of 90°. These results support the notion that a processing hierarchy exists that employs a stage of deformation extraction, and that it is this stage which is responsible for the derivation of slant estimates. The authors of this paper draw parallels between this study and that of Howard & Kaneko (1993) who found similar results in favour of a deformation analysis of disparity shears - the analogous stereoscopic stimuli.

The third experiment of Meese *et al* (1995) looked at slant estimates made about a vertical axis (tilt of 0°), and tested the following compressive gradients: horizontal compression, vertical compression, horizontal–vertical compression (*DefII*) and horizontal+ vertical compression (*Rad*). These four stimulus types are illustrated in the lower set of diagrams in fig.1.4. Again, the results largely support the hypothesis that a deformation analysis underlies slant estimation. Horizontal compression, *DefII*, and vertical compression conditions yielded robust slant impressions. However, for two out of the three subjects, the *Rad* conditions also seemed to produce reliable estimates, a finding that cannot be wholly accounted for by Koenderink's local deformation analysis. The flow fields that were generated in this condition (*Rad+Trans*) were not wholly inconsistent with instantaneous flow fields, which can be generated by slanted surfaces in the real world: a slanted surface (with a tilt of 0°) approaching along the line of sight.

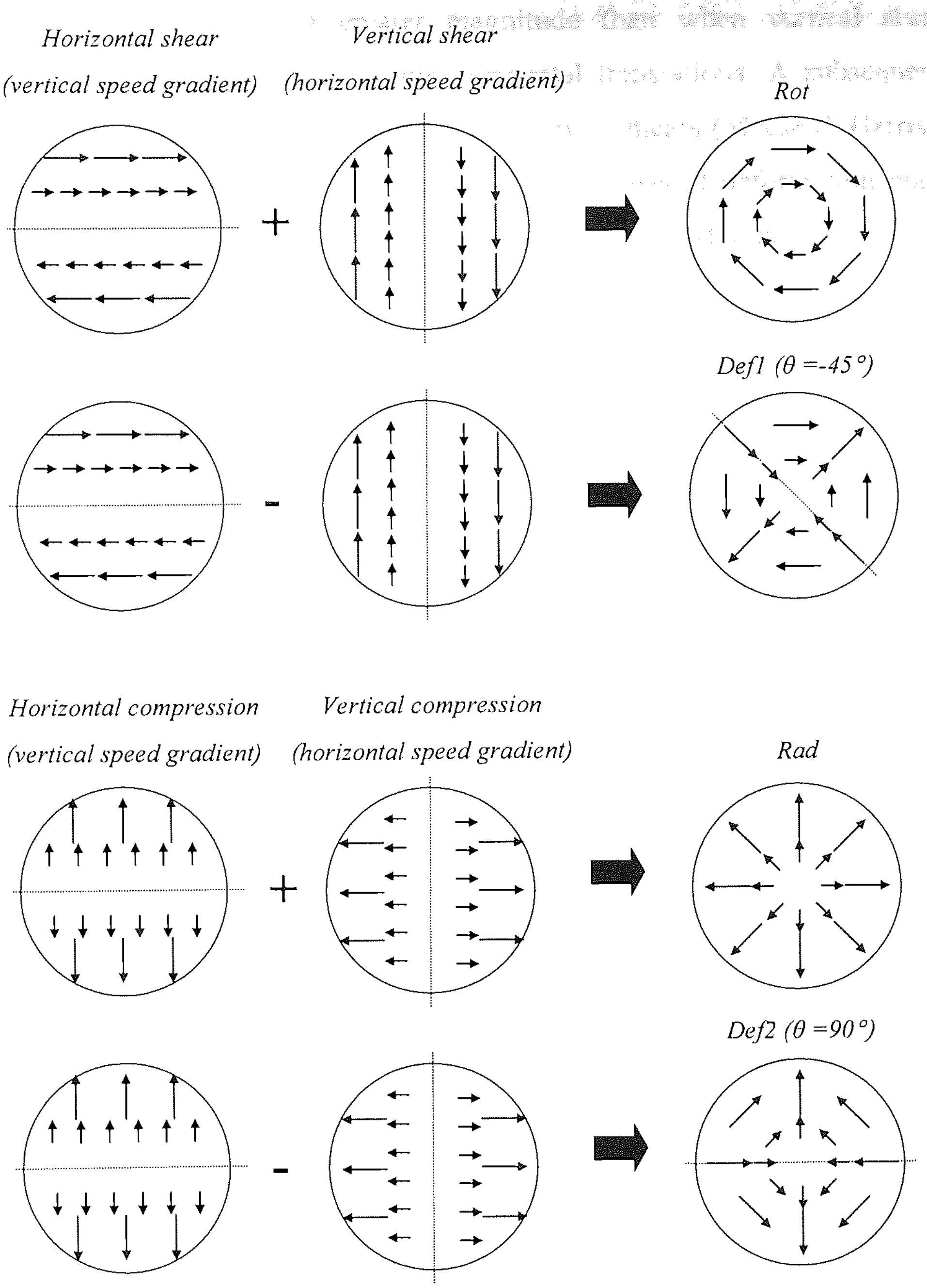


Fig. 1.4. Schematics showing how approximations to the four differential invariants, *Curl*, *Defl*, *Div*, and *DefII* (as proposed by Koenderink, 1986) can be constructed from 1-D vertical and horizontal speed gradients. Each configuration illustrates an instantaneous velocity vector field. Taking the sum of the orthogonal 1-D gradients results in 2-D gradients belonging to the generic 'spiral' group (e.g. *Rot* and *Rad*), while taking the difference between the two, results in gradients belonging to the generic 'deformation' group (e.g. *Def1* and *Def2*). The lighter lines that are included on the deformation and 1-D speed gradient schematics indicate the axes of negative compression. The resultant 2-D gradients will therefore generate dots that travel, on average, twice the distance of those that are generated by 1-D speed gradients. Note that the illustrative arrows attempt only to describe relative dot speeds and not absolute dot speeds (after Meese & Harris, 1995).

1.2.6 Slant and tilt estimation: A preceding stage of 1-D shear processing?

An additional anomalous finding in the study of Meese *et al* (1995) was that when horizontal shears and compressions were summed with a horizontal translation they produced slant impressions of a greater magnitude than when vertical shears and compressions were summed with the same horizontal translations. A subsequent study confirmed the asymmetry between horizontal and vertical shears (Meese & Harris, 1997). If slant magnitudes were determined exclusively on the basis of deformation magnitude, the orientation of the shear/compression should have made no difference to the impression of slant magnitude. The fact that it did suggests that there is an initial stage in which differently orientated 1-D shear gradients are processed before being compared in order to establish estimations akin to deformation.

1.2.6 The relationship between deformation, specific transverse velocity and surface slant

As deformation can be expressed as the product of the surface vector magnitude ($/F/$) and the specific transverse velocity (a_T), with $/F/$ being given by tan(S_ϕ), slant magnitude can be derived using the following expression:

$$S_\phi = \arctan\left(\frac{deformation}{a_T}\right) \tag{1.8}$$

Braunstein (1968) demonstrated that perceived slant (denoted, S_ψ from hereon) does not match the slants simulated by motion (S_ϕ) or texture cues alone. Instead, he showed that it is drastically underestimated. Braunstein made a comparison between perceived slant elicited from texture alone cues and that elicited from the same texture displays but with a homogeneous translation (equivalent to a motion indicative of a 0° surface slant) added. The slant impressions from the texture displays were found to be attenuated with the addition of the motion. It is likely therefore, that the absence of a depth cue does not contribute to an impression of flatness; rather, when the cue is absent it is simply not considered. Only when present is a depth cue able to act in concordance or conflict with other depth cues. From this result, perceived slant might be expected to be a complex function of simulated slant however, this does not appear to be the case. Results from display conditions using various combinations of motion and texture cues (simulating either 0°, 20°, 40 or 60°) found that perceived slant is reasonably well described by a linear weighted summation of the two cues. There is still a general underestimation however.

Harris *et al* (1992) obtained near veridical slant estimation when another pictorial depth cue, shape, was added in concordance with texture and motion cues. The underestimation for perceived slant from motion-only cues is, therefore, likely to be the result of both a lack of saliency (because of a lack of cueing) and the presence of conflicting cues to flatness. In the case of RDK stimuli, these conflicting cues to flatness could arise from the fact that all dots are of a constant size or that dot densities are, on average, uniform. Whatever the cause of this underestimation, a number of studies have confirmed that the relationship between perceived and motion-simulated slant is, to a first approximation, linear (Harris *et al*, 1992; Freeman *et al*, 1996a):

$$S_\psi = m_\psi S_\phi \tag{1.9}$$

The perceptual parameter, m_ψ, represents the gain associated with motion gradient cue. Acknowledging the physical relationship between deformation (*Def*), translation perpendicular to the line of sight (a_T), and slant angle, Freeman *et al* (1996a) went on to propose a model similar to the following:

$$S_\psi = m_\psi \arctan\left(t_\psi \frac{Def}{a_T} \right) \tag{1.10}$$

Where, t_ψ is a parameter representing the ratio of gains for the deformation and transverse translation estimates.

Freeman *et al* have provided empirical support for this perceptual model (Freeman *et al*, 1996a; Freeman & Fowler, 2000). The study by Freeman & Fowler (2000) showed that observers could make unambiguous slant estimates from a patch of relative motion translating across the screen whilst being tracked by smooth pursuit eye movements (note that the translation component was always perpendicular to the speed gradient of the relative motion). To a first approximation, the translating patches of relative motion did simulate the kinds of motion flows generated by real translating slanted surfaces. Locally however the retinal flow field should have shown no signature of the lateral translation because of it being completely annulled by the eye tracking, leaving only relative motion. If slant estimations were based only upon retinal flow then reliable estimates would not be expected because relative motion alone is insufficient for fully describing slant (see eqns 1.4 & 1.5). What this finding seems to confirm is that observers can make use of the proprioceptive feedback (in this case from eye movements) for slant estimation.

The fact that slant can be derived from independent measures of translation and relative motion components validates one aspect of the Freeman *et al* (1996a) model – that at least under certain conditions, slant judgements can be derived from independent estimates of translation (from proprioceptive feedback) and relative motion (from retinal flow).

The Freeman & Fowler (2000) study also revealed an Aubert-Fleischl phenomenon for surface slant. The phenomenon occurs when smooth pursuit eye movements are used to annul allocentric translations as in the case of tracking. However, these are known to underestimate the magnitude of translation to a greater extent than in 'fixed eye' conditions, in which estimates must be derived from the retinal flow field alone. If slant estimation were based upon these same translation estimates, then a tracked surface would be expected to be perceived as more slanted in the tracked condition than the 'eye fixed' condition (because of eqns. 1.8/1.10). Freeman & Fowler (2000) found exactly this, and went on to show their model could well describe the data, with t_ψ being capable of fully capturing the underestimating effect the Aubert-Fleischl phenomenon.

Speculation that surface information can be derived from a local deformational analysis has already been covered (Koenderink, 1986), but how might such an analysis be carried out biologically? Extraction could, conceivably, be derived from mechanisms with RFs selective for the orthogonal field structures of the two differential invariants *DefI* and *DefII* (fig.1.4) or from pairs of orthogonal 1-D mechanisms, as suggested by Meese & Harris (1997). Using responses from mechanisms with differently orientated RFs like these provides a basis from which first-order surface descriptors such as slant and tilt could be determined. In the following section a general overview of motion processing in the primate brain is provided. Research is included which bears a direct relevance to the above question of how surface information might be encoded from motion cues alone.

1.3 Hierarchical processing of motion

The motion processing hierarchy proceeds from simple local retinotopic motion estimates at the early stages to complex global estimates and interpretations at the latter stages.

The initial encoding of motion occurs in V1 simple cells. Their direction selectivity is achieved through an appropriate combination of spatially and temporally sensitive receptive fields (RFs). Some complex cells perform non-linear summation of the outputs from simple cells with overlapping RFs and a common orientation selectivity. This ensures responses with a degree of positional invariance (Adelson & Bergen, 1985).

After V1, processing proceedes through the areas of the dorsal stream, including the specialised motion processing areas, MT and MST and on to areas within the parietal cortex. The most significant connections are shown in fig. 1.5. Note that MT also shares connections with areas other than V1 and V2. For example, sub-cortical areas and area V4 of the ventral steam are connected to MT. In addition, many of the connections that are shown also have strong reciprocal connections.

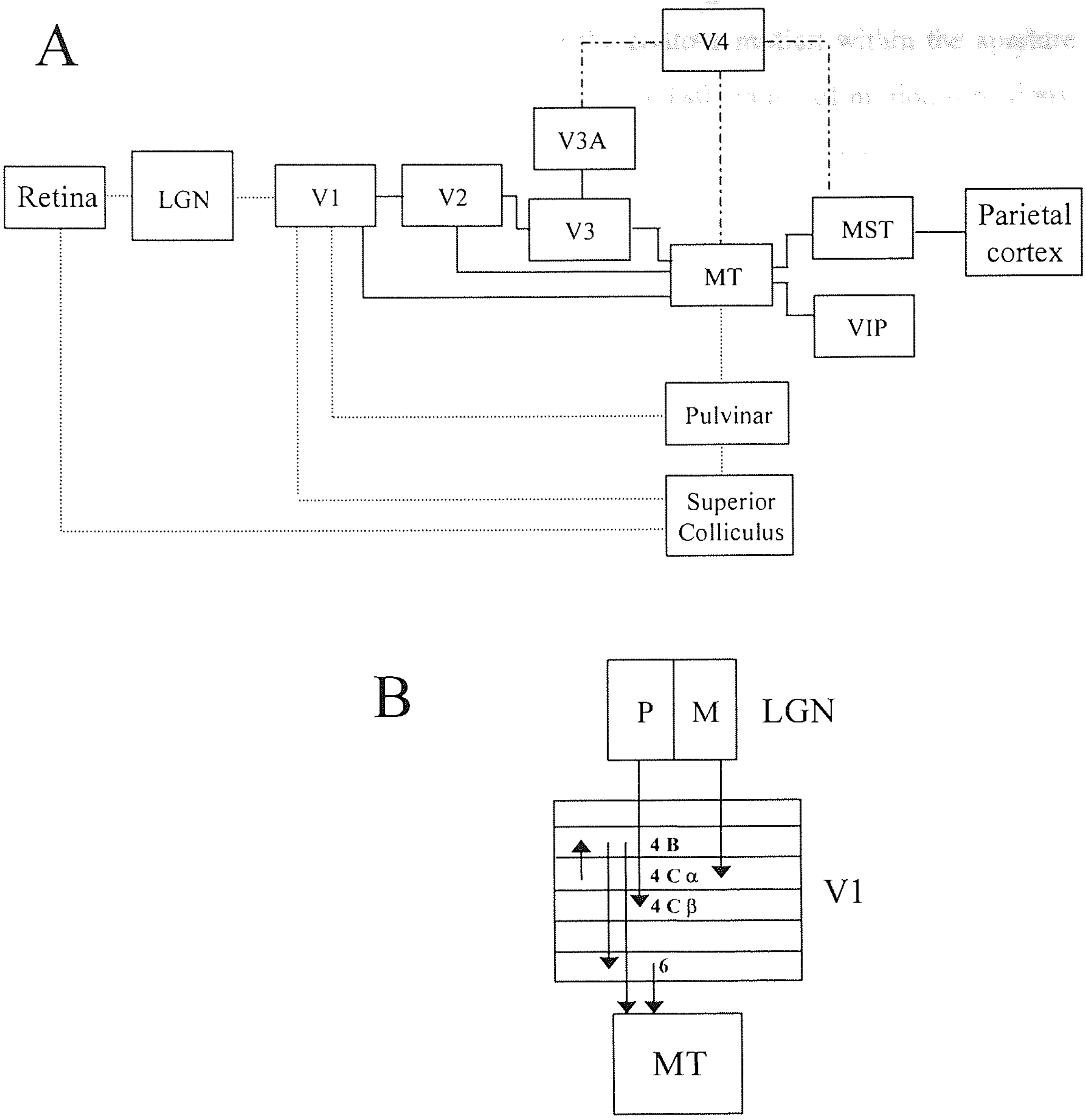


Fig. 1.5. Two schematics showing some of the most prominent connections with areas; MT and MST. (A) Cortical and sub-cortical areas with a potential key importance to the motion processing hierarchy. Solid lines denote connections between areas of the dorsal stream; dotted-dashed lines denote connections with V4 of the ventral stream and the dotted lines denote connections with sub-cortical areas. (B) A more detailed description of the forward flow of processing from LGN to MT via V1. P and M compartments of the LGN represent the parvocellular and magnocellular cell groups respectively. Layers 4B, 4Cα and 6 of V1 are also labelled (this review is based upon information obtained from Snowden, 1994, Crick & Koch, 1995, and Bruce *et al*, 2003).

1.3.1 Solving the aperture problem

At some stage, the local motions that are extracted at V1 need to be combined. There are two reasons for believing this occurs. One comes from the fact that the local motion signals are ambiguous with respect to the real-world stimuli from which they originate. This limitation can be visualised if the movement of a coarsely textured object viewed through tiny apertures is considered. The apertures permit only a partial viewing of the whole object, and if the apertures are small enough, just a single contour of its texture may be visible at one time. If this were the case then the contour motion within the aperture could conceivably have been derived from one of a near 180° range of motion directions. The ambiguity that emerges here is often referred to as the aperture problem. To establish the true motion direction of an object, the motions of other apertures covering neighbouring areas also need to be considered.

To obtain veridical estimates of object direction, the visual system might just consider the responses from cell types with 2-D spatial selectivities (e.g. those with end-stopped RFs). The alternative possibility is as discussed above, that local 1-D motion estimates are combined (Adelson & Movshon, 1982).

Welch (1989) provides evidence in favour of a solution based upon the integration of local 1-D motions and therefore advocates a solving of the aperture problem. Welch obtained speed discrimination thresholds for both grating and plaid stimuli and found peak sensitivities to occur for higher speeds with plaid stimuli than with grating stimuli. When the plaid data were re-plotted in terms of their component grating speeds, the peak sensitivities appeared to be in close correspondence with those of the grating stimuli, suggesting that the locus of limiting noise lay at an initial stage where component gratings are extracted. Exactly how component 1-D signals are integrated remains a topic of ongoing research but probably depends to some extent upon the exact stimulus parameters being used.

A control experiment carried out by Welch indicated that these plaid representations superseded those of its component gratings, so confirming that the previous performances could not have been attributable to the observers ability to selectively monitor single components (an potentially opposing view comes from Watanabe (Watanabe, 1995; Watanabe & Miyauchi, 1998)).

1.3.2 Spatial summation of local signals

Another reason to believe that local motion signals are combined at a latter stage comes from our ability to deal with motion signals that locally, are very noisy. Accurate judgements can be made concerning the presence of coherent motion and its direction, in the presence of noise. To permit such performance, it is necessary for subjects to sample numerous motion signals over an extended area or period of time, and combine them in some manner. The principles that govern the grouping of distributed local signals over an area of the visual field appear to differ depending upon the nature of the coherent global motion.

It is well established that different types of coherent motion exhibit different thresholds. For example, Freeman & Harris (1992) examined minimum speed detection thresholds for rotation, radial, translation and random motion stimuli, and found translation and random motion stimuli to have poorer sensitivity to speed than the two more complex motion types (see fig.1.4 for schematic depictions of rotation and radial). De Bruyn & Orban (1990) found an opposite pattern of results when thresholds were measured in terms of signal-to-noise (S/N) ratios (see section 2.8 for a full explanation of this measure): S/N thresholds for expansion/contraction were higher than for translation. An anisotropy has also been reported for radial motions (Edwards & Badcock, 1993), in which sensitivity was found to be higher for contractions than for expansions.

Differences in sensitivity between the different types of coherent motion, like those described above, could be easily explained if the visual system employed specialised mechanisms configured to certain spatial arrangements of motion. For example, a mechanism might be constructed in such a way so as to respond optimally to motion travelling in a particular direction (obliquely at +45° for example) and covering a certain area of the visual field. Other more complex mechanisms might impose further organisational constrains, additional to their extent of spatial sampling. Instead of simply pooling all motion signals of a particular direction, these might be selective for different directions, depending upon on their relative positioning within a sampled area. One that is optimally stimulated by a circular, clock-wise (Clk) motion, with the centre of rotation located at the centre of the sampling area, would be most selective for rightward motions along the lower half of the central vertical, while being most selective for leftward motions along the upper half (see fig.1.4).

Alternatively, a more dynamic processing mechanism, capable of configuring itself to numerous motion arrangements, might be responsible for the coherent motion detection. In this case, thresholds would be dependent upon how well the stimuli were adhering to abstract principles of organisation (such as common fate and symmetry) rather than the specific implementations of specialised mechanisms, as discussed above. If separate mechanisms do exist, each optimally tuned to a specific motion, then the processing of multiple motion flows might proceed independently. De Bruyn & Orban (1993) have suggested that this is not the case. They found that the identification of direction sign for radial and rotational flow is impaired when presented simultaneously under restricted viewing conditions (85ms stimulus duration with a 400ms backward mask). The possibility that this impairment had resulted from a perceptual masking between the two motions was ruled out in a control experiment, in which one of the motions would be pre-cued. The pre-cues eliminated impairments. Taken together, these results seem more consistent with the existence of general-purpose (matched filter) mechanisms capable of processing only a single type of motion flow at a time.

The above findings of De Bruyn & Orban (1993) notwithstanding, the most parsimonious account for the majority of the psychophysical and physiological data that has been brought to bear on this subject favours an account based upon the existence of highly specialised independent motion mechanisms. One of the first pieces of evidence to suggest this came from a study by Regan & Beverley (1978), in which an adaptation paradigm was used. They showed that human sensitivity to expanding and contracting stimuli decreased with prior exposure to the same stimulus. If, however, subjects had been given prior exposure to expansion or rotation, and the sensitivity to contraction was tested, no effect was observed. Adaptation effects are typically attributed to a fatiguing of mechanisms (at one or more stages) involved in the processing of the stimulus. The stimuli that these researchers chose allowed them to rule out the possibility that the adaptation was arising from local motion mechanisms and accordingly, they accounted for their results by proposing an adaptation of specialised radial detectors. In a subsequent study, measuring motion after effects (MAEs), both radial and translational motions were found to elicit after effects (Regan & Beverley, 1979). MAEs are also associated with mechanism fatigue and are believed to occur when mechanism activity influences perception through a larger population code (mutual inhibition between mechanisms might also play a role in the effect). A selective fatiguing of particular mechanisms is expected to disrupt the normal balance of activity, and therefore, even in the absence of a stimulus, perception is temporarily biased until the fatigued mechanisms recover. For example, the selective fatiguing of expansion mechanisms along with an assumption that spiral type motions (i.e. radial and rotational motions along with all the intermediate spirals motions) are uniformly represented, would predict a perceptual biasing towards contraction. While MAEs were observed for both radial and translational motions, larger decay times were observed for the MAEs when radial stimuli were adapted to than when translating motions were adapted to. Again, this favours an account in which these two motion types are processed by different sets of mechanisms.

More recent studies have looked at how increasing the area in which the motion signals reside affects S/N thresholds (Morrone *et al*, 1995; Burr *et al*, 1998; Meese & Harris, 2001b). A key paper by Morrone *et al* (1995) found evidence for linear summation with large field radial, rotational and translation motion. A latter study by Burr *et al* (1998) found that this high level of summation could continue in some observers for areas no smaller than 40° in diameter. To a first approximation, high levels of temporal summation also occur over prolonged periods - in the order of seconds rather than a few hundredths of a second, as tends to be the case for other stimulus types (Burr & Santoro, 2001). These effects persisted even when the signals were presented in discrete, discernable sectors. Moreover, if these signal regions were made continuous (by filling the intermediate blank sectors with noise) observers attributed the coherent motion to the whole field and not just the sectors containing signal. Summation and phenomenology of this kind are consistent with the existence of special purpose mechanisms latter in the processing stream, whose responses are more closely allied to perception (Bex *et al*, 1998). A fuller discussion of linear summation will be reserved for chapters 8 & 9, where a number of subthreshold summation experiments, involving 2-D complex motions (including radial and rotational), are discussed.

While significant performance improvements with increasing area have been found for S/N sensitivities they have not been found for contrast sensitivities (Morrone *et al*, 1995). As the responses of directionally selective cells in early visual areas (e.g. V1 and V2) are known to be contrast-dependent it seems likely that these summation effects were reflecting the properties of early direction selective mechanisms. The S/N improvements with RDK stimuli demand a more global analysis and are therefore likely to be reflecting the behaviour of higher-level mechanisms.

1.3.3 Complex motion extractors or complex motion templates?

Work has been carried out in an attempt to further describe these special purpose mechanisms. Hypothetically, they could operate in one of two ways. They might, for example, only detect motions of a particular configuration (e.g. expansion flows); any deviation from this configuration would fail to excite the mechanism in question. These will be referred to as templates for global motion. Alternatively, mechanisms might have preferred configurations of motion to which they are most sensitive, but deviations away from these lead only to gradual decreases in the mechanisms sensitivity. These will be referred to as extractors for global motion.

Adaptation and summation experiments have estimated a broad direction tuning for complex motions of the spiral group: the half-height bandwidths were estimated to be between 40-50° (Snowden & Milne, 1996; Meese & Harris, 2001a; Meese & Anderson, 2002). The implication here is that the mechanisms that underlie detection act like extractors rather than templates.

A question that leads directly on from this is how many differently tuned extractors are present in the visual system? The minimum number needed to encode all motion types depends upon the tuning of the direction bandwidths. For 2-D complex motion, attention has primarily been focused on the number of mechanism types that are used to encode the motions belonging to the spiral group. Morrone *et al* (1999) measured S/N thresholds for the detection and direction discrimination of many pitches of the spiral group motions. For the detection data they found sensitivity to be highest for the cardinal flows of radial and rotation, and lowest for the intermediate 'true' spiral flows. Principle component analysis was carried out and it revealed that much of the variance in the data could be accounted for by a set of cardinal mechanisms, tuned to the radial and rotational pitches. A subsequent study by Burr *et al* (2001) employed summation and masking techniques to test this cardinal axes hypothesis. The summation experiment entailed measurement of coherence thresholds (for direction discrimination) to pairs of orthogonal motions, both when presented alone and when presented together. Thresholds were generally lower when they were presented together, however these performance advantages were smallest when the two orthogonal motions were cardinal flows. This result is in accord with their cardinal axes hypothesis: the cardinal flows would excite entirely different mechanisms while both of the orthogonal, non-cardinal flows should stimulate a set of common cardinal mechanisms. A similar finding was obtained from their masking experiments ('adjacent masking') in which masks of cardinal flow were found to be most effective. However, there are other reports suggesting that additional sets of mechanisms, tuned to intermediate pitches, can better account for (a) the summation of complex motion components (Meese & Harris, 2001a; Meese & Anderson, 2002) and (b) complex motion adaptation (Snowdon & Milne, 1996).

In order to gain an understanding into how, algorithmically, selectivity for complex motion is achieved, it might be worth looking at the possible physiological substrate underlying higher-level motion processing. In the following subsection, a brief review of the specialised motion areas, MT and MST, is given.

1.3.4 Overview of two specialised motion areas

Medial temporal area (MT): an overview

MT is a major extrastriate area and part of the dorsal processing stream (see fig. 1.5. The human analogue, and that of Old World monkeys, is not actually contained within the mid temporal area. However, for the purposes of this thesis, the area will be referred to using MT as a generic name, irrespective of the primate species being considered. The area receives significant afferent connections from V1, V2, V3, and V4, with the magnocellular pathway (that originates in the LGN) appearing to be particularly influential (see fig.1.5). Unlike areas of the ventral processing stream however, its activity is not solely contingent upon V1 innervations. To a lesser extent, it is driven by input from the pulvinar via the superior colliculus. Significant efferent connections are also made to MST and other areas in the parietal cortex, as well as back to earlier areas.

MT seems to be highly specialised with respect to motion processing - a capability, which is probably afforded by the significant magnocellular input it receives (Zeki, 1974). The majority of its cells show selectivity for both motion direction and speed. In addition, this selectivity tends to be based upon perceived motion; irrespective of whether it is derived from first-order luminance defined motion or from exclusively second-order contrast defined motion. The directional bandwidths of these cells have been reported to be ±44° (Rodman & Albright, 1987), remarkably similar to estimates made from psychophysics (Raymond, 1993; Meese & Harris, 2001b). A retinotopic organisation is also apparent, but is anatomically less regular than that seen in preceding areas (Orban, 1997). This organisation is also maintained in its connections with V1, V2, V3 and V4. Some evidence suggests that different layers have different speed selectivity and it also seems probable that cells with opposing direction selectivity juxtapose (Snowden, 1994).

A number of insightful studies have suggested a close link between motion perception and the activity of MT cells. Newsome & Pare (1988), using a neurotoxin, selectively killed certain areas of macaque MT. This resulted in an increase in behavioural S/N thresholds, but only for quite specific areas of the visual field. These findings are in accordance with the idea that MT does retain some form of retinotopic organisation, and that the cells within it are pivotally related to perception. A follow up study by Newsome and colleagues measured individual cell responses and found them to be sufficient to explain observer performance in direction discrimination tasks (as before, via measurement of S/N thresholds). The third study of particular interest observed the effects of intervention on behavioural responses (Salzman *et al*, 1990). Directionally sensitive cells of MT were artificially stimulated, leading to predictable biases in the monkey's decisions about the displayed direction of motion. Together, these studies go some way to confirming a causal link between MT cell activity and motion perception. Conversely, the activity of V1 cells does not show such close correlations with perception. This has become particularly apparent from studies that have used random dot displays that were constructed from locally paired dots with opposing direction. To observers, these displays do not produce perceptions of global transparent motions, instead they appear noisy. Correspondingly, cells in MT tend to show a reduced response to these displays relative to displays constructed from unpaired dots. This is not the case with V1 cells however. They respond in the same manner irrespective of whether the displayed dots are paired (observer perceives no global motion percept) or unpaired (observer perceives transparency) (see Bruce *et al*, 2003. pp.232-236, for a brief review).

Consistent with the idea that MT cell activity closely reflects perception, many cells (~60%) have been found to respond to veridical real-world velocities rather than the component velocities that are extracted by V1 cells. It is at this point in the dorsal processing stream, therefore, that the aperture problem first shows signs of being solved. First- and second-order motion cues also seem to be integrated at this level. Pack *et al* (2001) found that selectivity to these real-world 'pattern' velocities take time to develop (100-150ms). Furthermore, when the animals (macaque) are anaesthetised this 'pattern' selectivity is drastically decreased (reducing the proportion of cells that were responsive to pattern motions down to ~7%). Both these observations suggest that top-down processes might also be crucial in the integration of motion signals. Other findings have also supported the view that top-down attention can have dramatic modulatory effects upon MT responses (Treue & Trujillo, 1999; Watanabe et al, 1998).

As processing hierarchies are ascended, the RF sizes of cells generally become larger. Consistent with this rule of thumb, MT field sizes tend to be larger than in preceding areas, V3, V2 and V1. If this early processing hierarchy is envisaged as being constructed from a network of feedforward projections, these size increases suggest a pooling of local finer-grained signals. Indeed, for MT cell responses to be capable of reflecting real-world velocities - as shown by Pack *et al* (2001) - some form of local component motion pooling must take place.

Velocity and the level of motion coherence also seem to be key factors in determining MT responses; the effects of luminance and contrast, however, quickly saturate. These characteristics fit well with the hypothesis that the motion pooling observed in the area summation studies of Morrone *et al* (1995) were subserved by mechanisms within MT. However, the average RF size of MT neurons is only about x3 of the directionally sensitive V1 cells, and is far smaller than the pooling areas needed to explain the areas of summation that were reported by Burr *et al* (1998) (subtending areas in excess of 40°).

MT cells with modulating surrounds: antagonistic & synergistic

The firing rates of many cells in primate MT are modulated by stimuli falling outside the classical RF (CRF). In the absence of CRF stimulation, stimuli in this surround RF (SRF) had no effect upon firing rates (Xaio *et al*, 1997). The size of the SRFs is probably dependent upon top-down factors such as attention. In anaesthetised animals the radii of these SRFs have been measured as being between 7-10 times larger than that of the CRF, while in conscious animals they only seem to be around 2-4 times larger (Buracas & Albright, 1996).

These centre-surround organisations can be divided up broadly into two categories; antagonistic and synergistic. Antagonistic arrangements are constructed in such a way that the surround is optimally selective for motion with the opposite direction to that with which the CRF is selective for. One way in which this centre-surround opponency could arise is through inhibitory connections. When the SRF is optimally stimulated there would be maximum inhibition to CRF; when there is an absence of stimulation the CRF would be released from this inhibition. Synergistic arrangements, conversely, have the surround optimally selective for the same direction as the centre.

Xiao *et* al (1995) have carried out careful mappings of these surrounds and found many to exhibit asymmetrical surrounds as well as radially symmetric surrounds (see fig.1.6). Furthermore, the relative speed tuning functions for the surrounds also varied. Some cells showed monotonic tuning functions (i.e. increasing the surround speed from less than that of the centre to more than that of the centre led to monotonically increasing or decreasing responses) (Tanaka *et al*, 1986). Other cells showed U-shaped tuning functions: firing rates increased as the speeds in the SRF became increasingly different from those in the CRF (irrespective of whether these speeds were slower or faster).

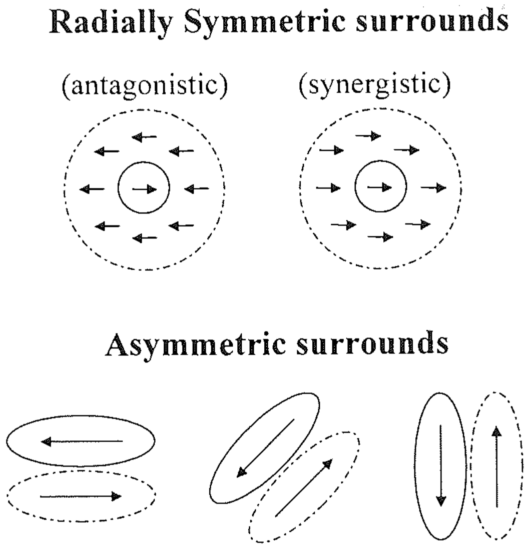


Fig. 1.6. Types of modulating RF organizations of MT cells. Solid outlines denote 'classical' RFs and dot-dashed outlines denote surround (modulatory) RFs. Cells can be divided into two broad categories: antagonistic and synergistic. They can further, but less neatly, be divided into radially symmetric (top) and asymmetric (bottom) arrangements. The lobes of asymmetric arrangements tend to be more elongated.

Medial superior temporal area (MST): an overview

The RFs of cells in MST are much larger than those of preceding areas, with diameters spanning between 10-100°. Consistent with the trend from V1 to MT, the retinotopic organisation of MST, is (anatomically at least) coarser than in MT. Careful physiological mapping of the RFs of cells in the dorsal part of monkey MST (MSTd) revealed many to be selective to frontoparallel translations and complex motions. Those selective for complex motions are primarily of the types that can be described by a spiral pitch (Graziano *et al*, 1994; Duffy & Wurtz, 1991a; 1991b). For example, a cell possessing a RF with a spiral pitch of 0° is one that has selectivity for outward radial (expanding) flows, while one with a spiral pitch of 90° describes selectivity for rotation (Clk) flows. Graziano *et al*'s (1994) study shows that RFs can also be found for intermediate spiral pitches; these MSTd cells therefore represent 'true' spirals as well as the cardinal directions of radial and rotation motion, although a general bias towards expanding flow has been observed. These properties are in close accord with those of the speculated psychophysical mechanisms, which were also shown to have large RFs, processing independence and preferences for complex motions of various spiral pitches (Morrone *et al*, 1995; Snowden & Milne, 1996; Burr *et al*, 1998; Meese & Harris, 2001b; Meese & Anderson, 2002).

The most obvious way in which these RFs could have been afforded to the MSTd cells would be through the appropriate pooling of 1-D motion signals from MT. This simple account appears, at first sight, to be hindered by the fact that many of these MSTd cells also exhibit invariance as to the positioning of their preferred flows within their RFs (Duffy & Wurtz, 1991a; 1991b). A theoretical account put forward by Zhang *et al* (1993) shows that if the MST cells are made to adhere to a certain set of criteria, the vector fields to which they are they are tuned can be extracted irrespective of the stimulus positioning within their RFs. In their model, the MST units take weighted inputs from MT units. They modelled these MT responses with the following equation:

$$a = cv\cos(\theta - \phi) \tag{1.11}$$

Where, a is the response of the hypothetical MT cell, c represents a linear speed tuning constant, v the stimulus speed, θ gives stimulus direction and ϕ represents the units preferred direction.

Below are listed some of the key properties that are implied by Zhang *et al*'s model, along with the physiological and psychophysical evidence available, for and against these properties.

Broad cosine direction tuning

Eqn.1.11 stipulates cosine direction tuning, implying broad direction bandwidths (±60°) for the MT units. Correspondingly, the complex motion units will operate as extractors rather than templates (templates would have been implied by a pooling of MT units with narrow-band tuning). The benefit of using cosine tuning functions is that a pair of MST cells with orthogonal RFs are capable of decomposing 2-D velocity flow fields into their constituent components. For example, a 'true' spiral could be decomposed into its radial and rotational components.

As mentioned in one of the previous subsections, behavioural data has estimated the direction bandwidths of the psychophysical mechanisms to be broad. These estimates are somewhat narrower than cosine tuning though, as are MT cell bandwidths (±45°) (Britten & Newsome, 1998). If MST mechanism responses were derived from MT inputs, similar bandwidths would have been expected.

Inhibited cell responses

In the model the cosine function is not rectified and so predicts that complex motion extractors should be inhibited by their antipreferred stimuli. Furthermore, this inhibition should be of an equal magnitude to that of the excitation elicited from preferred stimuli. Zhang *et al* did acknowledge that the magnitude of inhibition exhibited by MT cells is generally not as large as the excitation they exhibit.

Global motion opponency

With cosine tuning, a cells antipreferred direction would be 180° to that of the preferred direction. The model therefore predicts that superimposed global motions of the same type but opposing direction would prevent detection based upon global motion cues. There seems to be a paucity of evidence to suggest that this is the case. Meese & Harris (2001a) did find evidence in favour of such global opponency for deformation flows: S/N thresholds increased when deformation flows of an opposite direction were introduced, but for other global motion flows (i.e. vertical translation, horizontal translation, radial and rotation) they found no evidence for such opponency. Furthermore, for many stimulus configurations involving dots that are travelling in opposite directions, subjects report perceptions of transparency (i.e. global motions travelling simultaneously over each other, a phenomenon not expected if inhibitory tuning existed. Note that transparency is unlikely to arise directly from local motion detection, because, as mentioned previously, in cases where directionally opponent dots are paired within local regions of space, transparency is not observed (Braddick & Qian, 2001).

Another study to argue against the existence of a global motion opponency measured coherence thresholds, both for displays in which the noise was random and for displays in which a proportion of the noise dots were forced to take on directions opposite to that of the signal (Edwards & Nishida, 1999). No difference between the two was found, despite one condition having a greater amount of motion energy in the direction opposite to that of the signal.

Validity of the Zhang *et al* (1993) model

In a comprehensive investigation, Duffy & Wurtz (1991a; 1991b) looked at the RF properties of MSTd cells in some detail and found that many responded to more than one motion type (translation, rotational, and radial flows were tested). They also found that the amplitude of the response often depended upon the position within which the stimulus was presented in the RF. These findings are not consistent with a model of MST neurons, such as the one described by Zhang (1993).

It is clear therefore, that there are significant discrepancies between some of the implications of Zhang *et al*'s (1993) model and the empirical data supplied by psychophysics and physiology. It is unclear to what degree these assumptions can be relaxed before the positional independence breaks down completely, but Zhang *et al* do cite previous work in which positional invariance was obtained with a feedforward network receiving inputs from MT-like units with a narrower direction tuning.

The models simplicity makes it appealing and, at the very least, it illustrates that a positional invariant vector-field extraction can, in principle, emerge from a fixed weight, feedforward architecture.

The effects of aperture configuration upon the perception of motion

The connections of MT and MST to other areas such as V4 may have an importance upon the perception of motion. They may, for example, mediate phenomena described by Watanabe (1997), in which subtle changes in aperture configuration are shown to affect motion interpretation. In this study, two configurations were tested. In one, two partially superimposed squares formed three regions: two L shaped regions and one intersected square region. The intersecting region had the appearance of being the result of a transparency of the two squares (transparency configuration). In the second configuration, again there were two L shaped regions and a square region arranged in a similar manner to the first, but here the L-shaped regions were slightly misaligned with the corners of the square. This led to the square appearing not as the product of two intersecting transparent squares but as a distinct square occluding the corners of the two others (occluding configuration).

The same random dot motions were added to both configurations: the two L-shaped regions contained the orthogonal motions of 0° and 90° while the middle square region contained motion at 45° constructed from vector summing the two L-shaped motions. In the transparency configuration, the 45° motion of the central square appeared to decompose into the same two components contained in the L-shapes, while in the occluded configuration the motion at 45° did not appear to decompose. This shows that not only can subtle alterations to configuration drastically affect the perception of form, they can also drastically affect the interpretation of the motions within them.

1.3.5 Possible physiological mechanisms involved in determining surface descriptors

Evidence for specialised 2-D motion mechanisms in the primate visual system has suggested large RFs, covering areas that exceed 30°. Such large areas of sampling do not appear suited to the kind of fine-grain analysis that is needed to explain human performances in the extraction of surface from motion. Moreover, most of the cells that have been identified as selective for these 2-D flows belong to the generic spiral group and are located in MST, an area that is likely to have relatively poor location specificity. In addition, the RFs of some of these cells exhibit positional invariance, another property that cannot be easily reconciled with a fine-grained retinotopic analysis. It therefore seems unlikely that these global mechanisms are involved in the extraction of surface boundaries and structure. It is more likely that they are involved in the estimation of parameters for use in self-orientation and navigational behaviours.

Recently however, cells exhibiting potentially useful properties for the analysis of surface structure have been found in MT. Because these cells are much smaller than those of MST, they are more suitable for a local analysis, typically having CRFs of around 1° diameter at the fovea (Albright & Desimone, 1987). MT cells with modulating surrounds (SRFs) were introduced previously in section. 1.3.4, and it was noted that they can possess modulating SRFs that are either radially symmetric or asymmetric. Their speed tuning functions can be U-shaped or monotonic (either increasing or decreasing). Cells possessing these qualities make them ideal indicators for tilt/slant, because they can, in effect, act as orientated speed gradient filters. Indeed, Treue & Andersen (1996) and Xaio *et al* (1997) have identified such cells that do exhibit selectivity for velocity gradient.

In one computational study, Buracas & Albright (1996) developed model neurons based upon the aforementioned physiological observations and some necessary additional assumptions. Direction selectivity for both the centre and surround regions were modelled with cosine tuning functions (cf. Zhang *et al*, 1993) and their weighting functions described by Gaussian distributions. The researchers illustrate the biological plausibility of using neurons with radially symmetric SRFs (antagonistic and with U-shaped tuning functions) to compute slant magnitude and note that a property of these neurons is that their responses are invariant to the global translation. They also show that neurons with radially asymmetric SRFs might also be useful. These would permit estimates of tilt as well as slant. The tilt estimations could, for example, emerge from a winner-takes-all network in which the activity of neurons possessing a range of differently orientated SRFs is considered.

As well as providing 3-D structural descriptions (through the first-order components), it is possible that MT cells with SRFs could also be used to indicate surface boundaries (through the zeroth-order components). Gautama & Hulle (2001) have modelled the responses of real MT cells using artificial neurons and acknowledging the fact that variations occur in the SRF structure and local speed tuning. They found that the responses of these neurons could be used to reliably indicate surface boundaries when present but also slant and tilt when boundaries are not present. It is therefore possible that real MT cells are key to both functions.

1.4 Hierarchical processing of disparity

As with motion processing, disparity is processed in a hierarchical manner with increasingly complex and veridical representations subserved by increasingly higher brain areas. Processing specific to disparity is first observed in cells of the upper layers of V1. These cells have binocular receptive fields with selectivities to horizontal disparities. They were first observed by Barlow *et al* (1967, cited in Wandell, 1995) in the primary cortex of the cat and have subsequently been found in the primary cortex of the monkey. The disparity selectivities that these cells exhibit can be categorized into three main types: those with an opponent tuning to crossed or uncrossed disparities and those selectively tuned to zero disparities (Poggio & Talbot, 1981, cited in Wandell, 1995). Freeman & Ohzawa (1990) went on to study the properties of the cells in more detail, describing not only disparity selectivities but also their RFs in the left and right eye. It was found that although the two monocular RFs of these cells tended to be matched in terms of their orientation and size they were often mismatched in their positioning of the excitatory and inhibitory lobes. For example, a vertically orientated disparity sensitive cell with anti-symmetric monocular RFs, each consisting of two lobes, might have a left eye RF structure in which the excitatory lobe is on the left. In the right eye however, the RF structure could dictate the excitatory lobe to be on the rightward side.

How these cells contribute to disparity estimates that can actually be used to generate veridical stereoscopic perceptions does not seem to be a trivial matter. Cumming & Parker (1997) have shown that in macaque, while the activity of these early cells suggest selectivity to horizontal binocular disparities, they do not necessarily correspond to the disparities that are implied by subjective depth perceptions. In their study, they used anticorrelated RDSs. These did not produce perceptions of depth but did elicit activity in some of the disparity sensitive cells of V1. In other words, these cells appear fallible, capable of making erroneous 'false correspondences' between features, even when the global perceptual interpretation suggests otherwise. The RDSs developed by Julesz eloquently demonstrates that the human visual system clearly does have the ability to solve these problems of correspondence and do so purely within a stereoscopic modality (i.e. in the absence of any interactions with other form or depth cues). As discussed previously with respect to motion (section.1.3.1), a subsequent, more global analysis is probably required.

Marr & Poggio (1976) proposed three constraints that could be utilised by such a processing stage. The first was feature compatibility. In terms of RDSs this is simply a rule stating that potential correspondences must match identical features (e.g. a black dot on the left retina can only be matched with a black dot on the right retina). The second is a uniqueness constraint, in which a feature from one retinal image is only ever matched with one feature on the other retinal image. Thirdly, within a surface, there should be good continuity among the disparities of neighbouring features; this corresponds to the typical opacity and smoothness that surfaces have in reality; discontinuities in disparity tend only to occur at surface boundaries. Marr & Poggio went on to implement these constraints implicitly within a binary network by means of inhibitory and excitatory node connections. They found that it could be used to solve the correspondence problem in RDSs and so demonstrated that, by imposing just a few constraints that were consistent with the properties of the real world, the correspondence problem could be solved using a global cooperative analysis.

It is unclear at what stage the correspondence problem is solved for disparity but cell responses that are consistent with cooperative processes, like disparity-capture, have been observed as early as V2 (Bakin *et al*, 2000). In addition, these responses that are induced by cues outside the classical RF, take no more time to develop than the ones resulting from the non-induced disparities within the classical RF (~70ms). In contrast, the earliest point in the processing stream that the correspondence problem is solved for motion appears to be MT.

Using the need to solve the correspondence problem as the sole motivation for invoking an argument for a global analysis might be ill considered however. Models that calculate phase differences directly from the simple cells of V1 (at numerous spatial scales) could be used as a starting point for 'correspondenceless' processes that calculate depth (for a brief introduction to the idea that spatial phase could be used to extract disparity see Bruce *et al,* 2003. pp.180-184). How exactly the model would be implemented physiologically remains unclear, but the very existence of V1 binocular cells with monocular RFs, identical in all but one of the structural axes, does suggest that to some extent the extraction of phase at the early filtering level has a role to play in the calculation of depth (Freeman & Ohzawa, 1990).

Irrespective of the type of process that is employed for making depth estimates, the physiology suggests it to be unlikely that processes capable of calculating disparity differentials occur as early as V1. Imaging studies on humans and physiological recordings from the macaque appear to support this view. Backus *et al* (2001) used RDSs to produce displays containing two disparity planes. The upper and lower behavioural thresholds for interplane disparity differences were determined. They then measured BOLD responses with fMRI when subjects observed various interplane disparity differences, and found the activity of some areas to be dependent upon these differences. V1 did not seem to be particularly sensitive to these differences, unlike V3A whose activity increased as the disparity differences increased from the lower threshold and decreased as the upper threshold was breached. V3A rather than V1 may, it seems, be an area that specializes in the processing of stereoscopic depth and so be a possible loci in which disparity gradients are comprehended. The physiological study of Poggio *et al* (1988) is in accordance with the view that V3A specializes in stereoscopic processing: a higher proportion of cells in the monkey homologue of this area are disparity selective than in other areas like V1. The only area however, where cells have been found to signal 3-D properties from disparity are quite far up in the processing stream, in area TE of the parietal lobe (Janssen et al, 1999).

1.5 Cue-independent mechanisms involved in the processing of 3-D surface descriptors

The previous subsections of this chapter reviewed the possible loci for mechanisms that are involved in the representation of 3-D structure from motion and disparity. Here, some of the evidence that supports the existence cue-independent mechanisms which can encode 3-D surface descriptions is considered. This evidence has come mainly from cue integration studies. Many of these have used tasks in which the subject makes judgements about perceived depth/structure (e.g. Braunstein, 1968; Landy *et al*, 1995). Braunstein (1968) for example showed that the motion parallax cue could be summed with texture cues. The emerging view coming from these experiments is that cues sum linearly, albeit with their weightings being somewhat dependent on the prevailing conditions (e.g. the degree of discrepancy between cues). There have been other approaches aimed at probing the putative integrating mechanisms in other, possibly more direct, ways.

Rogers & Graham (1982), using RDKs and RDSs, constructed displays consistent with corrugated surfaces. They measured amplitude thresholds for sinusoidal structure at a number of corrugation frequencies, when they were defined either by disparity and motion parallax. Similarities were found between the two; both peaked at relatively low corrugation frequencies. Latter they carried out work in which observers were adapted to corrugations defined by motion parallax or stereopsis; this was followed respectively by the presentation of an ambiguous stereoscopic or motion induced corrugation. They found that the adapting stimulus had a profound affect upon the interpretation of the ambiguous test, tending to bias it to the opposite phase (Graham & Rogers, 1982, cited in Howard & Rogers, 2002, pp.472). Both lines of investigation are suggestive of an integrative depth process, which operates on both motion and stereo cues. Bradshaw & Rogers (1996) have found further evidence using a subthreshold summation technique. When this paradigm is used and the problem of depth cue integration is caste in terms of a simple multiple-analyser model (Graham, 1989) - with the depth cues being represented by independent analysers or groups of analysers - the levels of summation can be indicative as to its locus. For example, low levels of summation are usually taken as an indication of the probabilistic summing of analyser outputs at the decision stage. Larger, linear or $\sqrt{2}$ levels of summation are attributed to integrative mechanisms at the sensory stages. Bradshaw & Rogers found summation levels to be significantly greater than those predicted by probability summation and their work can therefore be taken as further evidence in favour of some kind of integrative process for depth cues.

Other depth cues have been investigated and similar results obtained. For instance, a recent preliminary study by Meese & Holmes (2003) has looked at the combination of pictorial depth cues using the same subthreshold summation technique as used by Bradshaw & Rogers (1996). They found consistent levels of summation for contrast and size gradients, greater than would be expected from any summation at the decision stage. However, it is the evidence for interactions between the motion and disparity cues to depth that are particularly strong (Bradshaw & Rogers, 1996). There is even physiological data that is suggestive as to the location of this interaction. Maunsell & Van Essen (1983) have reported MT cells that appear selective for both motion direction and disparity. Further studies that make use of ambiguous structure from motion have gone on to find correlations between the activity of these stereo-motion selective cells and what is perceived (Bradley *et al*, 1998). Although this correlation has been observed, whether such neural activity does actually reflect the processing of explicit 3-D representations, such as slant, remains unclear. It is possible that these cells are merely reflecting the activity of higher-level processes and that it is these that are determining the most appropriate interpretation. Which one of the two coherent 3D structures is perceived might be determined largely by top-down factors like expectation rather than an inherent bistability within MT (Treue, 2003).

1.6 Superior access to stimuli promoting 3-D interpretation?

Freeman *et al* (1996b) have investigated the accessibility that perception has to multiple motion representations. Unlike De Bruyn & Orban (1993) and Meese & Harris (2001b), who superimposed different flow patterns, Freeman *et al* placed multiple motion flows at spatially distinct locations of the display. The motion components that were tested included *Rad*, *Rot* and *Defl*, and the number of discrete motion flows that were present was always four. Within a single trial, all four patches contained the same motion type (e.g. all *Rad*). However, one patch would always contain an opposite sign of the motion to the other three – designated the 'odd-item-out'. Taking the *Rad* stimulus as an example, three of the patches may have contained expanding stimuli (positive *Rad*) while the odd patch by definition would have contained contraction. Equally likely would be the converse situation in which the odd patch contains expansion.

In one of their tasks (4AFC detection), the motion stimuli were displayed for a brief period (300ms) before disappearing. The observer's task was to make a response, indicating which patch had contained the motion with the odd polarity. In another task (4AFC reaction time), the motion stimuli were cycled continuously; here the observer's task was to make a response in as shorter time as possible (speeded response), again by indicating the odd patch while attempting to maintain a constant accuracy. There appeared to be some consistency in the performances across the two tasks. Odd polarity *Rad* stimuli were more rapidly and frequently detected than the *Rot* and *Defl* stimuli. In the detection task performances for these latter two were around chance.

In addition to the three stimulus types already mentioned, data was concurrently collected for a further three motion types: *Rad*, *Rot* and *Defl* vector summed with components of rightward translation. The construction of such stimuli is illustrated in fig.1.7. As in the other three conditions, the odd patch was defined in terms of a polarity switch in the 2-D complex motion component - the translation component always maintained a rightward direction. Again, a similar pattern seemed to emerge from both the detection task and the reaction time task. In the detection task the addition of translation worsened performance slightly for *Rad*, remained at around chance level for *Rot* (possibly improving slightly) and for two out of the three observers, improved performance quite drastically for *Defl*, raising it from chance level to around 50%-90%.

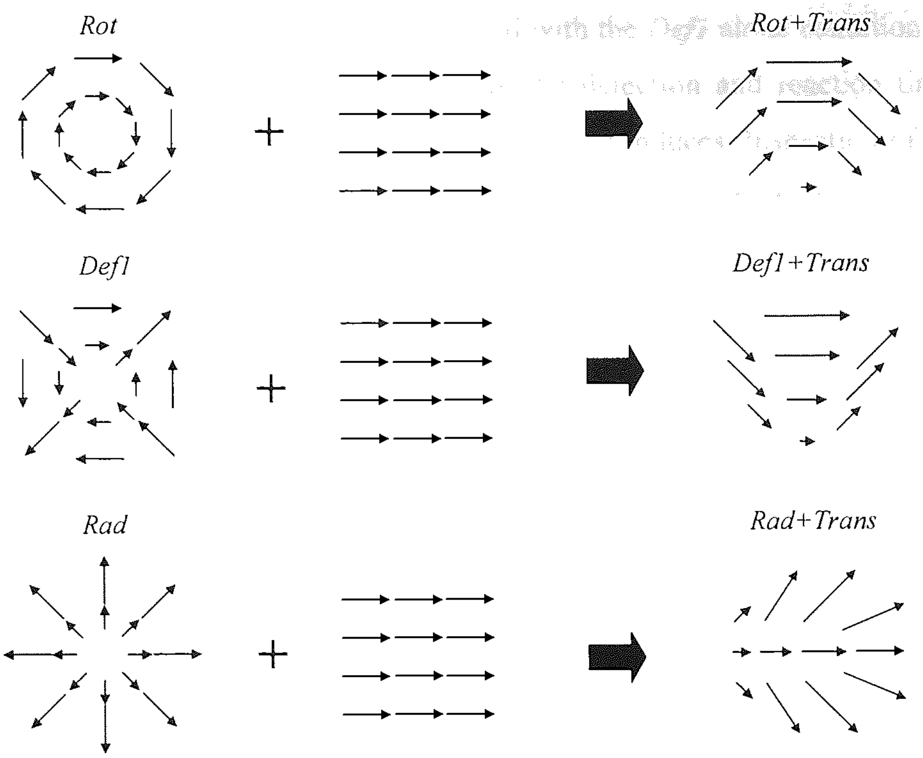


Fig. 1.7. Schematics showing the six stimulus conditions of used in Freeman *et al* (1996b). Assuming local deformation analysis underlies slant and tilt estimation, only the *Defl+Trans* stimuli would provide impressions of unambiguous slant (around the horizontal axis). However, a global deformation analysis might also indicate, to a lesser degree, unambiguous slant for the *Rad+Trans* stimuli.

A third, perceptual rating task was carried out to explore a potential explanation for the above results. Observers rated two 3-D aspects of each stimulus type: the degree of perceived motion-in-depth and the degree of perceived 3-D structure. Unsurprisingly, the *Rad* stimulus was rated high for motion-in-depth. Adding translation degraded this reported impression by a factor of two. Other conditions elicited no motion-in-depth impression. For 3-D structure both the *Rad* and *Rot* conditions were reported to give little to no impression, whereas the *Defl* conditions elicited strong impressions of 3-D structure, consistent with other reports (e.g. Meese *et al*, 1995; Domini & Caudek, 1999).

The authors argue that, with the exception of *Defl* alone, the performance measures from the first two tasks match reasonably well with the pattern of ratings obtained from the perceptual task. They hypothesize that the performance measures were determined primarily by the saliencies of the 3-D representations associated with each of the stimuli. Under this proposal, the superior odd-item detection found with the *Rad* and *Defl+trans* flows would be directly attributable to the strong impressions of motion in depth and 3-D structure, respectively. The poor detection performances with *Rot* flows show that good performance is not a general property of all 2-D flows. To account for the discrepancy between the high 3-D structure ratings obtained with the *Defl* alone condition and the poor performances obtained with this condition in the detection and reaction time tasks, we should consider the kind of 3-D representations that it induces. Inspection of eqn.1.10 and empirical evidence from other studies (Meese *et al*, 1995; Freeman *et al*, 1996a; Meese & Harris, 1997) demonstrates that reliable slant estimates can be derived from flows that contain both *Defl* and *Trans* components. A sign reversal of either the *Defl* or *Trans* component would similarly reverse the sign of slant. However, removal of the *Trans* component would leave an infinite number of solutions to eqn.1.10 and produce a flow field that will induce potentially ambiguous slant. Therefore, with *Defl* alone, reversing the polarity should not be expected to change 3-D impressions in any systematic manner, in which case they could not have been used as effective cues in odd item detection - consistent with the data. This idea will be referred to as the '3-D access' hypothesis.

Fig.1.8 outlines a simple feedforward hierarchical model of 3-D motion analysis. 3-D structural properties including surface descriptor estimates for tilt and slant are derived from the outputs of a second motion processing stage. The second stage entails an extraction of 1-D non-local translation and 2-D complex motion components. Included in the figure are the possible anatomical loci for each of the processing stages. Such a model is broadly consistent with the findings of Meese *et al* (1996), although they note the possibility of a preceding stage in which 1-D speed gradients are computed (Meese *et al*, 1997).

An alternative version of this model would be to have the 3-D analysis derived directly from local 1-D and 2-D mechanisms - a biological implementation of Koenderink's local analysis (Koenderink, 1986), possibly through the use of the MT cells such as those reported by Xiao *et al* (1995; 1997). Harris (1994) points out that while there may be computational advantage in having a local 2-D analysis, with respect to the segmentation of the flow field, there is little evidence to suggest the use of such mechanisms in the mammalian visual system. It should be noted however that one model of 3-D surface detection (Andersen, 1996) does posit a simple analysis of local deformation detectors. Whichever implementation is taken to be true, the key point that fig.1.8 is intended to make is that while the 3-D representations are readily accessible to the decision mechanisms involved in the two performance tasks, the representations of the intermediate stages are not.

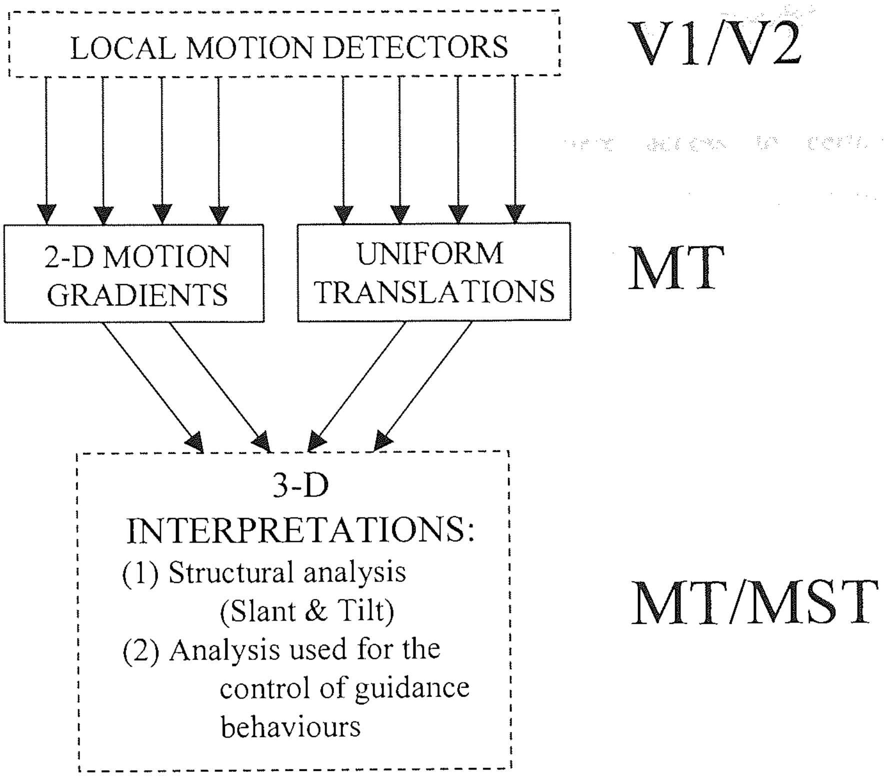


Fig. 1.8. Schematic diagram illustrating a possible hierarchical model for the computation of 3-D properties from motion. Boxes with solid outlines denote processes that cannot be readily accessed in the absence of attentional resources. The corresponding brain areas that have been associated with each of the processing stages are speculative and bear no importance to the rationale of the model.

1.7 Access to high-level and low-level representations

From an ecological perspective it might seem reasonable for decision-making processes to be given unrestricted access to only those representations of immediate behavioural relevance. While representations of preceding stages might still be accessible they probably demand a certain level of focused attention, either to modulate the perceptual processes that subserve these early representations or to enable a direct monitoring of these processes.

The idea of not having perceptual access to early stage analysers has already been reported, for example some very high spatial and temporal frequencies cannot be resolved by observers, but careful psychophysical probing reveals that they are in fact processed to some level within the visual system (possibly, only up to V1) (e.g. He & MacLeod, 2001; MacLeod & He, 1992). Welch's (1989) results (section 1.3.1) also showed that component motions (from a supposed first stage of processing) could not be accessed when they formed plaid stimuli.

Visual crowding is another possible example where access to certain stimulus representations becomes restricted. Crowding describes the phenomenon in which the addition of stimulus elements can lead to a reduction in the observers' ability to identify and describe (individuate) the individual elements. For example, the orientation of a patch of grating, presented parafoveally, might be easily identifiable on its own but, when presented with surrounding grating patches, the task can become impossible. In the past the effect has often been attributed to a lateral masking between the orientation detectors of early vision; this was questioned however, when He *et al* (1996) found that adaptation could be observed, both when individual grating patches were adapted to and when grating patches embedded in crowded displays were adapted to. Furthermore, the threshold elevations they observed for the crowded and non-crowded conditions were very similar. Clearly if lateral masking is the cause of crowding, it must be occurring at some stage subsequent to the extraction of line primitives.

Recently Parkes *et al* (2001) went on to show that the precision of global orientation estimates are independent of the number of oriented line elements within the display, despite observers being unable to report the orientations of these individual lines (i.e. despite a crowding)[1]. They hypothesize that crowding is requisite to the establishment of global representations so that ensemble statistics can be efficiently derived. If true this crowding is probably enforced by the coarseness of attentional resolution (He *et al*, 1996).

1.8 Summary and future direction

In this general introduction, literature from various areas of research has been reviewed. Each bears relevance to the experimental work of this thesis and/or its interpretation. The first research question that this thesis aims to address is the one posited in section 1.6: Are motion representations more readily accessible to the observer if they elicit impressions of 3-dimensionality? The preliminary work of Freeman *et al* (1996b) gave reason to suspect that this might be the case. However, in this study, extensive data was gathered from only a

---

[1] Dakin (2001) reached an opposing conclusion. He made internal noise estimates for the averaging mechanisms and deduced that increases in element density do lead to the degradation of local orientation estimates, implying the locus of crowding is at some point that precedes averaging.

single observer, and as will be described in chapter 3 this experiment was limited in its ability to assess accessibility to representations. Before proceeding with this line of enquiry chapter 2 will first describe some of the general methods and stimuli that were employed for the experiments of this thesis as well as some of their limitations.

Chapter Two: General equipment and stimuli

2.1 Chapter Introduction

This chapter documents equipment and observer details along with some aspects of stimulus generation that are common to all of the experiments that were carried out. In terms of methodologies, the experiments can be divided into three types: visual search, slant matching and coherent motion detection. The details pertaining to each of these will be introduced when required.

2.2 Equipment and displays

Stimuli were generated using a Cambridge Research Systems 2/3 card under the control of a PC computer (either a Pentium 120MHz, Pentium III 450MHz, or a Pentium III 500MHz) and displayed within an 25cmx25cm viewing area on either a Nokia Multigraph 445X, Sony 200PS, or Sony Trinitron 20seII monitor. Two experiments (experiment 3c from and experiment 4) required an extended framestore and were run on a separate set-up consisting of a Pentium 200MHz PC running a Cambridge Research Systems 2/3 card and extension card. Stimuli were displayed within an 18.5cmx18.5cm viewing area on an Eizo F553-M monitor. All monitors that were used operated at display rates of 120Hz and experiments were executed using programs written in Borland Pascal with Objects 7.0.

Observers viewed the displays in a darkened room from a constant viewing position. A fixed head position was ensured by use of a chin and headrest. All behavioural responses were made using a standard PC mouse.

2.3 Observers

Five observers were used in the course of the experiments, all with normal or corrected-to-normal vision (PAA, KAM, JBS, HRC, AEF). Only one of these, the author, participated in all experiments. Three out of the five observers had previous experience of psychophysical data gathering (PAA, KAM, JBS) and all, except the author, were unaware as to the exact purpose of the experiments.

2.4 Random-dot-kinematograms

Motions were displayed by use of random-dot-kinematograms (RDKs). These have the advantage of being able to provide motion signals in the absence of form cues.

Dots were positioned randomly within a designated stimulus area. Typically, hard aperture borders were used to curtail their visibilities although a more gradual curtailment is used in the summation experiment of chapter 9. All of the experiments in this thesis used circular patches of RDKs with a small region left blank at the patch centre. For both the visual search and slant matching experiments, all dots adhered to a global motion transform (transform matrices for the complex motions are described in appendix 1). In the summation experiments only a proportion of the dots followed trajectories consistent with a global motion transform; the remaining non-signal dots would follow random direction trajectories with the exception of one control condition in which the non-signal dots followed no specified trajectory, instead they were randomly re-allocated to other positions on each movie frame.

The perceived locations for individual dots were specified with sub-pixel accuracy using the algorithm of Georgeson *et al* (1996). The algorithm permits manipulation of perceived dot position by specifying each dot as a 2x2 set of pixels and adjusting the pixel luminance in an appropriate manner.

All displayed dots were bright with luminances set to approximately 110cd/m^2 and were displayed against a low luminance background of 2-3cd/m^2. The apparent motions that were induced from these high luminance dots did not induce any overt impressions of visual persistence or motion trails.

2.5 Dot trajectories

Dots always followed straight-line trajectories and maintained the same speeds throughout their lifetimes. Because of the geometry of the visual world, these dot trajectories are not true examples of the 2-D optic flow components that arise from real world planar stimuli. For example, a rotating surface of dots would generate curved dot trajectories and a looming surface of dots would contain dots that follow a positive radial trajectory (i.e. an expansion) and accelerate during their lifetimes. Dots following any other 2-D motion component - that can be described mathematically by differential invariants - would similarly exhibit velocity changes over its lifetime. However, the main benefit of the stimulus design used here is that it ensured all three complex motion conditions had the same average distributions of dot velocity, enabling meaningful comparisons to be made between them.

In the following subsections a number of limitations of typical random dot displays are discussed.

2.5.1 Straight-line trajectories can generate unwanted radial components

It was noted above that all dots followed straight-line trajectories even when describing global rotations. As a result the local trajectories will always contain a positive (expanding) component of radial motion and may therefore promote an undesired impression of an outwardly expanding spiral motion. To minimise the effects of such a component the randomly allocated dot positions were designated as mid-trajectory points, not starting points. In so doing, the first half of the trajectory will contain a negative radial component that is equal in magnitude to the positive radial component of the second half. Therefore over the duration of the entire trajectory the radial components will cancel. This method of dot allocation was applied for all complex motions.

2.5.2 Radial components add small amounts of noise to displays

The type of global motion will also subtly affect the amount of signal present in the display. When a global motion contains a positive radial component, a proportion of dots will be expected to travel beyond the outside aperture. Conversely, when a global motion contains a negative radial component, a proportion of dots will be expected to travel into the aperture from the periphery. The greater the magnitude of this radial component, the greater the proportion of dots will be expected to cross the aperture border. A rotational flow contains a smaller radial component than expansion or contraction and hence will lead to a greater proportion of dots crossing the aperture border. In the case of positive radial motion the outflow will lead to premature dot death while in the case of negative radial flow the inflow will lead to a late dot birth. In either condition it will add a small and possibly negligible amount of noise to the display that would decrease as the radial component is removed. This means that the ability to make direct comparisons between radial sensitivities and rotation sensitivities might be compromised.

2.5.3 Large dot trajectories generate dot density gradients

When RDKs contain global motions with radial components then subtle dot density gradients will emerge. The principle reason for the emergence of these dot density gradients is illustrated in fig.2.1. For simplicity, this figure considers dots allocated to a narrow annulus area of the aperture (boundaries denoted by dot-dashed circles). It is assumed that dot lifetimes are two frames long and the motion transform dictates a pure radial motion. It is clear from the diagram that the inner annulus region will on average contain the same number of dot instances as the outer region, yet the inner region covers a smaller area. This shows that dot density gradients will emerge for any complex motion containing radial components. The magnitude of these gradients will depend upon the magnitude of the radial component and the length of the dot trajectories. For reasonably short dot trajectories these dot density gradients are probably well below detection threshold. This view is supported from results of a control experiment carried out by Meese & Harris (2001b).

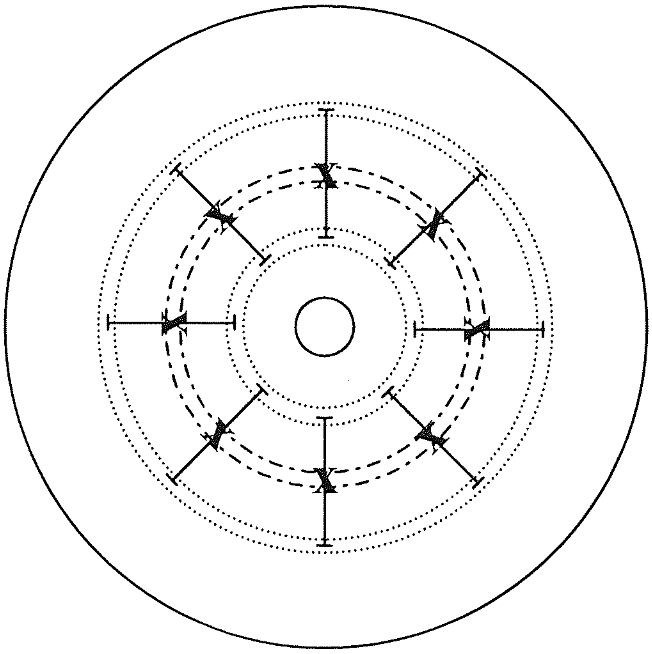


Fig.2.1. A diagram depicting how large dot trajectories can lead to dot density gradients. Solid lines delineate the hard borders of aperture. Dot-dashed lines denote an area in which all dots with trajectory midpoints falling within it are considered further. Dot midpoints are denoted with X's and for simplicity of this exposition they are made to be equi-spaced and have pure radial trajectories. Also illustrated are the positions of the two instances for each dot; the dotted lines delineate the areas within which these positions lie. Half of the dot instances will be positioned within the inner dotted annulus while the other half will be positioned in the outer dotted annulus; clearly the outer annulus covers a greater area than the inner, and therefore dot density will be greatest towards the centre.

2.6 Limited-lifetime dots

In all of the visual search experiments except experiment 1, limited-lifetime dots were used. With limited-lifetime dots a certain proportion of the dot trajectories terminate on each movie frame. These are replaced with new trajectories, ensuring that dot density remains constant throughout the display period. Limited-lifetime dot displays have several advantages over finite lifetime displays:

1) They ensure that unwanted radial components are, on average, balanced. In finite-lifetime displays all the signal dots would synchronously follow an identical temporal pattern of expansion-contraction or contraction-expansion but in limited-lifetime displays the pattern of expansion/contraction will, on average, remain the same throughout the display period.
2) They remove any dynamic dot density cues. In the finite-lifetime displays, complex motions will lead to systematic fluctuations in dot density. For example, at the end of a RDK movie displaying expansion, the dot density will tend to be lower than it was at the beginning (this is clear from the fact that, on average, more dots will have left the aperture than entered it during a display period).
3) They permit the presentation of continuous motion displays, avoiding the need to jump back abruptly to restart a movie. These jumps might contribute significantly to noise levels and be potentially confusing for the observer. Limited-lifetime movies avoid this because they can be constructed from dots with staggered cycles of 'birth' and 'death'.
4) A corollary of the above point is that short dot lifetimes may be used. This can be advantageous for minimising the presence of undesired motion components and static cues because shorter dot trajectories will better approximate the ideal trajectories of the differential invariants.
5) Short trajectories might also be used to discourage the tracking of single dot trajectories and so force observers to base their judgments upon a global motion analysis.

One of the disadvantages with this method of dot generation is that the increased 'birth' and 'death' rates of the dots add to the display noise. For example, signal dots with a lifetime of just two frames will possess half the signal of those of finite-lifetime displays. Choosing appropriate dot lifetime lengths depends upon the degree of saliency that is required. If sufficient impressions of global motion / 3D structure can be elicited with short dot lifetimes then they should be used. In many cases however, longer dot lifetimes are needed to avoid excessive reductions in the signal.

2.7 Speed gradients

Dot speeds within an RDK can either be fixed at a constant value or varied linearly with nominal start position. For 2-D motions the presence of a linear speed gradient enables closer approximation to their differential invariants. For the experiments involving visual search (experiment 1-3b), slant matching (experiment 3c) and the summation of orthogonal motion components (experiment 4), linear speed gradients were employed (with the exception of the translation condition the first visual search experiment). In the area summation experiments (experiments 5a-e), however these speed gradients were removed.

2.8 Coherence thresholds and their measurement

In experiments 4-5e, coherence thresholds are used as a performance measure. These are defined as the minimum proportion of signal dots that are needed to reliably detect a coherent motion. In these experiments, reliable detection is taken to be 50% above chance level (i.e. 75% correct). Previous studies have confirmed that thresholds are largely independent of dot density (Barlow & Tripathy, 1997).

A temporal two-interval forced-choice (2IFC) paradigm was typically used to gather the data. One, randomly allocated, interval would contain only noise dots (N interval), while the other would contain some proportion of signal dots (S+N interval). The observers' task was to indicate which interval contained signal, by using the left and right buttons of a PC mouse. Two control experiments of this thesis did not use the above paradigm. Instead, they used a single-interval forced-choice (1IFC) direction discrimination paradigm. In these cases, a single test interval was used, which always contained some proportion of signal dots. The observers' task was to indicate the polarity of the motion (for example in a radial motion condition, the left button denoted an expansion and the right button a contraction). Once a response had been made, immediate auditory feedback was given to indicate its correctness.

The proportion of signal dots in the S+N interval were varied in logarithmic steps using a three-down, one-up staircase (Wetherill & Levitt, 1965; Meese, 1995). Signal levels started at maximum to give observers a chance to familiarise themselves with the stimuli. For each staircase, data were gathered during either 8 (spatial summation experiments) or 12 (orthogonal component summation experiment) staircase reversals, and a single threshold estimate was based upon data collected from a pair of interleaved staircases (Cornsweet, 1962). Probit analysis used these data to estimate thresholds and corresponding standard errors (SEs) (Finney, 1971).

If the SEs from a single session (each session consisting of approximately 110 trials) were equal to or above 3dB then the corresponding threshold estimates were discarded. The majority of sessions yielded SEs below 3dB. In those that did exceed this level there seemed to be no systematic relationship with the magnitude of the threshold estimation, suggesting that these high errors were reflecting noisy data and not artefacts of the curve fitting to the data.

Chapter Three: Visual Search - Search paradigms and literature review

3.1 Chapter Introduction: A need to manipulate set-size

In the 4AFC detection and reaction time tasks used by Freeman *et al* (1996b) comparisons were made between performance measures of complex motions and those same complex motions vector summed with translation. The addition of a translation component to rotation and radial components tended to degrade performance while its addition to deformation tended to improve performance. It was posited that improvements in the later condition are attributable to an observers' ability to access mechanisms that process 3-D interpretations more rapidly than mechanisms that process 2-D interpretations. In the specific situation of *defl+trans* it was conjectured that odd-item detection had been based upon the 3-D descriptor: slant sign.

The inferior performances with the *defl* alone condition were attributed to the fact that slant sign in these flows were ambiguous and so could not have been used as a reliable cue. Instead observers would have had to resort to a detection based upon 2-D interpretations, which they proposed were accessed with poorer efficiency.

One possibility is that 3-D representations are accessed simultaneously while the 2-D representations are accessed serially. The left plot of fig.3.1 illustrates RTs that would be expected under this hypothesis. Set-size denotes the number of motion patches present in the display. For *defl* alone, performance is shown to be based upon a serial access while for *defl+trans*, performance is based upon simultaneous access. One limitation of Freeman *et al*'s study is that only displays containing four motion patches were tested. As shown in the right plot of fig.3.1, this means that there is an alternative account for their findings. In the alternative account *defl+trans* performance is based upon a serial access while *defl* performance is given simultaneous access. Also reflected in the plot are differences in the processing times of the individual stimuli: *defl+trans* latencies are shorter than those for *defl* latencies. Taken together these differences conspire to set performances at a set-size of four, to levels identical to those in the left plot.

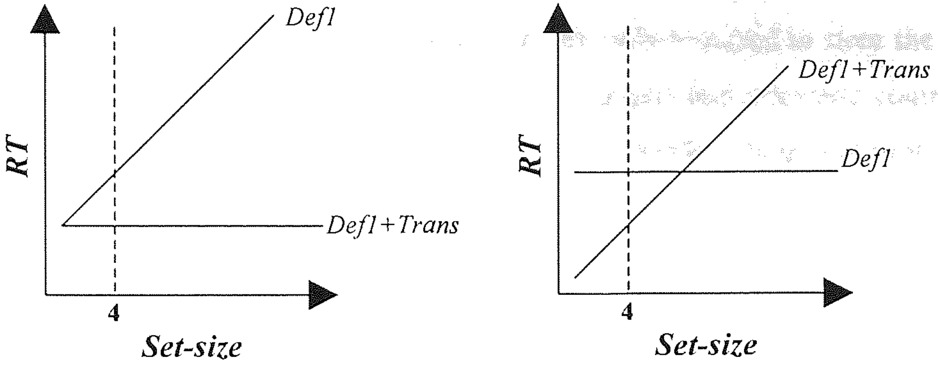


Fig.3.1. Two extreme scenarios that could account for the data obtained in the *Defl* and *Defl+Trans* conditions of Freeman *et al* (1996b) in two different ways. The dashed lines mark set-sizes of 4 (corresponding to the set-size used by Freeman *et al*) RTs at this set-size are the same in both plots.

Therefore, to obtain a clearer picture as to whether access is more rapid for one stimulus type than another, performances need to be measured over a range of set-sizes. The remainder of this chapter will concern itself with the studies and the paradigms that are typically used when ranges of set-size are measured. Also described are the possible interpretations that can be made from the magnitudes of the set-size effects that are obtained from such paradigms. Throughout this and the three subsequent visual search chapters there is extensive reference to the literature. The primary reason for such an extensive and broad coverage is that over the last two decades, since set-size effects began to be measured and used as a diagnostic for the underlying search strategies being used, there have been various concerns raised about the interpretation of these set-size effects. Within these chapters I hope to point out the potential caveats and possible criticisms that should be considered when interpreting set-size effects.

3.2 Paradigm and theory

By far the most common visual search paradigm to have been used is what has become known as standard singleton search. In this paradigm the observer is required to detect the presence or absence of a single target stimulus within a display of other distractor stimuli. On 50% of the trials a target will be present, on the remaining 50% no target will be present only distractor stimuli. The total number of stimuli in the display (the set-size) is varied from trial to trial by varying the number of distractors, but the number of possible responses is always kept the same: 'target is present' or 'target is absent'. Estes & Taylor (1964) were the first to introduce a detection paradigm of this sort. Its purpose was to assess the number of search items identified while minimising the potentially confounding effects of memory (e.g. Sperling, 1960) and response complexity (Hicks, 1954). The effects of memory are minimised because observers are only required to store the status of a single stimulus type (which will be what defines a target) and response complexity is minimised because that status can be one of only two possibilities (target-present or target-absent). This is in contrast to the early experiments of attention research that are often referred to as selective filtering experiments and are characterised by large response sets and comprehensive report by the observer.

In visual search tasks the observers are instructed to make a response in as short a time as possible while attempting to maintain a constant accuracy irrespective of set-size. The dependent measure has traditionally been reaction time (RT), with search performance being defined as the magnitude of the set-size effect. The magnitude of the set-size effect is taken simply to be the slope of the best-fitting line when mean RT is plotted against display set-size. This has typically described the data well.

Many stimulus types used in standard singleton search paradigm were found to give zero or close to zero set-size effects. In other words increasing the number of distractors had little to no effect on the time taken to detect the target. These searches tended to be found for targets that were distinguishable from their distractors by a simple feature/primitive such as colour or line-orientation (Treisman & Gelade, 1980; Nakayama & Silverman, 1986). For example searches for a patch of leftward motion amongst patches of rightward motion or a red line amongst green lines yield set-size effects close to zero.

3.2.1 Feature-integration theory (FIT)

Triesman & Gelade (1980) proposed that there was a limited number of simple features within the visual system each with their own arrays of specialised detectors, arranged in mosaics that tile the entire visual field. Often these 'features' were assumed to be synonymous with the stimulus properties of early cortical mechanisms such as those exhibited by the cells of V1. Many of the cells in early cortical areas and beyond form comprehensive mosaics that can operate in parallel with their retinotopic organisation retained in the anatomy of the cortex. Retinotopic maps could be constructed for various features like orientation, colour, motion and size: ones that reflect the dimensions of selectivity, that such cells possess. In Feature-integration theory (FIT) Treisman & Gelade

(1980) attributed the simultaneous, parallel nature of feature search to the presence of corresponding feature maps that were free to operate in parallel.

Other searches involving multiple-features yielded what seemed like a qualitatively different group of slope magnitudes. They produced steep set-size effects of 20ms/stimuli or greater. In these, targets would be defined by the presence of two or more 'simple features'. The distractors could take on the various stimulus features common to those of the target but a single distrator would be prevented from taking on the exact combination to that of the target. An example of a feature-conjunction search would be search for a target motion defined by its direction (upward) and colour (red) amongst distractor motions that were downward-red, upward-red or upward-green. FIT attributes steep slopes to a sequential search strategy – what has been termed Serial Self-Terminating Search (SSTS) – in which display stimuli are analysed one by one until a target is encountered. It claims that an SSTS strategy is resorted to when there is an absence of specialised detectors in the brain, capable of analysing the feature-conjunctions. In order to analyse them attention needs to be deployed but because attention is thought to be a limited-resource it can only be focused on a single stimulus at a time.

In summary, it is the neural substrate that is held as being the limiting factor that degrades search to inefficiency (those with steep positive set-size slopes), not attention. In this theory, attention is only ever deployed to a single stimulus item at a time irrespective of the stimulus type. Under FIT, the set-size effects associated with a stimulus dimension or conjunction of dimensions can be used to test for the presence or absence of specialised detectors selective for that stimuli. If they are present, efficient search will result (<5ms/item) if not, inefficient search will result (>20ms/items). Table.1.1 summarizes the properties thought to be associated with each of the two categories.

	Efficient (<5ms/item)	**Inefficient (>20ms/item)**
Phenomenology	Effortless (Pop out)	Laborious investigation
Instantaneous region of processing	Display-wide	Single stimulus item (Attentional spotlight)
Attentional demand	Preattentive	Subject to attentional bottleneck
Nature of processing	Parallel / Automatic	Serial / Controlled
Level of processing that search was based on	Early	Late
A target is defined by a	Feature	Conjunction of features
Stimuli processed by	Specialised detectors	Attention binding

Table. 1.1. The various differences that FIT would infer from the two search slope categories.

3.3 Features and Surfaces

We will now turn to some of the work that has been carried out in attempts to establish the status of particular stimuli and hence gain an insight into the nature of its processing.

3.3.1 Translational motion is featural

Direction selectivity is known to be a property of some complex cells in V1. As the features of FIT are often thought to correspond to early retinotopic maps, 1D motion direction should be a likely candidate for having a featural status. Set-size slopes of approximately zero have been found for searches in which the targets contain motion but the distractors do not (Driver & McLeod, 1992). While these do suggest some aspect of motion has featural status it is unclear from these as to whether that is motion speed or motion direction. Other studies have based their target-distractor differences solely on motion direction (i.e. a target might possess rightward motion while the distractors possess leftward motions) or motion speed and have found that both of these factors are capable of producing set-size slopes of zero (e.g. Driver, *et al*, 1992; Nakayama & Silverman, 1986).

Additional support for motion direction as a feature comes from subjective reports of 'pop-out' in figure-ground displays (Julesz & Hesse, 1970). Figure-ground displays are a variant on the single target displays of classical visual search. In them, a figure, defined typically by a large number of stimuli/elements is embedded in distractor stimuli of the same element density. As in standard search tasks, targets and distractors differ along some feature dimension or conjunction of feature dimensions. In the displays of Julesz & Hesse the stimuli differed only in motion direction. For example a rectangular figure could be constructed from leftward translating lines while the surrounding lines would be rightward translating. They found such rectangles to be effortlessly discernable (i.e. pops out from the surround) and concluded 1D motion direction therefore to be featural. The theory behind this is very similar to that of FIT in that the textural 'pop-out' is believed to arise because the stimulus elements were processed in parallel, by specialised detectors and without attentional requirements (Julesz, 1986). While in the Julesz & Hesse (1970) study, performance was judged only on the subjective reports of the observers, more objective measures can be gained from texture segmentation tasks if observers are required to make some kind of judgement concerning the textural figure. Typically, performance would be measured by varying stimulus duration (or SOAs) so as to obtain a measure of search efficiency and not set-size (e.g. He & Nakayama, 1994).

In summary both visual search and texture-segregation suggest that at least for translational motion directions processing is parallel and efficient. An exception does seem to occur however if the stimuli are isoluminant (Luschow & Nothdurft, 1993).

3.3.2 Complex motion representation: Featural or feature conjunctions?

In chapter 1, psychophysical and physiological evidence was reviewed suggesting that specialised global motion detectors exist, albeit higher up in the processing hierarchy than where local and global translational detectors are first observed (MST). It remained unclear from the discussion whether multiple complex motion stimuli are processed in parallel or rapid serial (Meese & Harris, 2001b; DeBruyn & Orban, 1993). It is therefore of interest whether, under the definition of FIT, complex motions have featural status along with homogeneous translations, or not. If they do, the implication would be that arrays of complex motion detectors exist and are capable of operating in parallel. If true, what level might we expect this featural representation to be at: local or global? Beverly & Regan (1979) reported specific adaptation effects of radial motions to be apparent for patches of motion subtending only up to 1.5°, suggesting analysis to be reasonably local. This contrasts with other psychophysical reports suggestive of much larger RFs for the putative radial and circular detectors (Burr *et al*, 1998). The physiology also argues in favour of large RFs at least with regards to the known complex motion selective cells of MSTd. Many of these cells however, also appear to be position invariant; not a property typically expected from detectors constituting retinotopic feature maps. It is therefore unclear at what level a feature representation might occur. Fig.3.1 describes one scenario, in which complex motion feature maps might be derived. Using expansion as an example, it proposes feature detectors to be constructed from a two-stage hierarchy. A model of this sort finds support in the detection experiments of Morrone *et al* (1995). However the model does not intend to make any assertions about how the feature detectors might be activated and come to affect search performance. The feature detectors may, for example, perform linear summation of local motion signals irrespective of the distribution of their locations (as in Morrone *et al*'s model). Alternatively their activity might be dependent upon the distribution of their locations (e.g. diametrically opposed signals may affect detector activity more than spatially adjacent signals).

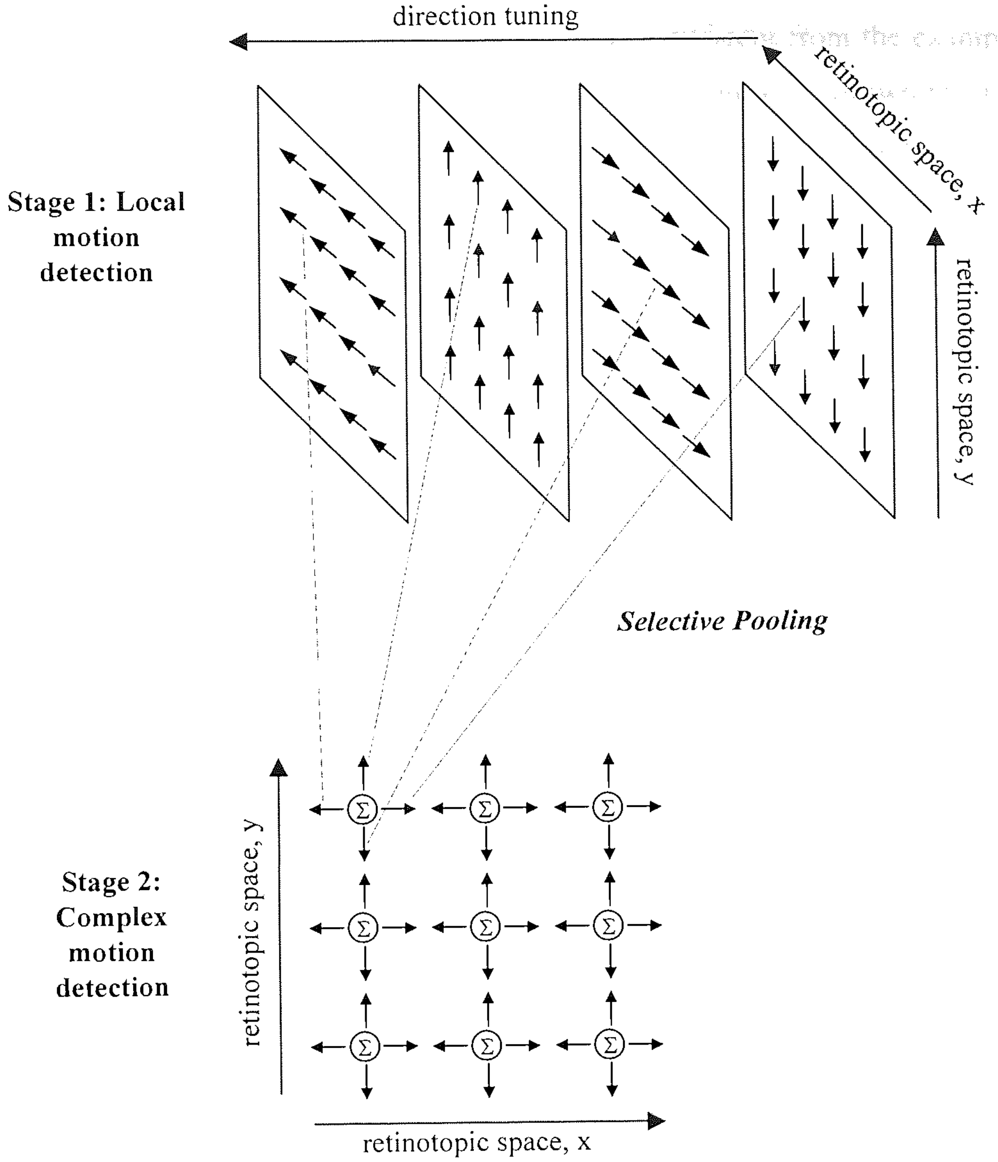


Fig.3.2. Feature detectors for expansion, constructed from a two-stage hierarchy. Illustrated is a simple two-stage feedforward model for complex motion processing. In this model, arrays of complex motion detectors tile the visual field, each constructed from specific arrangements of local motion detectors. This selective pooling is illustrated here only for expansion, although a number of these detector arrays may exist for various complex motions (see Chapter 1 for a discussion on the number of complex motions that may be explicitly represented in the brain and their tuning characteristics). For clarity the construction of only one of these expansion detectors is shown (one with an RF that is centred on the top-left of the visual field). From the perspective of FIT the complex motions that are explicitly represented by such detector arrays would be considered featural.

Alternatively if complex motions are not explicitly represented in the brain by arrays of specialised detectors that tile the visual field then processing must, in terms of classical FIT, be considered as attentionally demanding and therefore as some type of feature conjunction. The nature of this conjunction is clearly different from the example given in section.3.2.1. In that example targets were defined by the presence of two features (motion direction and colour). All distractors shared one of the two features with the target but with the constraint that not all distractors could share the same feature. Searches such as these, in which targets are defined by some unique conjunction of distractor features will be referred to as *across-item conjunctions* in general keeping with Duncan & Humphreys' (1989) terminology of, *across-object conjunctions*.

The visual search and texture segmentation experiments that have been carried out with complex motion stimuli have all defined their targets in the same way that Freeman *at al* (1996b) did: by allocating the target an opposite sign of motion to that of the distractors. For example in an odd-item-out search display containing distractors that contract, there would be a target of expansion. Both distractors and targets would be in possession of the same local component motions, just in a different spatial arrangement. As was first noted by Braddick & Holliday (1991) these targets and distractors can be considered as different conjunctions of the same sets of local motions. Searches such as these, in which targets are defined by their particular configuration within individual stimuli, will be referred to as *within-item-conjunctions*, again in loose keeping with the oft-used term *within-object conjunctions*. There has been confirmation that this type of conjunction reliably yields inefficient search both in the fronto-parrallel plane (Logan, 1994; Palmer, 1994; Davis *et al*, 2003; there are a few possible exceptions, for example, He & Nakayama, 1994) and depth (Moore *et al*, 2001).

The first study to investigate the status of complex motions was Julesz & Hesse (1970). Again they used figure-ground displays, restricting their investigation to rotation stimuli. While segmentation was found for figures defined by faster rotating lines they found no such segmentation when figures were defined by lines rotating in an opposite direction to those of the surround.

A visual search study by Braddick & Holliday (1991) investigated two other complex motions, radial flow and deformation (def2). They constructed stimuli from simple line drawings that would change shape in an appropriate manner during a movie sequence (i.e. they expanded, contracted or deformed). Movie sequences were cycled until responses were made. Targets and distractors were defined in the conventional way – distractors would take on the motion with the opposite sign to that of the target. To make sure stimulus size could not act as a reliable cue, the starting points of the movie sequences were randomly selected. The set-size slopes they obtained from their main experiment far exceeded the 20ms criterion for both divergence (Hit slope of 96ms/patch; C-R slope of 210ms/patch) and deformation (Hit slope of 276ms/patch; C-R slope of 542ms/patch) and the ratios between 'hit' slope magnitudes and 'correct-rejection' magnitudes were found to be around 2:1. Both these points suggested to them that these motions were processed as within-item-conjunctions, not features (slope ratios are shown in table 4.1). The 2:1 ratio is suggestive of a conjunction search because the SSTS strategy predicts that for target-present displays half the number of stimuli will need to be inspected before the target is found.

Takeuchi (1997) reinvestigated search performance for radial flows. His stimuli differed from the above in that the stimulus patches were of a fixed size and contained continuously drifting luminance gratings (a total of 4 gratings, placed in the upper, lower, left and right quadrants). How the trials were executed also differed: sessions were blocked with respect to target type and so in one session the target might always have been expansion (amongst contracting distractors) and on another, always contraction. Search for an expansion amongst contractions was found to be highly efficient (Hit slope of 4.3ms/patch; C-R slope of 5.1ms/patch) and therefore it was concluded that the expansions were processed in parallel and hence had featural status. A control experiment using 1-D expanding stimuli (i.e. positive horizontal compression) confirmed search for this condition was much less efficient (Hit slope of 23.2ms/patch; C-R slope of 76.5ms/patch). Interestingly, search for contractions did not produce the same results as those for expansions (Hit slope of 32.4ms/patch; C-R slope of 58.0ms/patch), although they were clearly more efficient than the slopes found by Braddick & Holliday (1991). Takeuchi speculates that the asymmetry might be the product of ecological factors. If approaching objects have held greater evolutionary significance to humans in the past then approaching objects might in general be expected to be processed faster than receding ones. Indeed, from Takeuchi's data, RTs for the detection of single expansions are lower than those for the detection of single contractions. This corroborates the findings of a previous study by Ball & Sekuler (1980).

A more recent visual search study by Thornton & Gilden (2001) addressed the issue of complex motion processing using a multiple-target search paradigm that they claim is able to distinguish between serial processing and some special cases of parallel processing.

Using this they find evidence that parallel processing can be confirmed both for expansion and contraction flows but not rotational flows. The reasons why parallel processing might yield positive search slopes will be discussed in chapter 7 along with the potential caveats of using multiple-target search paradigms. Included in their extensive study were single-target conditions (slope magnitudes are displayed in table.4.1). In these, all four motions tested - rightward translation, clockwise rotation, expansion and contraction - yielded positive slopes. The finding of a 20ms/patch slope for translation is clearly contrary to previous reports in which slopes were near zero. Possibly the most notable differences between this and previous studies are the ranges of set-sizes used and the method of data collection. Thornton & Gilden used the unusually low set-sizes of 1,2 and 4. It is debatable whether higher set-size ranges tend to yield more shallow slopes (Pashler, 1987; Palmer, 1994). They also intermixed trials containing single targets with trials containing multiple targets, possibly adding a degree of task uncertainty. Another discrepancy emerged between the expansion and contraction efficiencies of this experiment and those of Takeuchi's (1997). Thornton & Gilden found no sign of a radial anisotropy (expansion Hit slope of 16ms/patch; contraction Hit slope of 13ms/patch) contrary to the findings of Takeuchi (1997).

3.3.3 Other complex stimuli with featural status

Nakayama & Silverman (1986) displayed stimuli on a common stereoscopic fronto-parallel depth plane and found they could be efficiently searched even when target-like stimuli are displayed on other fronto-parallel depth planes. This was found to be the case for targets defined either by their motion direction (upward/downward) or their colour (red/green). At the time, Nakayama & Silverman took this to imply the existence of multiple feature maps selective for both motion direction and colour at various disparities. They drew support for this view by citing the physiological evidence for the existence of cells in MT that are both selective for motion direction and disparity (see section 1.4). It therefore seemed plausible that feature maps for stereo-motion were residing in MT and those responsible for stereo-colour were pending discovery (Maunsell & van Essen, 1983). There are other psychophysical data that could be drawn upon to support the view that mechanisms exist that are selective for both motion direction and disparity. Anstis & Harris (1974) found after-effects dependent upon the motion direction and disparity conjunction to which the subjects were adapted.

Other complex stimuli have been found to produce efficient search. Wolfe *et al* (1989) for example found near zero search slopes for some colour-form conjunctions and Theeuwes & Kooi (1994) have discovered that conjunctions of shape and contrast polarity also yield near zero slopes. Another similar finding has been found for the conjunction of motion and form (McLeod *et al*, 1988).

The above findings pose two challenges to the exposition of the previous sub-section. The first challenge concerns itself with the assumption that the neural substrate for feature maps occurs only within the striate cortex, whereas the earliest anatomical location for the processing of the complex stimuli of Nakayama & Silverman (their motion-disparity conjunction, 1986) and Takeuchi (expansion, 1997) appear to be extrastriate (MT and MST respectively). The second challenge derives mainly from studies such as Wolfe *et al*'s (1989) in which the conjunctions of the primitives producing efficient search appear arbitrary. It is difficult therefore to envisage why the visual system would have specialised detectors dedicated to such stimuli. Each issue will be dealt with in turn starting with the anatomical location of these hypothetical feature maps.

3.3.4 Feature maps are not synonymous with the stimulus dimensions of V1 cells.

In addition to the fact that parallel processing is found for stimuli having properties that far exceed the processing capabilities of V1 cells, another line of evidence also argues against the idea that feature maps are located exclusively within V1.

This evidence comes from the findings that second-order stimuli can be searched, in some cases just as efficiently as first-order stimuli. For example orientations defined by motion (Cavanagh, Arguin, & Treisman, 1990) or even illusory contours (Gurnsey, Humphrey, & Kapitan, 1992) can yield parallel searches. This occurs despite the fact that second-order processing has in the past not been thought to occur at stages as early as V1.

It should be conceded however that recent imaging and physiological studies have suggested that a certain amount of mid to high-level processing like second-order processing may actually be occurring in early areas, including V1. Imaging studies on humans have shown that differences in the activity of the early visual areas can correlate with differences in perception and it has been demonstrated that these are unlikely to be the result of low-level feature differences (e.g. Humphrey *et al*, 1997). Activity of some V1 cells in Macaque seem to be driven by illusory contours as well as non-illusory luminance defined contours. Sugita (1999) found this to be true for illusory contours defined by amodal completion while Grosof *et al* (1993) also found it to be true for modally completed contours. Zipser *et al* (1996) have also identified V1 cells whose activity is modulated by texture differences occurring outside their RF's; moreover these modulations could be induced by different cues; for example the texture differences could be defined by colour, luminance, disparity or motion. Bakin (2000) similarly showed that the activity of Macaque V2 cells (another area early in the processing stream) can be correlated with modally defined contours irrespective of whether they are induced by luminance or disparity cues, modally or amodally. Findings such as these have started to challenge the long held belief that early areas (like V1 and V2) are concerned only with retinal image extraction and not second-order processing and inference.

It is unclear whether the cell properties described above (i.e. responding to stimuli that require more than simple low-level feature extraction) are independent of any top-down attentional influence. Modulations responding to the 'higher-level' stimuli did tend to appear a few hundred milliseconds after stimulus onset, a delay sufficient enough to make feedback from higher areas a possibility. Indeed the delayed modulations of Zipser *et al* (1996) seem to disappear when the Macaques are anaesthetised, despite the fact that they continue to respond to stimuli driving their classical RF's (Lamme *et al*, 1999). It remains to be seen, exactly to what extent areas V1 and V2 play in higher-order processing and whether there is a strong correlation between the types of stimulus processing going on in these areas and the types of stimuli that can be searched efficiently.

Although a convincing argument about higher-level processing complexity in the early visual areas can be invoked, a convergence in the various lines of evidence suggests that feature maps are not synonymous with the stimulus dimensions of V1.

3.3.5 Parallel search for ostensibly arbitrary feature conjunctions are the exception, not the norm

It is easy to imagine that some conjunctions of local components have greater ecological significance attached to them[1] and therefore a visual system with detectors specifically

---

[1] For example expansive flows are indicative of surfaces or objects that may or may not be about to collide with the observer; it is conceivable that contraction flows in general would represent situations of lower behavioural significance.

dedicated to them, would, in computational terms be advantageous. As long as the number of these 'special' conjunction detectors remains small then it should pose no problem for FIT. Clearly if too many conjunctions appear to have featural status and the conjunctions seem arbitrary, the combinatorial implications for neural hardware would make models like FIT untenable. Some models have attempted to account for the good performance of certain conjunctions without resorting to the explanation of specialised conjunction detectors. For example in Revised Feature Integration Theory it is suggested that sometimes when the features of the target are highly salient they actively inhibit the corresponding features of neighbouring distractors (Triesman & Sato, 1990). This feature differencing could conceivably arise in a bottom-up fashion through lateral, inhibitory connections. Wolfe's Guided Search model (1989) has a similar theme but instead of feature inhibition it suggests feature summation; the decision mechanism makes use of the feature summation maps that result (typically referred to as a saliency/activation map). A review of the literature however suggests that the majority of complex stimuli that yield parallel-like search are often distinguishable because they do hold ecological significance usually in terms of surface-like attributes. This implies that the feature maps of FIT, originally thought to be exclusive to early, primitive attributes are also found for some surface level attributes.

3.3.6 Search performances based on surface-level representations

Convex shapes that are defined by shading are efficiently searched (Kleffner & Ramachandran, 1992), as are inferred Kanizsa figures (Davis & Driver, 1994). A number of other studies have confirmed surface slant to be searched efficiently. Enns & Rensink (1991) showed that line drawings consistent with slant impressions could be rapidly searched on the basis of slant sign. As well as pictorial depth cues disparity depth cues have also been tested. Holliday & Braddick (1991) produced stimulus elements with strong impressions of slant by presenting outline parallelograms dichoptically. Each parallelogram would have its own monocular representation but there would be a slight difference in the degree of shear of the figures between the two eyes. This generated stimuli with vertical disparity gradients. Accordingly they were perceived as flat slanted surfaces with a 90° tilt. Using a standard odd-item-out search task they tested whether an element having a disparity gradient of an opposite sign to the others could be searched efficiently. In the same way that complex motion searches were caste in terms of a series of attentionally demanding *within-item-conjunctions* (of local motions), disparity gradients might also necessitate a *within-item-conjunction* (of local disparities) search. Moore *et al*

(2001) have confirmed that the apprehension of relative disparities for two arbitrary but featurally distinct stimuli (green plus symbols and purple squares) is highly inefficient. In their main study the task was to search for a stimulus pair in which the green plus symbol had the greater disparity and therefore appeared in front of the purple square. If no specialised mechanisms exist for processing disparity gradients or general slant representations then a similar result would have been expected for Holliday & Braddick's disparity gradient stimuli. They, in contrast, observed efficient parallel search (Hit slope of 4.4ms/element; C-R slope of 7.7ms/element) indicating that multiple mechanisms capable of distinguishing between signs of disparity gradient do exist.

3.3.7 Early level representations cannot be directly accessed

In addition to the arguments above, which suggest the feature maps employed by visual search might also reside in higher areas, several other experiments suggest that access to early feature-level representations is actually restricted under certain stimulus conditions (Wolfe & Franzel, 1988; Susuk & Cavanagh (1995); Rensink & Enns, 1998; He & Nakayama, 1992; 1994; Davis & Driver 1994; 1998). In the key experiment of Wolfe & Franzel's (1988) study they investigated human ability to perform a search, based only on utrocular orientation differences. Search for lines differing in orientation by 90° would be rapidly detected under non-rivalrous binocular conditions (e.g. Foster & Ward, 1991). Wolfe & Franzel went on to question whether this rapid search would persist when binocular rivalry is used to produce stimulus items that lead to similar local image alternations irrespective of whether they are targets or distractors (see fig.3.3. for an illustration of the kind of dichoptic displays they used). They found that subjects could not carry out the task to any reasonable degree of accuracy. As both monocular views contained sufficient information to distinguish the target from the distractors, their conclusion was that the decision stage had access only to mechanisms subsequent to the locus of rivalry. The fact that certain cyclopean cues such as disparity can support efficient parallel search demonstrates that the combination of monocular information need not necessarily be detrimental to search efficiency; it can often be conducive to it. Another experiment in the same paper found gloss (induced by binocular luster) to be processed in parallel. This may not be surprising from the perspective of the aforementioned hypothesis, that search performance is closely linked to surface descriptors and not the activity of low-level analysers, because gloss may also be an informative surface property.

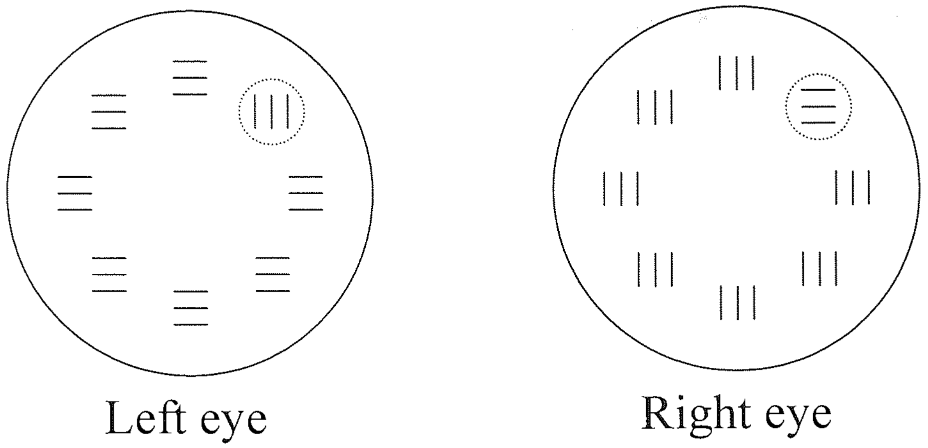


Fig.3.3 Illustration of the dichoptic displays used in the utrocular rivalry experiment of Wolfe & Franzel (1988). In this example a target is present and indicated by the dashed circle (not present in the actual displays).

Other search and texture segregation studies have demonstrated other instances in which primitive monocular features are rendered inaccessible to the decision stage; only higher-level cyclopean representations seem to be available. In these experiments disparity cues were used to induce amodal completion; amodal completion being the inference that some form/surface continues behind another occluding form or surface. Disparity can be a powerful inducer of amodal completion when consistent with interposition cues and is capable of altering the observer's perceptual experience as well as their behavioural responses (Nakayama *et al*, 1989). Using white 'L' and mirrored 'L' shaped stimuli that always abutted black squares, He & Nakayama (1992) carried out odd-item-out search tasks for two experimental conditions. In one condition all the 'L' shape stimuli were given the same crossed disparities and so perceived as 'in front' of fixation. In the other they were given uncrossed disparity and so perceived as behind fixation. The black squares always maintained zero disparity irrespective of the condition or whether they were accompanying target stimuli or not. In the 'in front' condition the stimulus arrangements give rise to distinctive impressions either of an 'L' shape or mirrored 'L' shape depending on the monocular images presented. The upper diagram of fig.3.4 illustrates this perceptual impression for an 'L' shape stimulus item. In the 'in back' condition amodal completion occurred and the 'L' shapes, whether mirrored or not, tended to appear as two corner edges of a white square, the rest being occluded by the abutting black square. The lower diagram of fig.3.4 attempts to illustrate this impression of amodal completion, again for an 'L' shape stimulus element. He & Nakayama reasoned that if search performance was based only on representations that were subsequent in the processing stream to amodal completion then search performance in the 'in back' condition would be poorer than in the 'in front' condition because all the stimuli would be represented as a homogenous set of white squares, even if one of the stimuli was actually a target and monocularly different from the rest. This was indeed what they found: steep search slopes for the 'in back' condition and near zero slopes for the 'in front' condition.

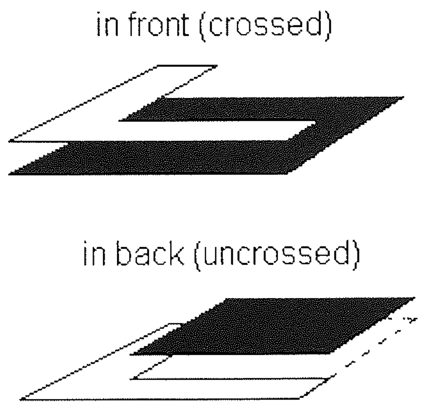


Fig.3.4. Schematics illustrating the impressions of depth that were induced by the two disparity manipulations for the 'L' shapes used by He & Nakayama (1992; 1994). The dotted lines for the 'in back' condition illustrate the perceptual completion that took place.

The latter finding might seem surprising given that there is no physiological evidence for specialised detectors selective for 'L' and mirrored 'L' shapes. The authors acknowledge this but nevertheless attribute them featural status, instead of instances of *within-item-conjunctions* (of two orthogonal lines). Previously, Olson & Attneave (1970, cited in Bruce *et al,* 2003, pp.128-129) have confirmed that displays containing 'L' and mirrored 'L' shapes do usually require effortful analysis. Using a texture segregation task they found that a quadrant defined by 'L' shapes was not easily discernable, as it should have been if the shapes had featural status. This discrepancy between the two studies may have something to do with the fact that in He & Nakayama's (1992) 'in front' displays the 'L' shapes had crossed disparities. This alone could not have been used as a search cue as both target and distractors would have had crossed disparities. But it may have been that the presence of relative depth made both the target and distractors more compelling (salient) and in turn enabled them to be searched more efficiently. As Olson & Attneave's 'L' shapes had no disparity they may have appeared less compelling. It is possible therefore that the highly efficient search of He & Nakayama's (1992) 'in front' condition is attributable to the enhanced saliencies that the crossed disparities afforded to the stimuli.

A follow up study went on to confirm the existence of an 'in back' inferiority effect with a texture segregation task (He & Nakayama, 1994). Slightly different monocular stimuli were used in this case; vertical bars and 'L' shapes, but in the 'in back' condition both would appear amodally completed to white squares as in the previous stimulus set. These two experiments confirm the findings of Wolfe & Franzel (1988) with a different stimulus:

that decision processes employed in visual search tasks are prevented access to low-level representations if pre-empted by higher-level representations (Rensink & Enns, 1995).

Davis & Driver (1998) have further tested the claim that amodal completions are carried out in an obligatory parallel manner using an even more elaborate stimulus. An example of one of their target-present displays is depicted in fig.3.5. Instead of using luminance-defined occluders they used white subjective Kanizsa squares induced with a set of black pacman-shaped inducers arranged in an appropriate fashion. The task for the observer was always the same: to detect the presence or absence of a large notched brown disk amongst unnotched brown disks. A notched target disk would be judiciously placed at one corner of the subjective square so that if it were given uncrossed disparity it would be entirely consistent with the squares presence. In the example display of fig.3.5 the notched target disk is located within the top-right stimulus cluster. Associated with each distractor configuration was a complete brown disk which would either replace one of the black pacman inducers or be placed a small distance away from the Kanizsa figure altogether.

Again two disparity conditions were tested, one in which the brown circles (whether notched or whole) were given crossed disparity and the other in which they were given uncrossed disparity. If subjective figures could be induced (by modal completion) in parallel and go on to act as occluders in the same way the black squares did in He & Nakayama's studies (1992; 1994), then the same 'in back' inferiority effect would be expected. Indeed this was what they found, with parallel-like search for the 'in front' condition (Hit slope of 2.8ms/element; C-R slope of 9.3ms/element) and an attentionally demanding serial search in the 'in back' condition (Hit slope of 57.5ms/element; C-R slope of 159.1ms/element).

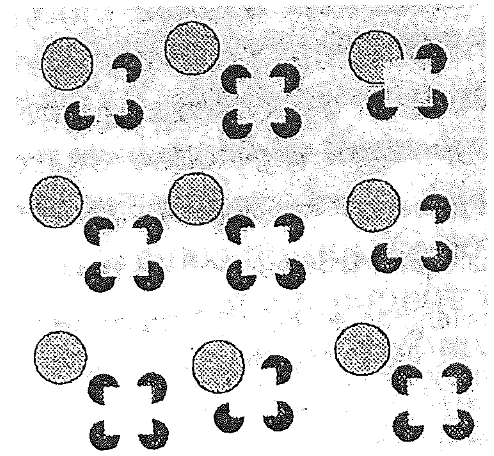


Fig.3.5. A depiction of a typical search display used by Davis & Driver (1998). This example has a set size of 9 and a target is present (the target being the larger notched circle of the top right stimulus configuration). Note that this illustration makes no attempt to represent a search of particular disparity condition. Taken from Davis & Driver (1998).

Rensink & Enns (1998) also found a similar inferiority search effect when interposition cues alone were used to induce amodal completions. In summary this selection of studies illustrates the importance of higher-level representations to the performances of humans in visual search and texture segregation tasks. Moreover it seems that the use of these representations are obligatory even if they do result in performance decrements. The latter studies that described parallelism for the processing of amodal/modal completions support the conclusions of the previous studies, in which higher-level surface descriptors were found to be processed in parallel (e.g. slant sign, Holliday & Braddick, 1991). Certainly in terms of the Marrian view of visual perceptual processing amodal completion would be considered as being part of the $2^{1/2}$D sketch (Marr, 1982). Whether such completions are processed earlier in the visual stream than surface slant remains unclear. But as some of the literature covered in chapter 3 will suggest, explicit surface descriptors (e.g. tilt, slant, curvature) are probably computed in the more distal extrastriate areas.

This concept of compulsory access to late perceptual representations is reminiscent of the crowding phenomenon and the studies of He & MacLeod (2001) and MacLeod & He (1992) that were discussed in chapter one. Visual crowding seems to provide a good analogy to the findings of these search tasks. The study by Parkes *et al* (2001) demonstrated that while it may not be possible for subjects to identify individual line orientations when they are embedded in a group of lines, the orientation of each of the lines are processed and can even have an effect upon certain behavioural performances based on higher-level representations. Both the processing of these individual lines and the 'L' shapes and bars of He & Nakayama (1992; 1994) can be considered as early feature extraction. Under the right conditions perception has full access to this early feature information; in the case of crowding this occurs only when single lines are presented and in the case of He & Nakayama's 'L' shapes, it occurs only when they appear with non-crossed disparities. However in other stimulus conditions direct access to this featural information is prevented in favour of higher-level representations; in crowding this higher-level representation probably emerges from a global compulsory orientation averaging and in the He & Nakayama experiments because of compulsory amodal completion. The proposal made here therefore is that while early feature-level detectors may undoubtedly limit performance it is not these that are directly monitored by the decision mechanisms rather it is some later-level detectors sufficiently sophisticated, to be capable of inferring surface properties. One implication of this for visual search is that steep positive set-size slopes should not be used as definitive evidence against the parallel processing of stimuli. It is possible that only under certain circumstances, for instance if the stimuli contributed to the generation of surface-like impressions, would such parallel processing be revealed by an absence of set-size effect. In relation to the work of Freeman *et al* (1996b) it is possible that the def1 stimulus they used was processed by specialised mechanisms in parallel both in the alone condition and when summed with translation. Only in the later condition however, when the extracted deformation components contributed to unambiguous surface interpretations, could efficient target detection be achieved.

3.4 Review summary

The study by Freeman *et al* (1996b) (section 1.6) provided preliminary data consistent with the hypothesis that perceptual access to multiple representations of motion is improved if these motions elicit impressions of 3-dimensionality. One limitation of this study was that performance measures were only obtained for displays containing four patches of motion. In section 3.1 it has been pointed out that data obtained for a single set-size cannot provide an unambiguous measure of search performance. A less ambiguous measure can is given by the magnitude of RT/set-size slopes, requiring odd-item-out RTs to be obtained for a number of set-sizes.

The general theory behind the use of RT/set-size slopes as a measure of search performance was also discussed. One model of visual search (FIT) that has been central to this review has proposed that efficient search occurs when arrays of special-purpose detectors are available to process the individual display stimuli. Contrary to FIT, in this chapter it is speculated that the existence of these multiple special-purpose detectors may not necessarily lead to efficient search, rather perception might only have unrestricted access to detector outputs that are of direct relevance to the observer, such as detectors responding to higher-level surface descriptions. This would be consistent with the '3-D access' hypothesis proposed by Freeman *et al* (1996b). A number of examples were given to support this view, in which efficient search was found for surface representations derived from stereoscopic and pictorial cues.

The search experiments described in this review, therefore, give support to the'3-D access' hypothesis, but it has also been shown that in order to obtain reliable estimates of search performance, set-size needs to be manipulated. Experiment 1 of this thesis addresses this issue so that a more rigorous test of the '3-D access' hypothesis can be carried out with respect to motion cues.

Chapter Four: Experiment 1 - Complex motion components

4.1 Chapter Introduction

The following chapter describes a singleton search experiment that was carried out on the same complex motion flows as were used by Freeman *et al* (1996b): rotational (*Rot*), radial *Rad*) and deformational (*Def1*), both with and without translation (*Trans*). The aim of this experiment was to obtain direct measures of search performance for these various motion flow types. There are several questions of interest. The first is concerned with the nature of complex motion searches. Only one study has previously investigated deformation flows and this claimed search to be serial (Braddick & Holliday, 1991). However, the study used deformations of the type, *Def2*, not the *Def1* components used in the Freeman *et al* study (see fig.1.6 for stimulus definitions). Moreover, the stimuli used by Braddick & Holliday were unusual, in that they were constructed not from formless textures or dots but from figure outlines. Two other studies have looked at rotations and are in agreement with each other that they are not processed in parallel (Julesz & Hesse, 1970; Thornton & Gilden, 2001). The nature of radial search remains unclear: one study has reported it to be serial (Braddick & Holliday, 1991) while another parallel (Takeuchi, 1997) and yet another parallel but subject to capacity limits (Thornton & Gilden, 2001). This last proposal will be considered in more detail latter.

One purpose of experiment 1 is to reassess search performances for *Rot*, *Rad* and *Def1* flows by using RDKs. This will ensure that performances are based solely upon motion cues. Previous singleton searches have typically found homogenous translation to yield near zero search slopes (Driver *et al*, 1992; Nakayama & Silverman, 1986; Braddick & Holliday, 1991; for one exception see Thornton & Gilden, 2001: they found Hit slopes of 20ms/item.). A *trans* condition will therefore be used as a control, to ensure that for a motion previously confirmed to be searched in parallel, the chosen method of stimulus generation is also capable of producing pop-out in ideal circumstances.

The second question that this experiment aimed to address was whether search performance for deformations could increase if summed with appropriate translation (*trans*) components. The preliminary results from Freeman *et al* (1996b) suggested that this might well be the case. They had however, made the assumption that there was parity in the RTs for individual patches across motion type. Measuring RTs at various set-sizes will enable measures of search performance to be obtained that could not have been confounded with general sensitivity. If it were found that performance improves with the addition of *Trans*, then this would add weight to the working hypothesis of 3-D access, described in section 1.4 to account for the data of Freeman *et al* (1996b) (See also section 3.1 for the motivation behind using the singleton search paradigm to further test this hypothesis). To summarise, the '3-D access hypothesis' predicts that *Defl+Trans* and *Rad* stimuli will be searched more efficiently than *Defl* alone because the former two can base their searches upon 3-D representations (slant sign for *Defl+Trans* and motion-in-depth for *Rad*). Search for *Defl* alone has to resort to 2-D representations. Access to 3-D descriptions is presumed to be more rapid than 2-D descriptions.

4.1.1 Small set-sizes may be an exception

In the study of Freeman *et al,* a single small set-size was used (four RDK patches). There is however the possibility that, for small numbers of any stimulus type, processing might be able to proceed in parallel. It could occur by use of general-purpose mechanisms that, with dedicated attentional resources, can be configured to independently process any stimulus of reasonable complexity. At higher set-sizes, when either the attentional resources or the neural machinery supporting this processing becomes limited, search would revert to a serial processing (see Pashler, 1987).

Many of the early experiments that claimed evidence for parallel processing, did so for letter stimuli, an undoubtedly complex stimulus in terms of classical FIT (see for example Shiffrin & Gardner, 1972; Shiffrin & Schneider, 1977). In many of these studies, observers were required to detect target letters in the presence of multiple distractor letters. These detection performances would be compared to ones derived from another condition, in which letters were presented sequentially. No performance differences were observed between the two conditions for set-sizes of four or less; at higher set-sizes, performance in the sequential condition surpassed that of the simultaneous condition. The researchers therefore concluded that parallel processing occurs for up to four letters.

It seems unreasonable to posit the existence of retinotopic maps for letter stimuli, and even if they did exist in some form they would almost certainly be occurring at a high level of processing and not at the low-mid levels at which feature maps are thought to reside. More recently, Klein & Lane (1986) have found reason to believe that the apparent parallelism in many studies can be attributed to feature detection rather than letter detection. For example a target letter 'N' might not only be unique in terms of it status as a letter but also by the fact that it contains an oblique line. This criticism notwithstanding, the general conclusion that up to four letters can be processed in parallel, remains well founded.

To minimise the likelihood that the performances in this experiment will be based upon general-purpose mechanisms, set-sizes larger than four will be included.

4.1.2 Learning affects perceived slant

Practice is known to affect the perceptibility of slant when it is induced by random-dot motion displays. Anecdotal accounts report that the impressions of slant become more compelling as the observer becomes more familiar with the stimulus type. Domini & Caudek (1999) have reported empirical evidence confirming differences between practised and unpractised observers, albeit with slants induced from object rotations rather than object-observer relative translations. It was found that the unpractised observers tended to underestimate slant to a greater extent than the practised observers do.

In the visual search literature, practise effects have been frequently reported. Practice can not only lead to general RT decreases but also to concomitant increases in search efficiency (i.e. decreasing set-sizes effects) (Heathcote & Mewhort, 1993; Steinman, 1987). In light of Domini & Caudek's findings, this fact raises concerns that if search were based upon slant impressions for *defl+trans* stimuli then the resulting search slopes might decrease with practice. To try to circumvent any such confound in this, and subsequent experiments involving possible slant impression, all observers were given extensive training with the stimuli before data gathering. This, it was hoped, would improve the perceptibility of slant to some stable level.

Most models of visual search, including FIT, suppose that stimuli having neural machinery dedicated to their processing would make search more efficient. Set-size effects might therefore be able to decrease (possibly to zero) if perceptual learning permitted the recruitment or reconfiguration of specialised mechanisms. While there is evidence that learning can lead to large improvements in search efficiency, there are no reports of set-size effects shallowing to zero and searches becoming 'feature-like' (Heathcote & Mewhort, 1993; Steinman, 1987; Wang *et al,* 1994). Shiffrin & Schneider (1977) have, however, reported pop-out for certain letters appearing in locations that were supposed to be ignored; it occurred only if the letter had been a target on a previous search task that had been carried out over an extensive period of time. (This therefore is somewhat analogous to the classic 'cocktail party effect' that was described by Moray (1959; cited in Styles, 1998.) Of course, some attentional models would posit that all stimuli are processed in parallel at all stages of perceptual processing, irrespective of whether dedicated mechanisms are available to process them (e.g. Deutsch & Deutsch, 1968 cited in Styles, 1998) but there is little evidence available from the visual search literature to support such a view.

Alternatively, practise effects might be due to learning at the latter levels of decision or response. For example, it could be reflecting the observers' ability to become more adept at monitoring the mechanisms relevant to the search or at extracting signal from noise (tending towards ideal observer performance). Domini & Caudek (1999) speculate that the general improvements they report in slant perceptibility arise because observers learn to ignore the conflicting dot density cue to flatness, also present in the stimulus. If this were true then it is likely that any concomitant increases in search performance are due to some high-level decision stage learning.

4.2 Methods

4.2.1 Equipment and stimuli

The display was viewed monocularly to avoid any binocular cues to flatness (the non-preferred eye was covered with an eye patch) from a distance of 114cm. A notional grid at the centre of the display divided an area of 10°x10° into nine squares. Circular stimulus patches were randomly allocated to the centre points of these squares, the only constraint being that no more than one patch could be allocated to a single square. As the display configuration ensures allocation area remains constant with set-size, confounding eccentricity effects are avoided. Each stimulus patch contained a cycling RDK movie constructed from nine static images presented in a continuous sequence. Each movie image was displayed for four frames; this led to movie durations of 300ms. This duration was sufficiently short to permit the 2-D motions to be good approximations of their ecological ideals. To ensure the movies produced motion impressions of a single sign, 300ms blanking periods were inserted between these movie presentations.

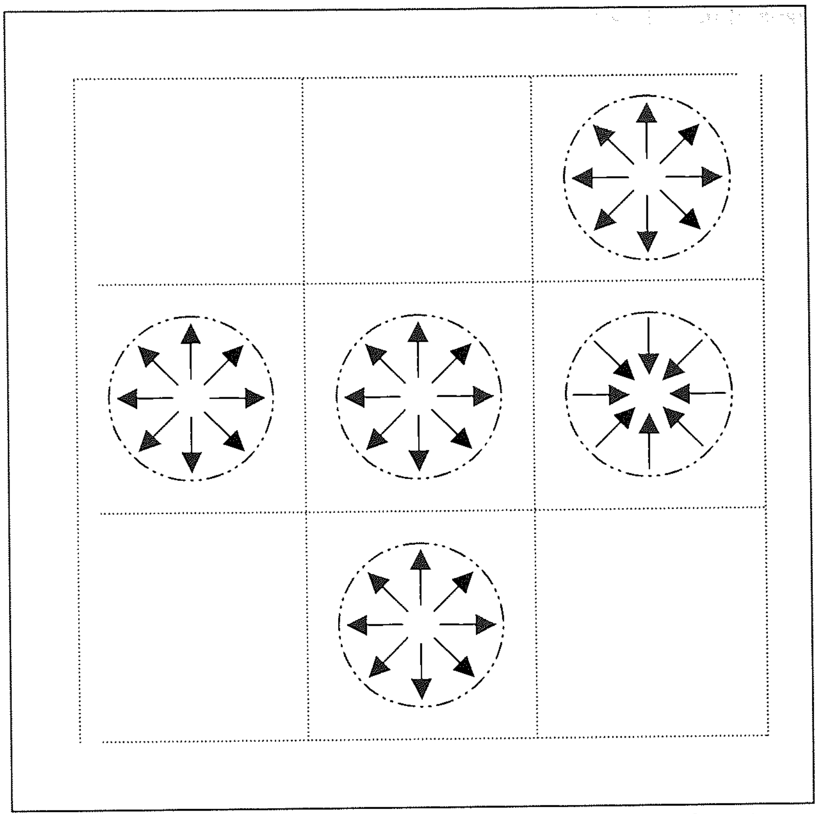


Fig. 4.1. Schematic showing a typical display configuration for a 'target present' trial with a set-size of 5. The dotted grid lines are included only to illustrate the use of a notional matrix in the stimulus patch allocation. Similarly the dashed lines outlining the patches are also purely illustrative and do not appear on actual displays; the patches are demarcated exclusively by the dots' absence.

Seven motion conditions were tested: *trans*, *rad*, *rot*, *defl*, *rot+trans*, *rad+trans*, *defl+trans* (see fig.1.7). The translations were always horizontal, travelling from either left-to-right or right-to-left. When translation components were summed with the complex motions, again the translation was homogeneous, but always travelled from left to right. The stimulus patches had diameters of 3° and contained an average of 144 dots. Complex motion components had linear speed gradients of 0.22sec^{-1} and translation components had homogeneous speeds of $0.67°\text{sec}^{-1}$. This led to maximum dot speeds of $1°\text{sec}^{-1}$. For the translations alone, a homogenous speed of $0.33°\text{sec}^{-1}$ was used.

4.2.2 Procedure

The subject was given extensive experience with all seven motion stimuli until he felt the perceptibility of any 3-D interpretations had stabilised (i.e. the impressions of motion-in-depth for the *rad* stimuli and the impressions of 3-D structure associated with the *defl* and *defl+trans* stimuli). During this period of training, no data were gathered.

Data collection was carried out in 70 experimental sessions. In each session, data were gathered for one of the seven motion conditions. Sessions were carried out in a pseudorandom order, each session containing 110 trials. Eight of these were randomly selected practice trials placed at the beginning in order to familiarise the observer with the motion stimuli. The results from these trials were subsequently discarded. The remaining 102 trials were experimental and logged on to a data file. A standard singleton search task was used, so that in 'target present' conditions only one target (odd-item-out) would ever be present. In the complex motion conditions a target would be defined as a patch containing a complex motion component having an opposite motion sign to that of the distractors. If, for example, in a divergence session the distractors were expanding, the target would be contracting. Note that in conditions upon which complex motion was summed with translation, the translation component always moved rightward, irrespective of whether the patch was a target or not. In the translation-alone condition, the direction of horizontal translation could be either leftward or rightward. A target would be defined as a patch containing a direction of translation opposite to that of the distractors.

The actual motion signs allocated to the distractors were random and so there was always a degree of target uncertainty. For example, within a session containing trials with odd-item-out radial motions, the observer would not know whether the target was an expansion or contraction. A corollary of this is that there is a three patch lower limit on the set-size.

The observer was required to report the presence or absence of a target by pressing the appropriate button on a PC mouse (left indicated 'present' and right indicated 'absent'). They were given feedback as to the correctness of their response using auditory beeps. Both their responses and RT's were logged. Subjects were instructed to make speeded responses but to try to maintain a constant level of accuracy throughout the sessions irrespective of set-size.

Data were gathered for three set-sizes (3, 5 and 9 motion patches). Within a single experimental session, 17 responses were collected for each set-size for each target type (i.e. 'target present' and 'target absent'). The order in which these responses were collected within a session was random.

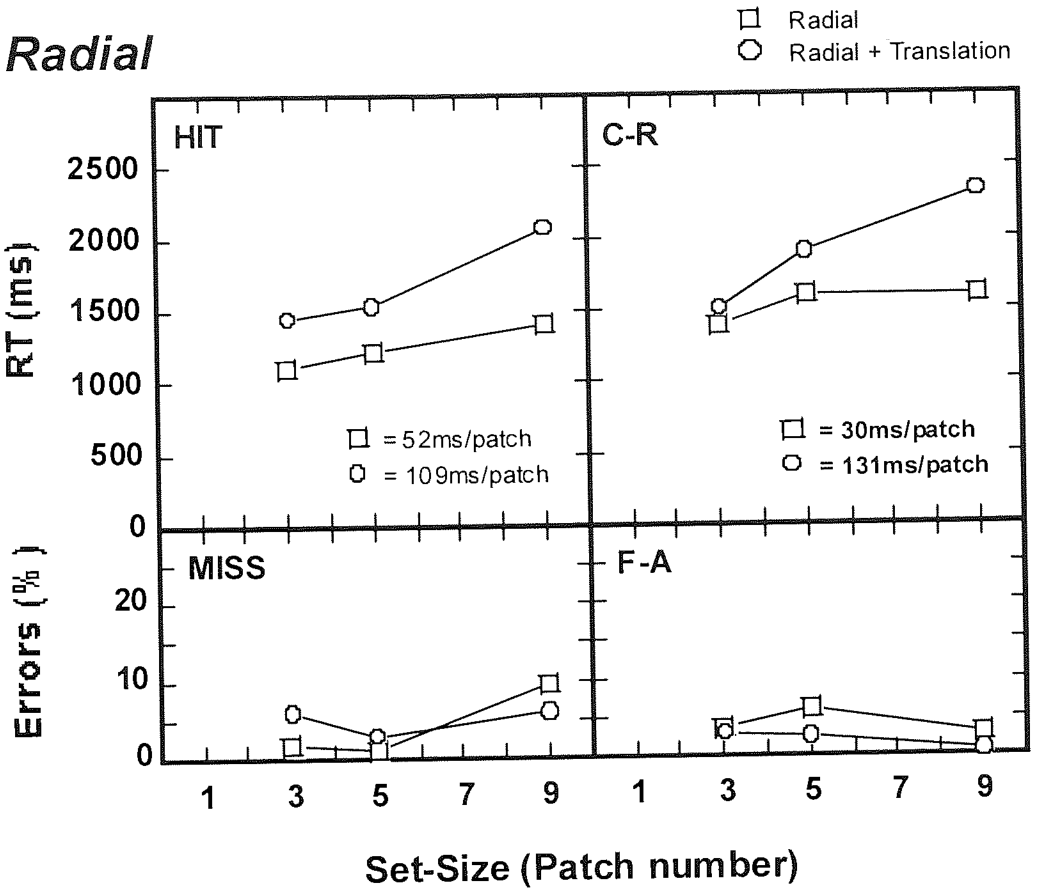


Fig.4.2. Mean RTs for Hit and C-R responses (upper panels). Miss and F-A error rates (lower panels).

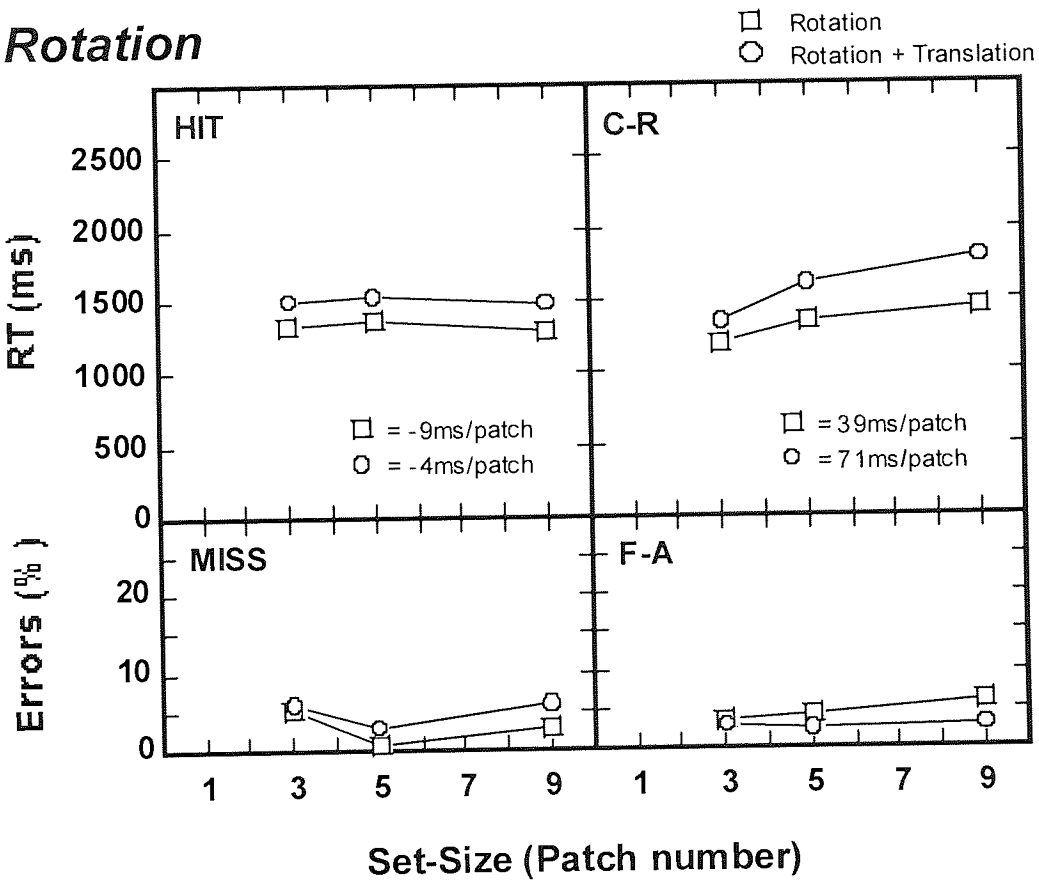


Fig.4.3. Mean RTs for Hit and C-R responses (upper panels). Miss and F-A error rates (lower panels).

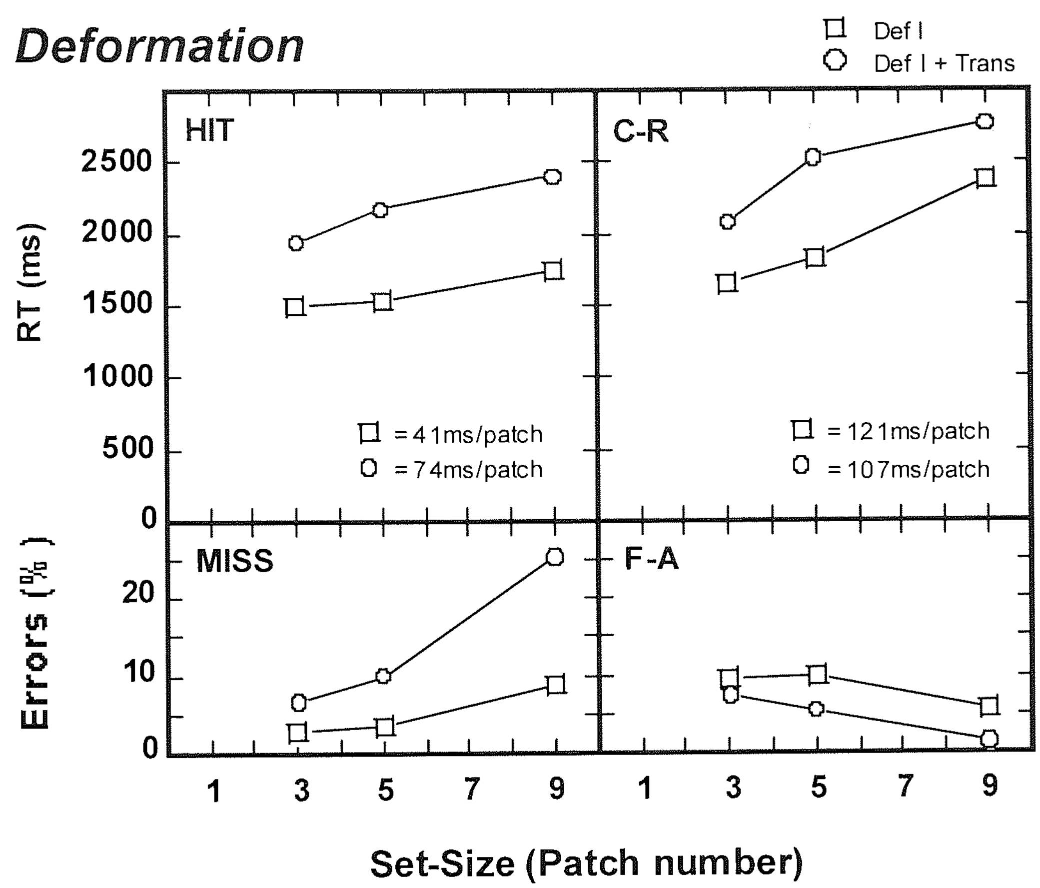


Fig. 4.4. Mean RTs for Hit and C-R responses (upper panels). Miss and F-A error rates (lower panels).

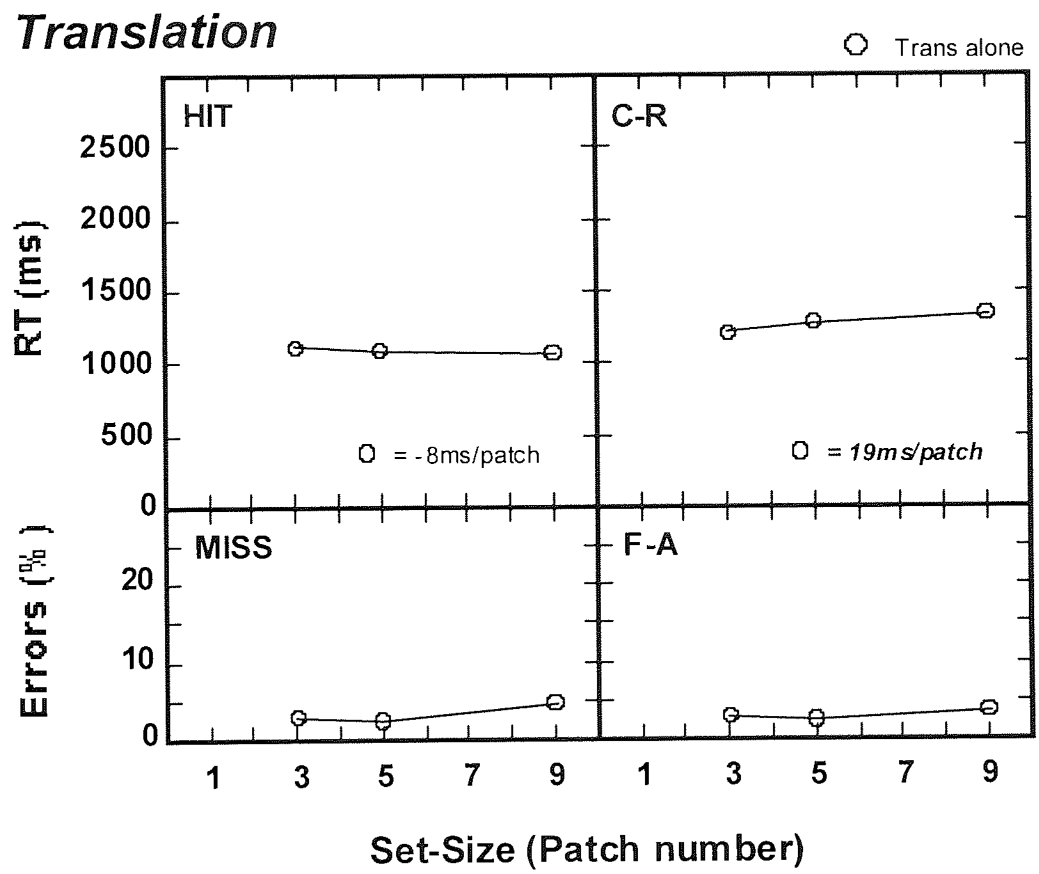


Fig. 4.5. Mean RTs for Hit and C-R responses (upper panels). Miss and F-A error rates (lower panels).

4.3 Results

Subject reports indicate that the sign of motion could be easily discerned for all stimulus types when selectively attended. The mean RTs for Hit and C-R responses are plotted for each motion type, along with error rates (fig.4.2-4.5). Linear regression was used to find lines of best fit and the slope magnitudes were taken to be definitions of set-size effect. For the complex motion conditions, each plot includes data for the complex motion alone, and for when summed with translation. The same axes scales have been used in each of the plots to enable easy comparison between motion types.

Four out the seven motion types yielded Hit slopes greater than 20ms/patch and therefore, in terms of FIT, would be classed as having been searched in serial. These motions were *rad* alone (Hit = 52ms/patch), *rad+trans* (Hit = 109ms/patch) *defl* alone (Hit = 41ms/patch) and *defl+trans* (Hit = 74ms/patch). The three remaining motion types yielded search slopes that were lower than 5ms/patch and would therefore be attributed with parallel search. In fact, all three had slope magnitudes slightly less than zero: rot alone (Hit = -9ms/patch), *rot+trans* (Hit = -4ms/patch), and *trans* (Hit = -8ms/patch).

C-R slopes were, with the exception of *rad* alone, steeper than Hit slopes. The slope magnitudes were as follows: *rad* alone (C-R = 30ms/patch), *rad+trans* (C-R = 131ms/patch), *rot* alone (C-R = 39ms/patch), *rot+trans* (C-R = 71ms/patch), *defl* alone (C-R = 107ms/patch), and *trans* (C-R = 19ms/patch). While C-R slopes tended to be steeper than the Hit slopes, they all were in ratios that differed somewhat from the 2:1 slope ratio predicted by the SSTS strategy of FIT. This 2:1 slope ratio would have been predicted for the four conditions in which Hit slopes indicated a serial search (i.e. slopes greater than 20ms/patch); it is unclear what slope ratios would be expected for the three conditions with parallel-like Hit slopes (i.e. slopes less than 5ms/patch). The one condition in which the C-R slope was found to be less than the Hit slope (*rad* alone) was, in terms of FIT's diagnosis, indicative of being a serial search. Absolute RT performances were however superior for the Hit's than C-R's.

For the *trans* and both of the *rad* and *rot* conditions, error rates tended to be much lower than 10%. The Miss rate for the *rad* alone condition at a set-size of nine did however come close to 10%. For both *defl* conditions, Miss and F-A rates were a little higher. This was especially true for the Miss rates where they increased monotonically with set-size, in one condition reaching as high as 25%. As both the Miss rates and the Hit RT's increased with set-size, the possibility that the high set-size effects were due to a speed-accuracy trade off can be ruled out.

For five out of the six stimulus conditions that contained complex motion components there were conspicuous outward kinks present in their C-R slopes. The one exception was with the *defl* alone condition. These kinks may indicate a violation in the assumption of linearity that is an implicit feature of FIT. To confirm this, a more extensive sampling of the set-size function would be necessary.

For the two complex motion conditions yielding steep, serial-like search slopes (*rad* and *defl*) the addition of a *trans* component made both their Hit and C-R slopes even steeper. Addition of a *trans* component to the *rot* motion condition made little difference to the Hit slope, both being slightly negative. It did however increase the C-R slope from 39ms/patch to 71ms/patch. In all three complex motion conditions and at all set-sizes, RTs increased with the addition of *trans*. This was true for both Hits and C-Rs.

Table.4.1 summarises the search efficiencies for the complex motion alone conditions of the present experiment to those of previous findings. Note that there are differences between the complex alone conditions despite having on average the same distributions of dot velocity.

Translation

The *trans* condition was used as a control to confirm that, in principle, the display configuration and method of stimulus generation could yield parallel-like search slopes. In accordance with previous studies, the slope was near zero (Braddick & Holliday, 1991; Driver, *et al*, 1992; Nakayama & Silverman, 1986).

Radial

The Hit slopes for the radial condition fell somewhere in between previous estimates. Braddick & Holliday (1991) report least efficient search (Hit = 96ms/item; C-R = 210ms/item) while Takeuchi (1997) and Thornton & Gilden (2001) found much shallower search slopes and in the case of the latter study reported a Hit slope of 4.3ms/patch was observed for expansion[1].

---

[1] Note that both Takeuchi (1997) and Thornton & Gilden (2001) report search efficiencies for both signs of motion unlike Braddick & Holliday (1991) and the present experiment which collapsed data over the two signs.

Rotation

The highly efficient search performance of rotation is contrary to the conclusions of Julesz & Hesse (1970) and Thornton & Gilden (2001). The observer's subjective experience in the target-present trials did tend to be of an effortless segmentation.

However, one possibility is that the cue to this segmentation was not the target patch's unique rotation sign in itself, but a cue emerging from interactions between the flows of neighbouring patches. If areas with common translation components can group together across patches for example, the area of this grouping might be used as a cue to the presence of targets. Fig.4.6 illustrates such a grouping. Indeed this account is consistent with subjective reports. Cavanagh *et al* (1990) has already shown size to have a feature-like status even when defined by motion.

If inter-patch grouping is being utilised then the denser displays may be expected to be more effective at doing this because of the shorter distances involved. To test this, dense and sparse displays were compared for a single set size (see fig.4.6 for examples of dense and sparse displays). A dense display was defined as one in which the target is neighboured by two others (i.e. ones residing in grid elements immediately above, below, left or right) and sparse displays defined as any other configuration. Using a two-tailed t-test, no significant difference could be found between the two (p>0.05). If some kind of patch interaction was responsible, whether in the form of grouping or otherwise then it seems not to have been heavily dependent upon patch spacing.

Deformation

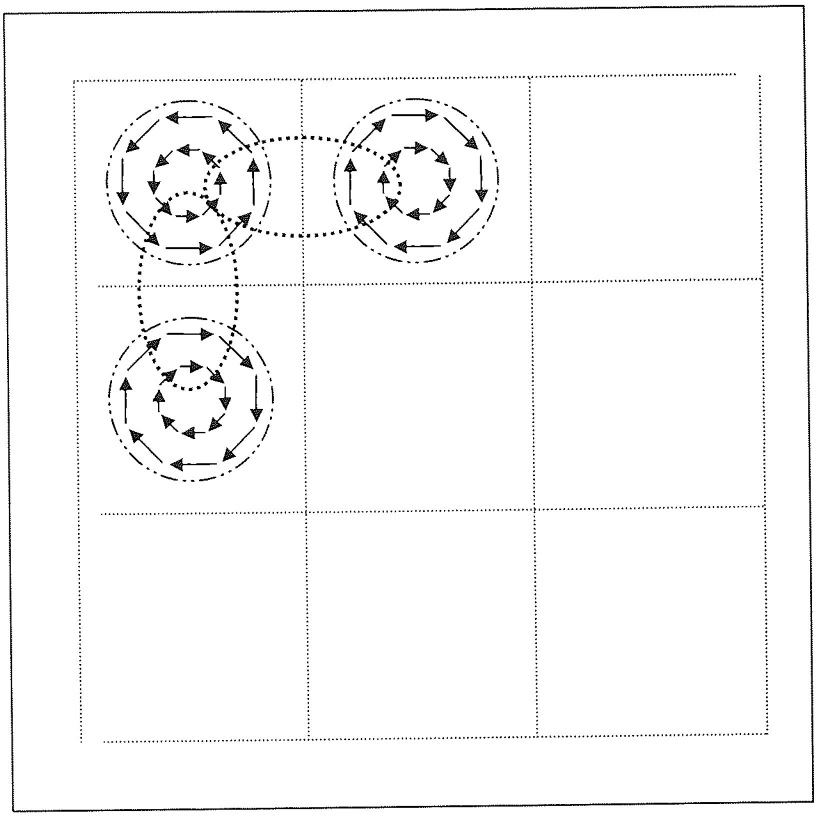
While the *def1* stimuli were searched more efficiently than *rad* stimuli, search was still highly inefficient. Braddick & Holliday (1991) reported the converse: in their study, *rad* stimuli were searched more efficiently than deformation stimuli, *def2*. This may not be an entirely valid comparison. In terms of motion coherence thresholds, Meese & Harris (2001a) did find differences between *def1* and *def2*.

Subjective reports also suggested that when the patches were clustered together in the *def1* and *def1+trans* displays search become easier. To test whether this was the case, a t-test was used to compare performance between sparse and dense display configurations at a set-size of three. Significant differences were found between the RTs of both Hit (p<0.05; 277ms mean difference) and C-R (p<0.01; 200ms mean difference) for the def1+trans condition with performance being superior for dense displays. No significant differences were found for the def1 condition.

	Rotation			**Radial**			**Deformation**		
	HIT	**C-R**	**Ratio**	**HIT**	**C-R**	**Ratio**	**HIT**	**C-R**	**Ratio**
Braddick & Holliday (1991)				96	210	x2.2	[Def2] 276	[Def2] 542	[Def2] x2.0
Takeuchi (1997)				[Cont] 32.4 [Exp] 4.3	[Cont] 58.0 [Exp] 5.1	[Cont] x1.8 [Exp] x1.2			
Thornton & Gilden (2001)	[Clk-wse] 56			[Cont] 13 [Exp] 16					
Expt 1	-9	39	n/a	52	30	x0.6	[Def1] 41	[Def1] 121	[Def1] x2.9

Table.4.1. Set-size slopes derived from singleton visual search experiments. Slopes are in ms/stimulus. Due to the blocked nature of some of these experiments, slopes for the two signs of motion are considered separately. Also included where appropriate are the slope ratios between 'Hit' and C-R'. Note that the deformations used in Experiment 1 of this thesis and those used by Braddick & Holliday (1991) differed in type casting some doubt on the validity of making direct comparisons between the two. Moreover however, the general nature of the stimuli and display configurations also differed substantially between these four experiments.

A



B

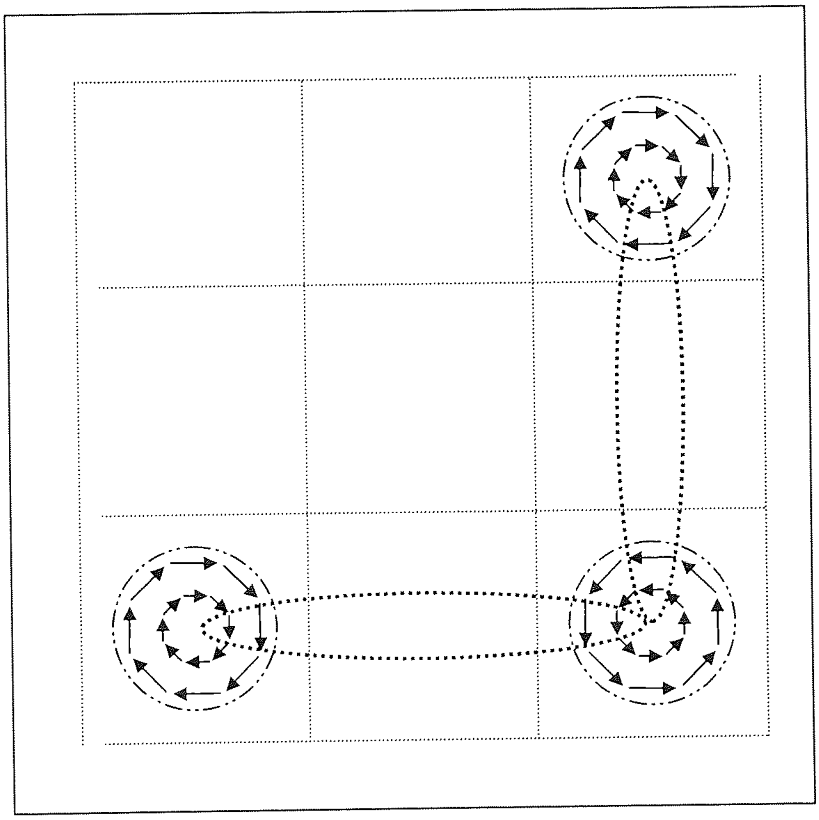


Fig.4.6. Two schematics showing the putative interactions between patches of rotation. Both show set-sizes of three with a target present (the cclk-wse rotations). **A**, is a dense configuration because the three patches are clustered together while **B**, is a sparse configuration. All dotted lines are only for illustrative purposes and do not appear on the actual displays. The dotted ovals represent the general areas in which the speculative grouping that was mentioned might be operating.

4.4 Discussion

4.4.1 Density effects

Subjective reports suggested that the density at which stimulus patches are presented affects the RTs for *defl+trans* motions. A comparison between dense and sparse display configurations did reveal more efficient search for dense configurations than sparse configurations. Braddick & Holliday (1991) have also reported superior search efficiencies for displays with smaller inter-stimulus spacing: Hit and C-R slopes for *rad* flows were observed to decrease by more than a half for one subject (Hit slope reduced from 77ms/item to 30ms/item; C-R slope from 159ms/item to 40ms/item). The researchers did not speculate as to the reason for these performance decrements, but one possible explanation for both these findings and those of the present experiment is that the distance travelled by the putative attentional spotlight is less for smaller stimulus spacing. On average therefore, the time needed to serially scan all stimuli would be lower.

4.4.2 Inter-stimulus interactions

Other subjective reports raised concerns that inter-stimulus interactions were responsible for the high search efficiencies of the *rot* and *rot+trans* conditions and not the *rot* motion components in themselves. Braddick & Holliday (1991) also recognised the possibility of interactions between the stimulus items of a search display, but considered them only to be deleterious to search performance. The interactions proposed here would instead lead to improvements.

It is puzzling however, why such cues were not also observed in the displays of the previous two studies that presented multiple rotation flows simultaneously (Julesz & Hesse, 1970; Thornton & Gilden, 2001). This may be less so for the study of Julesz & Hesse (1970) because they used very different stimuli but not so for Thornton & Gilden (2001) who had also used RDKs, albeit in a spatially smoothed form.

4.4.3 What insights do the search slopes give to the existence of specialised mechanisms?

Out of the six conditions containing complex motion components, only the two *rad* conditions showed signs of parallel search, and this was suspected to reflect the utilisation of inter-element cues rather than the existence of specialised complex motion mechanisms.

The steep, serial-like search slopes observed in the *rad* and *defl* conditions cannot be used as definitive evidence against the existence of multiple complex motion mechanisms (Braddick, 1993). It may be the case that the outputs of such mechanisms are simply unavailable for use in the search tasks (see fig. 1.8)

4.4.4 Is the '3-D access hypothesis' supported?

In the introduction to this chapter (section.4.1), it was pointed out that the '3-D access hypothesis' would predict search in the *defl+trans* and *rad* conditions to be more efficient than the *defl* alone condition. The pattern of results for two out of the three subjects in the Freeman *et al* (1996b) study were consistent with this. However, the results from experiment 1 found the converse to be true: *defl* search was more efficient than both the *defl+trans* and the *rad* search.

Upon initial observation, the data from experiment 1 appear unanimously at odds with the '3-D access hypothesis'. Subjective reports from the observer suggest that this may not necessarily be true however. To this observer, not only did the *defl* flows appear three-dimensional but its' two polarities of motion appeared reliably different from each other. The two 3-D impressions are illustrated in fig.4.7. *Defl* with a positive sign consistently provoked subjective impressions of a rigid surface carrying out a rotational movement (roll) around an axis tilted to -45 degrees (the bottom-right corner moved towards the observer). Negative *defl* consistently provoked the same impression but with the perceived roll in the opposite direction (the bottom-left corner moved towards the observer) and with the axis tilted to +45 degrees. These impressions of 3-dimensionality were consistent with structure-from-motion (SFM) interpretations rather than the motion parallax interpretations of the *defl+trans* stimuli.

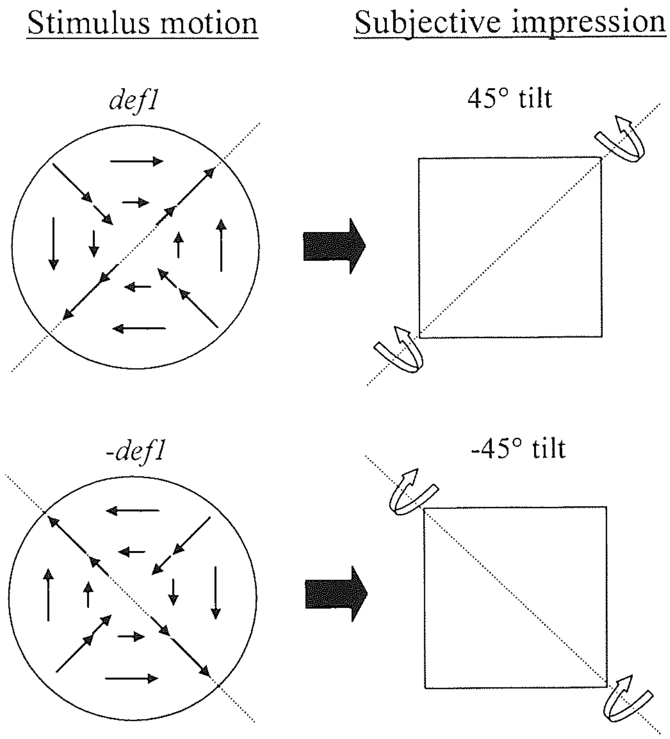


Fig.4.7. 3-D impressions generated from defl motion for subject PAA. The diagrams on the left illustrate the two stimulus types that were employed in the '*defl* alone' search displays. The notional aperture boundaries have been included along with the axes of negative compression (dashed diagonal lines). On the right are the corresponding impressions of 3-dimensionality that the subject reported. Clearly, the impressions are not wholly consistent with the motion flows that would actually be generated from a rigid planar surface rotating about an oblique axis. Firstly, the positive compressive speed gradient component of the *defl* stimuli seemed not to contribute significantly to the 3-D percept. Secondly, unlike the flow fields that would be induced by planar surfaces rotating about an axis perpendicular to the line of sight, these did not develop over time (i.e. because the deformation vector remained constant: pure *defl*).

Because, in the *defl* condition, the angle of axis and direction of roll were readily available to the observer, these cues – emerging from 3-D interpretations - might have been used as the basis for search. Indeed, the observer felt that this had been the case; if true the '3-D access hypothesis' would not have been disproved.

Other observers in the *defl* condition might also base their search performances upon cues derived from 3-D interpretations. The value of comparing *defl* conditions to *defl+trans* and *rad* conditions is therefore compromised because the experimenter will be uncertain as to whether *defl* search had been based upon 2-D or 3-D stimulus descriptions.

4.4.5 How well does classical FIT account for these results?

For parallel searches, FIT only makes firm predictions about the Hit slope being close to zero, saying nothing about what the magnitude of C-R slope should be. For serial searches, FIT dictates a SSTS strategy. In this the set-size, slope is assumed to be solely dependent upon the dwell time; i.e. the time taken for a single attentional fixation to be made to a stimulus patch and a decision be made as to whether it should be accepted or rejected as a target. It is conceivable that, for some conjunctions, dwell times would be small, but a general rule that seems to have stemmed from Treisman & Gelade's (1980) original findings for conjunctions is that search is classed as serial if the target-present (Hit) slopes are 20ms/stimulus or above. This translates into dwell times of 40ms or above, because on average only half of the stimuli need to be searched in order for the target to be detected. Therefore, in terms of FIT, both the deformation and radial conditions would have dwell times in excess of 40ms.

The two other predictions made by SSTS are (1) set-size effects should be linear and (2) the C-R and Hit slopes should be in a 2:1 relationship. This is expected because for the detection of a target, only half of the stimuli need to be searched, whereas when a target is not present, all stimuli will need to be searched in order to confirm target absence. There were some instances when this first prediction appeared to be violated. RT/set-size relationships often seemed somewhat off linear, but because RTs were measured for only three set-sizes it is unclear whether this was genuinely the case. The second prediction certainly seems to have been violated, as no condition came close to a 2:1 slope ratio. Pashler (1987) has previously pointed out that both predictions are sometimes violated when the set-size range covers small set-sizes like in this experiment (Pashler found this was so for set-sizes below 8). Results showed that the C-R slopes tended to be shallower in this range. This also seems to have been the case for the two *rad* conditions of this experiment. The *defl* conditions showed the opposite trend with C-R slopes, being more than twice as steep as the Hit slopes. This might however be the due to a speed-accuracy trade-off (Miss rates tended to increase monotonically in these conditions).

It seems therefore to be the case that, for the conditions in which there is inefficient search, SSTS is not a sufficient model for the data. The veracity of SSTS has been questioned on other occasions since being first proposed by Triesman & Gelade in 1980. Ward & McClelland (1989), for example, have previously found RT variances for C-Rs to be inconsistent with the SSTS. An assumption implicit in SSTS is that the observers adopt an ideal sequential search strategy in which each stimulus item is checked only once. Klien (1988) has shown that the visual system does have an aptitude for knowing which stimuli have been analysed already. Such ability would indeed be necessary for large set-sizes (for example Takeuchi's (1997) maximum set-size was 25 patches).

4.4.6 Search efficiency may be limited by stimulus quality

Table 4.1 shows that there are large discrepancies in the search slope estimates between the same motion conditions of different search experiments. For example, the *rad* motion condition has been diagnosed as both serial and parallel in different studies. The slopes obtained from the present experiment did not corresponded well with the results of the previous experiments, rather they tended to fall somewhere in between.

What might be the cause of this large variability across studies? They may in part be due to subtle differences in experimental conditions, such as the use of different set-size ranges (Pashler, 1987) and display configurations. Another possibility is that they are due to differences in stimulus saliency. Stimuli that induce more compelling impressions of global motion might be searched with greater efficiency. Indeed, an empirical study by Treisman & Gormican (1988) has confirmed, more generally, that increasing target and distractor saliencies can lead to improvements in search performance. This potentially important point of stimulus saliency will be the subject of experiment 2, where more compelling impressions of surface slant are generated from 1-D speed gradients.

Chapter Five: Experiment 2 - 1-D speed gradients

5.1 Chapter Introduction

The results of experiment 1 brought into question several aspects of the stimulus design. It became apparent that surface representations could have been providing the observer with reliable cues in the *defl* alone condition, as well as the *defl+trans* and *rad* conditions. The *defl* motions elicited impressions of a rolling planar surface, in which both the direction of roll and the axis upon which the roll took place were candidate cues for search. It was therefore concluded that comparison of the slope magnitudes for the *defl* and *defl+trans* conditions were an inappropriate means for assessing the '3-D access' hypothesis. While in some observers the differences in slope might reflect the utilisation of different representations of motion (e.g. 2-D representations for *defl* and 3-D representations for *defl+trans*), in others it might simply reflect differences in the efficacies with which 2-D or 3-D cues are used (e.g. slant representations being more readily accessible than the direction or axis of roll).

An alternative approach to the assessment of the '3-D access' hypothesis is to seek stimuli which can only be searched using 3-D surface descriptions/motion differentials but produce near zero search slopes. This would indicate a parallel processing and a simultaneous accessing of representations.

The steep search slope for the *defl+trans* condition of experiment 1 provided no evidence that the surface descriptor slant can be accessed in parallel. However, it has been shown that when slant sign is defined by other cues, such as stereoscopic (Holliday & Braddick, 1991) or pictorial cues (Enns & Rensink, 1991) parallel search is observed (see section 3.3.6). This prompts the following question: were the above parallel slant sign searches based upon the observers' ability to simultaneously monitor cue-specific representations of slant[1] (stereoscopic slant representations in the Holliday & Braddick, 1991, study and pictorial slant representations in the Enns & Rensink, 1991, study) or their ability to simultaneously monitor cue-independent representations of slant? If the latter were true then similar parallel search might also be expected for slants defined by motion gradients. The failure to find evidence of parallel search for the *defl+trans* condition of the last experiment might be attributable to the stimulus' inability to produce sufficiently compelling impressions of horizontal slant (Treisman & Gormican, 1988).

---

[1] In chapter 3 (section.3.3.4), it was noted that search based upon 2-D orientation is processed in parallel, irrespective of the cue used to define the orientation.

The points above raise two issues. (1) Is there any evidence to suggest the existence of cue-independent 3-D descriptor mechanisms (such as those selective to slant)? In section 1.4, studies were discussed that suggest such mechanisms probably do exist. (2) If true, can sufficiently compelling impressions of slant - generated from motions with other types of speed gradient - stimulate these mechanisms sufficiently to permit parallel search? In the discussion of the previous chapter (section 4.4.6) it was speculated that stimulus saliency might well be a factor in visual search efficiency. To explore this issue, in the following experiment 1-D vertical speed gradients were used. These generated impressions of planar horizontal slant, which, subjectively, appeared more salient than those of *defl+trans* stimuli and provided a much closer match to the flat planar surfaces that were generated in the Holliday & Braddick (1991) study.

If parallel search can be found for 1-D motion gradient sign then an argument could be made for considering slant *per se* to be a "feature". This assertion might be contested by proponents of models such as Revised FIT and Wolfe's (1989) Guided Search model (see section.3.3.5) in which they might argue that some seemingly arbitrary feature conjunctions are searched in parallel (e.g. Wolfe, 1989, found near zero set-size slopes for a colour-form conjunction search). As Holliday & Braddick (1991) have already pointed out though "these have all involved conjunctions between different stimulus dimensions (e.g. colour and motion) [*across-item-conjunctions*], and the finding that conjunctions within a single dimension [*within-item-conjunctions*] give steep search-time functions has so far been robust".

5.2 Stimulus and experimental design

The experimental design of experiment 2 is similar to that of experiment 1, in that it also uses a standard singleton search paradigm with RDK patches for the individual stimuli. However, in this experiment, the emphasis is shifted from looking at the relative differences in search efficiency to seeking near zero slopes for when the motion stimuli contain 1-D and not 2-D speed gradients. These 1-D gradients were used because they produced (a) unambiguous impressions of slant, and (b) subjectively more salient impressions of slant.

While it is acknowledged that there have been various criticisms levelled against the use of set-size slopes in the diagnosis of search strategy, these have primarily been directed at the misdiagnosis of parallel search; the implication being that parallelism can sometimes yield positive slopes. This might be due either to the fact that high-threshold theories (HTTs) of visual search are inappropriate, or because search is limited by a fixed attentional capacity (these issues are discussed in section 7.2.4). Slope magnitudes are therefore incapable of distinguishing between two types of limited-capacity search (serial versus inefficient parallel) but they are capable of identifying parallel searches with no capacity limitations.

A number of additional modifications have also been made because of observations from the last experiment. These are also detailed below:

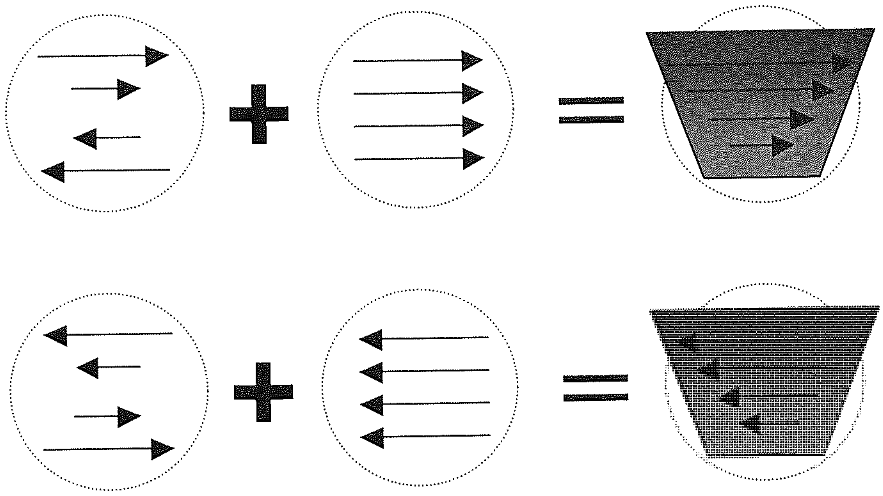
5.2.1 Use of 1-D motion gradients to produce compelling slant impressions

Preliminary experimentation was carried out with single RDK patches. The use of 1-D vertical speed gradients summed with horizontal translations were found to give subjectively more compelling slant impressions than either of the *defl* or *defl+trans* stimuli of experiment 1 (see fig.5.1 for the possible stimulus construction). In terms of the surface representations produced, these are analogous to the stereoscopic stimuli of Holliday & Braddick (1991): flat, slanted surfaces with a 90° tilt. The compelling impression of slant from such motion stimuli was first noted by Braunstein (1968) and has been confirmed numerous times since. It is these stimuli that will be used to test the working hypothesis - when stimulus saliency is increased to levels comparable to those of the stereoscopic slants used by Holliday & Braddick (1991) near zero set-size effect will be observed.

5.2.2 Reducing inter-patch interactions

For two of the conditions in experiment 1 (*rot* and *rot+trans*) it was suspected that search had been based upon spatial interactions between the motions of the RDK patches. If two neighbouring rotations had the same sign, then a shear would be present at their boundaries. If they had opposite signs however, they would produce unidirectional motion. If observers can efficiently comprehend the areas of unidirectional motion then they may also permit the efficient detection of 1-D motion gradients of an odd sign. Indeed preliminary experiments on a single subject, looked at 1-D motion gradients (horizontal and vertical) both with and without translation and found no difference between the two in terms of search performance. Both yielded near-zero set-size slopes. This strengthened concerns that inter-patch cues had been responsible for search performance, as the 1-D motion gradient conditions (without translation), did not produce unambiguous slant impressions. To ensure that the only reliable cues to the presence of a target comes from within the patches themselves, surface representations were constructed in two ways. In the case of a positive shear (floor-like) stimulus, they could be constructed by summing a positive slant with a leftward translation or a negative shear with a rightward translation. In the case of negative slant (ceiling-like) stimulus they could be constructed by summing a positive shear with a rightward translation or a negative shear with a leftward translation (see fig.5.1 for a schematic). The method of construction for each patch was random and independent of other patches. Because of this, the presence or absence of shearing at patch borders can no longer act as an effective cue to the presence of a target.

Top slanting towards observer (ceiling-like)



Top slanting away from observer (floor-like)

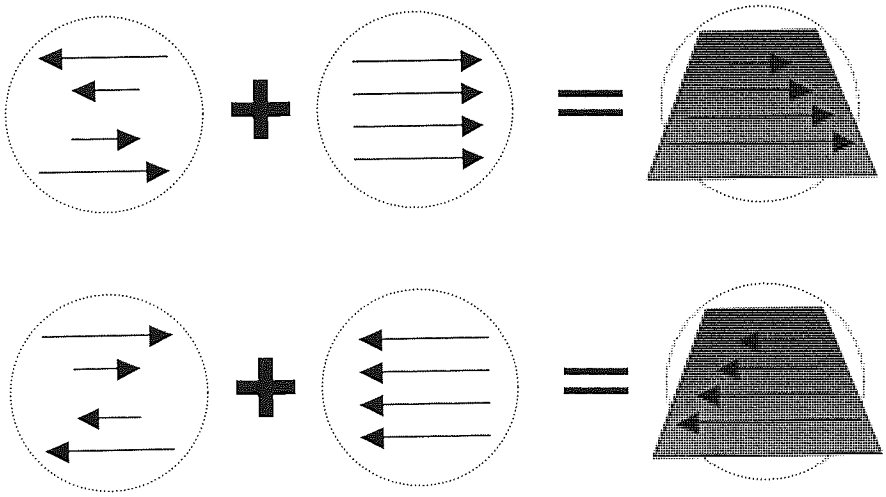


Fig.5.1. Construction of the 1-D stimuli. These two schematics illustrate the construction of two motion parallax stimuli and their resulting perceptual surface impressions. On the left are the two motion components that are summed: shears (with a vertical speed gradient) and lateral homogenous translations. On the right are the resultant parallax flows. The shaded trapezium shapes were included to illustrate the slant impressions accompanying the flows (i.e. a square planar surface in perspective). Note that the motion vectors have not been drawn to scale.

In an attempt to lessen the effect of patch interactions, the spacing between the patches was also increased. This was achieved by reducing the size of the patches. The degree to which they could be reduced was limited by the saliency of the slant impressions that small patches could produce: if the patch diameters were too small, the saliencies of these impressions were compromised.

5.2.3 Reducing inter-patch grouping

Distractor grouping has been described previously (Duncan & Humphreys, 1989). For example, when an observer searches for a vertical line amongst a set of oblique (e.g. 45°) lines, the oblique distractors will tend to group together based on their similarity (i.e. a common orientation). This seems to enable them to be rejected by the observer as a group rather than as a number of individual stimuli and is one explanation as to why search performance is often better with homogenous distractors than heterogeneous distractors. It may also account for the curious situations - like the rotation conditions of the previous experiment - in which the RTs for easy search tasks actually decrease with additional distractors. Negative set-size slopes may therefore be indicative of some kind of texture segregation rather than search based on the comparisons of independent analysers. It is possible however, that distractor grouping may, under some circumstances, have the opposite effect tending to decrease search efficiency. This could arise from a grouping actually masking the unique attribute defining the target and so the saliency of the target-distractor differences would decrease as a function of set-size.

It was speculated that use of the same method of patch allocation as in the last experiment could actually facilitate distractor grouping even more so in the present experiment. This would occur because each of the distractors would possess common magnitudes of slant (each tilted to 90°). Because of the Gestalt law of 'good continuity', grouping by slant might therefore occur on a row-by-row basis. To lessen any such grouping, the co-alignment of patches was prevented through, (a) the tilting of the allocation grid by 22.5° and (b) spatially jittering the location of each patch by a small amount (see Fig.5.2).

5.2.4 Clustering of RDK patches

Following on from the observation that dense and sparse displays are searched with a different efficiencies (the *defl* and *defl+trans* conditions of experiment 1) it was decided that patch density should remain the same irrespective of the set-size. To do this, patches were not allocated randomly to the grid squares of display as in the first experiment but were allocated in clusters, although the configuration and the allocation of these clusters remained random. To achieve this, the allocation followed a set of rules that acted to ensure the amount of space that goes un-neighboured by the displayed patches is minimised. For example, for patches belonging to a display with a set-size of four, each would be allocated to grid squares sharing a border with two other grid squares containing patches. In addition, this cluster of four would be allocated randomly within the display grid. As there are nine potential allocation squares in total for this particular set-size, there are four positions the cluster could appear in. If a target is present, then it could appear as any one of the four patches (see an example display for a set-size of six in fig.5.2).

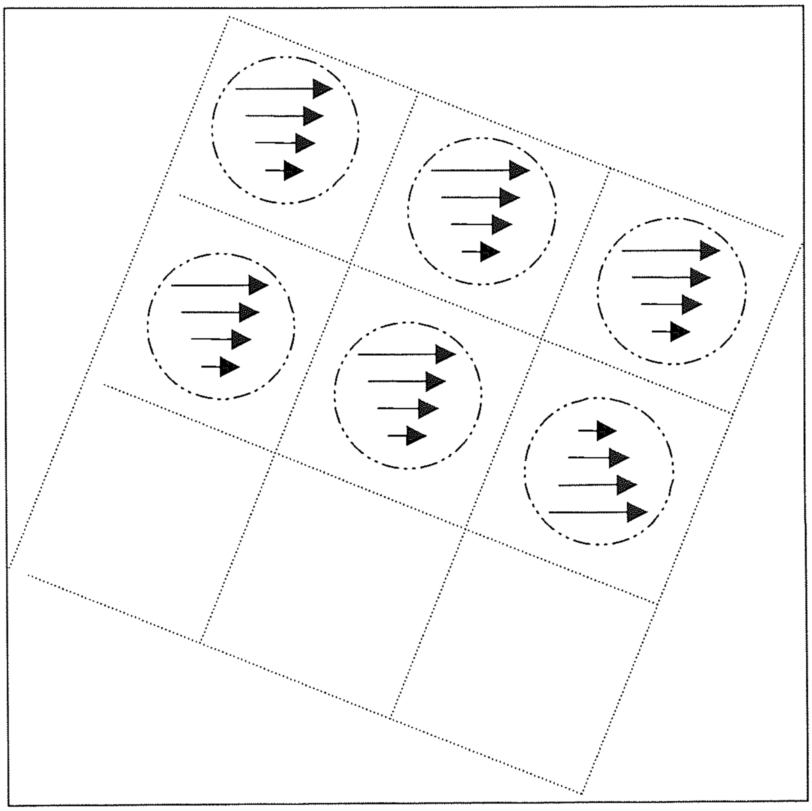


Fig.5.2. A schematic showing a typical display configuration for a 'target present' trial with a set-size of 6. The target is located in the mid-right grid sub-section. The grid was reduced in size and rotated by 22.5°. The patches were clustered as shown for every trial.

5.2.5 Use of a continuous motion stimulus

Both the studies of Braddick & Holliday (1991) and experiment 1 of this thesis used stimuli with a cyclical nature. They found large set-size effects for radial motions. However, a study by Takeuchi (1997) found parallel search when an expansion was to be detected amongst contractions. A possible reason for the discrepancy between studies may have been due to the different types of stimuli used. In the former two studies, a short movie would be presented and then restarted once it had finished[2]. The stimulus cycling would continue until a response was made. Conversely, Takeuchi (1997) used continuous

---

[2] In experiment 1 of this thesis brief blanking periods were inserted between the end and the restarting of the movie. This was found to be necessary in order to maintain impressions of motion of a single direction.

periodic stimuli. This may have improved the saliency of the stimulus either by reducing the amount of noise in the stimulus or by eliminating the possibility of apparent motion in the opposite direction. Consequently, search efficiency may have improved. To avoid the potentially deleterious effects of cyclical stimuli, experiment 2 used a continuously cycling movie of limited-lifetime dots.

5.2.6 Greater sampling of set-sizes

In experiment 1, concern was raised as to the linearity of the *Hit* slopes. FIT predicts linear relationships between set-size and RT. However, an outward kink often seemed to be present in these functions. It was unclear as to whether this was a genuine property of the data or not. Previous studies have suggested that this assumption of linearity may sometimes be violated (Pashler, 1987). For one of the observers in this experiment a more detailed description of the set-size functions was obtained by collecting data from all the set-sizes between one and nine. To obtain RTs for the set-sizes of one and two, sessions needed to be blocked in terms of target type, as for these set-sizes, targets could not be identified on the basis of being an odd-item-out. Therefore, for all trials of a single session the target would remain the same (i.e. either floor-like or ceiling-like).

5.2.7 Search asymmetry between floor-like and ceiling-like targets

Data from the two target types, floor-like and ceiling-like were not collapsed as was done by Holliday & Braddick (1991). The reason for this originates from a report by Morita & Kumada (2003), who found superior search performance when stimulus items adhered to ground-like surfaces as opposed to ceiling-like surfaces. Their study used pictorial cues to induce these impressions of slant, but McCarley & He (2000, 2001) found the same asymmetry in search times when the display items adhered to stereoscopically defined surfaces. Although the nature of these search tasks were fundamentally different from the present experiment, it is possible that the same asymmetries could be present when searches are based on local regions of slant and the task is to detect an odd slant sign.

5.3 Equipment and stimuli

A VD of 114cm was used with monocular viewing. Differences between the display stimuli of experiment 1 and experiment 2 are discussed below where appropriate.

The allocation grid covered an area of 9.3°x9.3° and was tilted 22.5°. When present, stimulus patches were 2° in diameter and as before consisted of high-contrast random dots moving with 100% coherence. Patches were clustered together, as described in section 4.2.4. The motion transforms were constructed from a horizontal shear component summed with a horizontal translation component (see fig.5.1). These generated flows consistent with flat slanted surfaces, which would be either floor-like or ceiling-like.

The motion components that seemed to produce the most compelling slant impressions were homogeneous translations of $\pm 1.15°\text{sec}^{-1}$ (speed) summed with shear components of $\pm 1.61\text{sec}^{-1}$ (vertical speed gradient). These components were used in the 'slow' condition. While this condition produced unambiguous slant perceptions with both subjects, strictly speaking they were not representative of any stimuli that can be generated from relative object-observer translations alone. This is because, the use of these motion components within an aperture of a 2° diameter leads to motions in which not all the dots move in the same direction (i.e. there are stationary points towards either the top or bottom borders). Observer PAA also carried out a second 'fast' condition. This also contained a shear component but in this case all dots moved in the same direction: homogeneous translation $= \pm 2.15°\text{sec}^{-1}$ (speed) and shear $= \pm 1.08\text{sec}^{-1}$ (vertical speed gradient). The slant impressions elicited from these stimuli did not appear as compelling however, possibly because they appeared less slanted.

Dot lifetimes were limited. On each movie frame, a certain proportion (14.3%) would disappear and reappear in new, random locations. Because the 'births' and 'deaths' of the dots were staggered throughout the movie, this enabled movies to run continuously. Each dot had a lifetime of 175ms (a dot trajectory would consist of 7 images each presented for 3 frames) unless terminated prematurely because of movement outside the patch boundary. Because the speed gradient is one dimensional, the dot velocities did not need to change over time; this permitted the use of dots with larger trajectories. By using large dot trajectories the innate noise associated with the limited-lifetime dots (the flickering caused by dot birth and death) is minimised, and so promotes stimulus saliency.

To avoid the abrupt onsets and offsets of dot luminance due to their birth and death, the luminance of each dot were smoothed by ramping them on and off with a sinusoidal function during their lifetimes. This produced displays with compelling slant impressions without distractive twinkling effects that occurred without the luminance ramps.

5.4 Procedure

Two subjects participated in the experiment. One was the author (PAA) and the other a naïve observer (HRC) who was unaware as to the purpose of the experiment, but did have some previous experience with psychophysical tasks. PAA gathered data for both slow and fast conditions. HRC gathered data only for the slow condition.

The search task was carried out in blocks of either 175 (subject HRC) or 195 (subject PAA) trials. At the beginning of each block there were 15 randomly selected practice trials so that subjects could re-familiarise themselves with the task and the stimulus condition. These trials were identical to the experimental trials but the RTs were not recorded. Data were gathered from PAA for all set-sizes between 1 and 9 in attempt to obtain a detailed description of the set-size slope. The second subject, HRC recorded from a limited number of set-sizes: 3, 4, 7, and 9. Each block consisted of 10 (PAA) or 20 (HRC) trials for each set-size, for both target-present and target-absent displays. Trials were intermixed and presented in random order. The target remained the same throughout a block (i.e. a floor-like slant or a ceiling-like slant). This enabled RTs to be obtained for the two lowest set-sizes: 1 and 2 (as there was no target uncertainty).

In this experiment, two practice sessions were carried out for both target conditions so as to familiarise the subjects with the task and stimuli. After this, subject PAA, carried out 20 experimental sessions in a pseudorandom order (10 on each target condition, resulting in a total of 100 RT estimates per condition) and subject HRC carried out 14 experimental sessions (7 on each target condition, resulting in a total 140 RT estimates per condition).

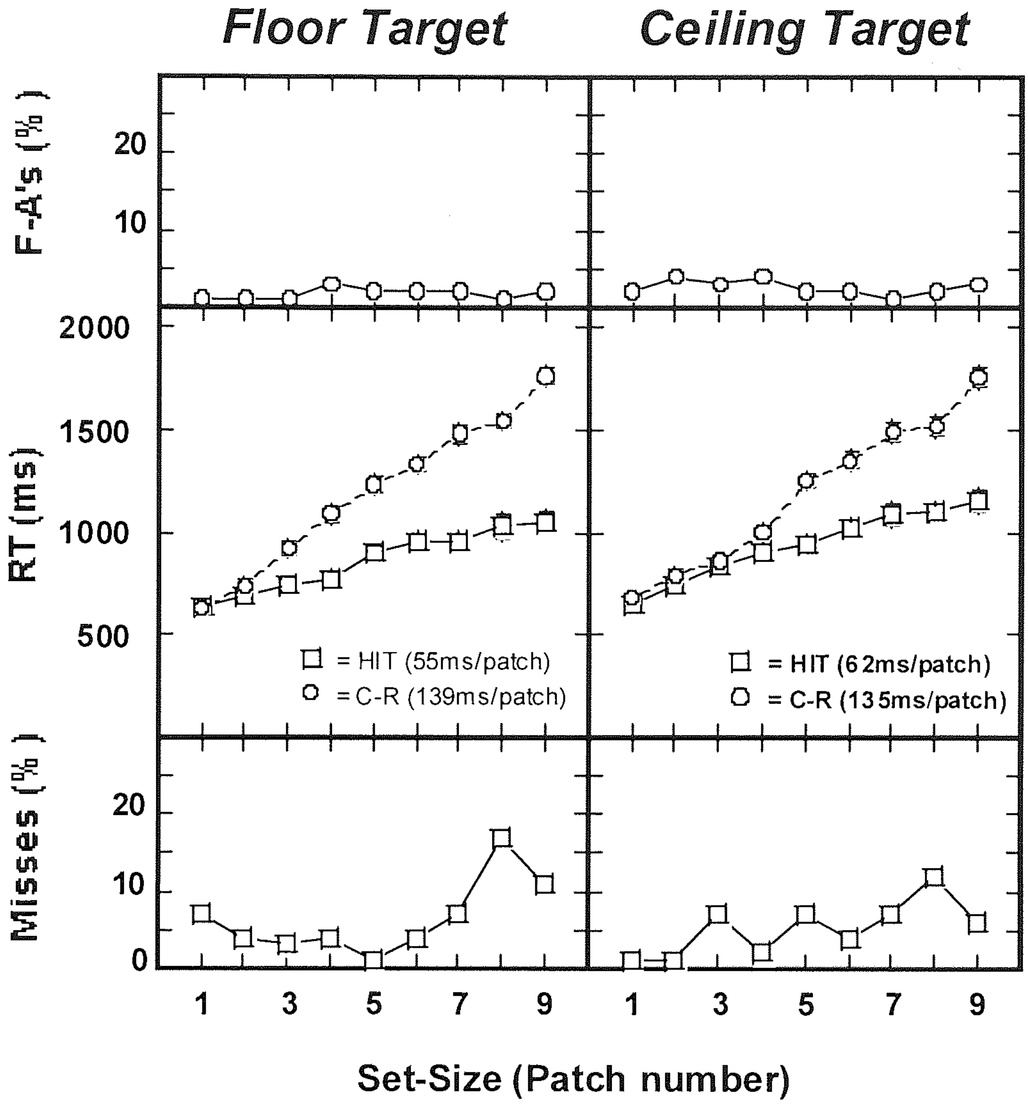


Fig.5.3. Slow condition (Homogeneous translation = ±1.15°sec^{-1}; shear component = ±1.61sec^{-1}) for subject PAA. The left-hand column shows data for searches where the targets were most consistent with floor-like representations and hence the distractors were ceiling-like. The right-hand column shows the converse. Corresponding error rates are plotted in terms of error type (Misses below & F-As above. Linear-regressions were carried out on the reaction time data for Hits & C-Rs. The slope estimates are inset in plots (bracketed values). SE bars are included, but because of small variabilities they are often concealed by the data symbols.

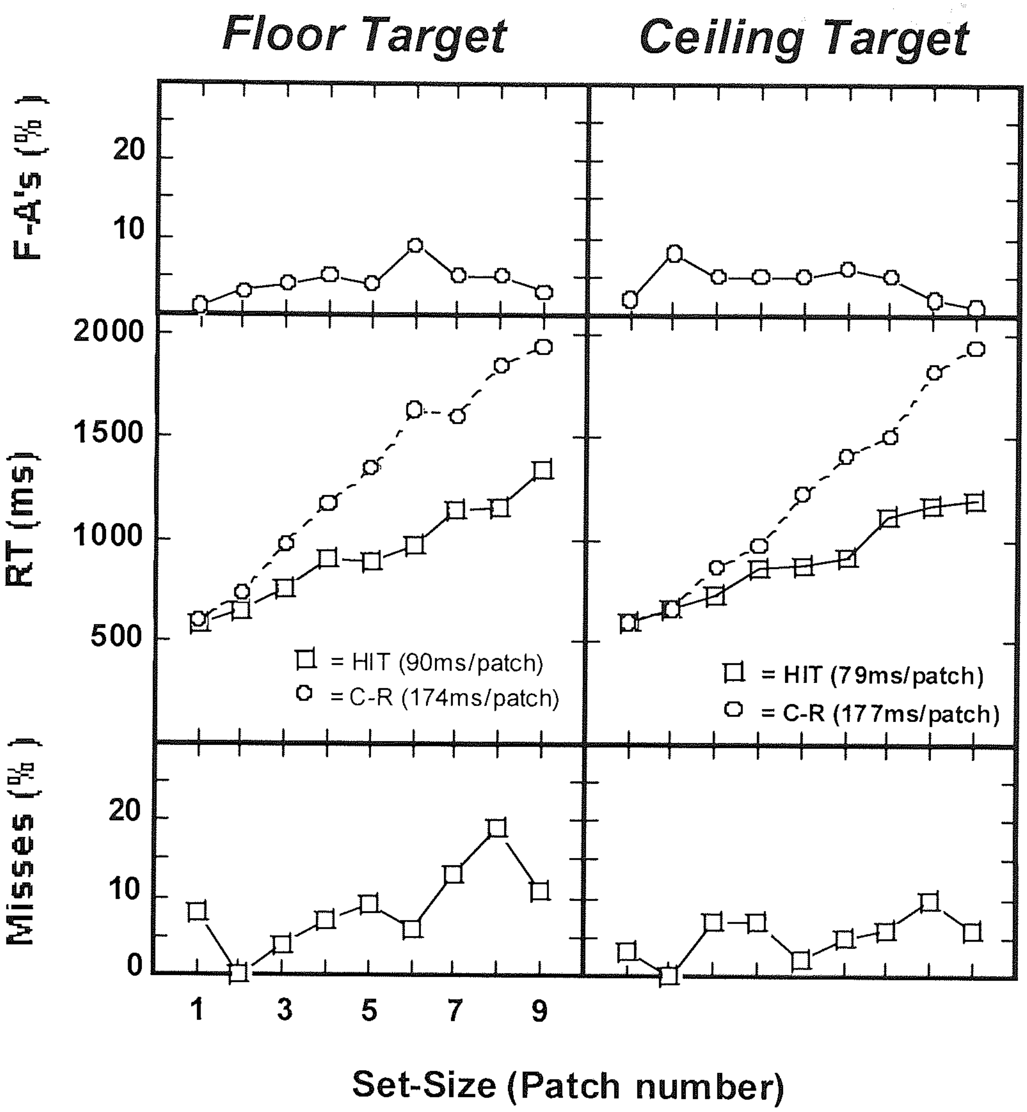


Fig.5.4. Fast condition (Homogeneous translation ±2.15°sec^{-1}; shear = ±1.08sec^{-1}) for subject PAA.

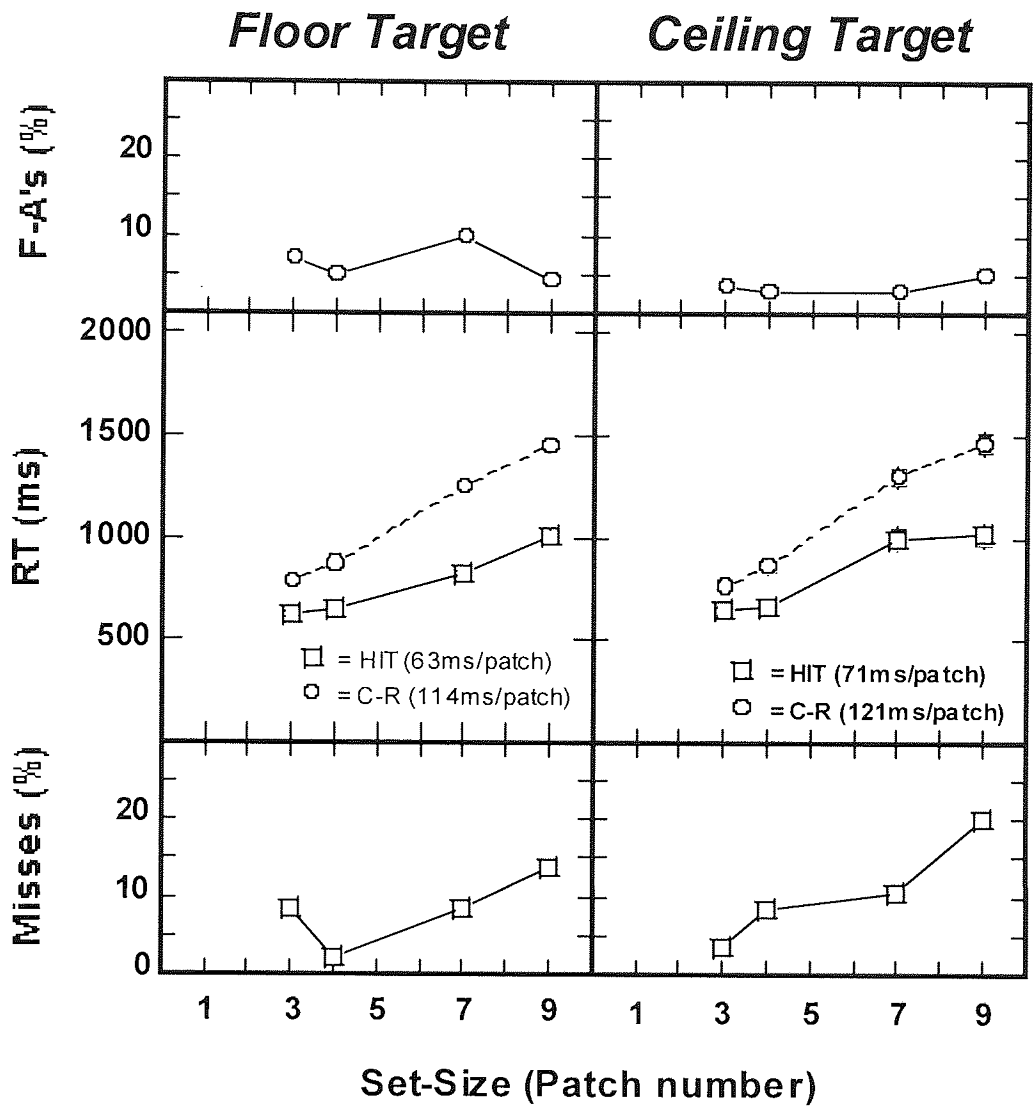


Fig.5.5. Slow Condition (Homogeneous translation = ±1.15°sec^{-1}; shear component = ±1.61sec^{-1}) for subject HRC.

5.5 Results

Mean RTs and error rates are plotted for each condition and observer (see fig.5.3-5.6). Within a single plot, both the *Hit* and *C-R* slopes are included. Table. 5.1. summarises the slope magnitudes from the best linear fits and the ratios between them.

	Floor-Like			**Ceiling-Like**		
	HIT	**C-R**	**Ratio**	**HIT**	**C-R**	**Ratio**
PAA (Slow Condition)	55 (45-65)	139 (130-148)	x2.5	62 (53-72)	135 (126-145)	x2.2
HRC (Slow Condition)	63 (52-74)	114 (103-124)	x1.8	72 (58-86)	121 (107-136)	x1.7
PAA (Fast Condition)	90 (77-102)	174 (162-186)	x1.9	79 (68-89)	177 (166-188)	x2.2

Table 5.1. Slope magnitudes for HITs and C-Rs are given for each observer and condition (ms/patch) along with 95% slope confidence intervals.

The more comprehensive data set from PAA (all set-size conditions from 1 to 9) confirmed that the set-size functions were, to a first approximation, well described by linear functions. The slope magnitudes obtained from both observers were large, but for PAA the slow condition produced more efficient search. Unlike the slopes of experiment 1, the *Hit* and *C-R* slope ratios clustered around 2:1 relationships, consistent with serial search. No clear evidence for the existence of a search asymmetry between floor-like and ceiling-like stimuli was observed, either in terms of search performance or general sensitivity.

F-A errors tended to be lower than the Miss errors. For subject PAA in the slow condition the F-A rates remained particularly low for all set-sizes (always less than 5%). This is a similar pattern to the one that was found in experiment 1. Another similarity is in the apparent trend for Miss rates to increase with set-size number. Although this trend is often not monotonic, prima facie analyses of the various conditions, both for the present and the previous experiment, suggest that the highest Miss rates tended to occur for the largest set-sizes.

5.6 Discussion

In the introduction it was speculated that the parallel search observed for disparity gradient (Holliday & Braddick, 1991) may have been based upon generic (cue-independent) 3-D descriptor mechanisms. Improving the saliency of the slants produced by motion might therefore have permitted parallel search. This was not the case for the two subjects tested. Neither showed evidence of parallel search when these subjectively more compelling slant representations were used. In fact, for observer PAA, search performance had been better in the *defl+trans* condition of experiment 1 than for either of the stimulus conditions of this experiment. There are three possible reasons for the failure of this experiment to produce parallel search like that found by Holliday & Braddick (1991) with stereoscopically defined slant. (1) Search can be based upon cue-independent representations of slant but the motion gradients that are generated in RDKs cannot evoke sufficiently compelling slant representations to permit parallel search. (2) While there are specialised mechanisms selective for disparity gradients, no such mechanisms exist for the processing of motion gradients. (3) Specialised mechanisms selective for both disparity and motion gradients exist but only the former are readily accessible to the observer. These three possibilities are further explored in the following subsections:

5.6.1 (1) The motion gradients generated by RDKs cannot evoke sufficiently compelling slant representations

It is possible that neither of the motion stimuli tested were capable of attaining the same levels of saliency as those achieved by Holliday & Braddick with stereoscopic stimuli. The most obvious reason simply being that motion is a weaker depth cue or that the impressions they produce are not salient enough when viewed peripherally. But were there any other confounding differences between the two experiments that may have been responsible for saliency differences? Four possibilities are discussed:

Display representation of the differential gradients

A major difference between the motion parallax and the stereoscopic displays was the way in which they represented differential gradients. The stereoscopic displays generated disparity gradients using pixel-wide outline drawings whereas the motion parallax displays generated differential motion gradients using RDKs.

a) While neither attempted to avoid conflicting pictorial cues to flatness the latter might have suffered from them to a greater extent. The RDKs consisted of dots that could be allocated anywhere within the stimulus aperture, leading typically to well spaced dot distributions. The absence of vertical dot-density gradients may have provided much stronger cues to flatness than did the absence of perspective in Holliday & Braddick's line parallelograms.

b) The burden of the correspondence problem might have been more acute for the RDK stimuli than the stereoscopic outlines. If RDSs - the closest stereoscopic analogy to RDK stimuli – are considered then because the features to be matched (in this case white dots) are randomly allocated, any informative cues to form would be absent. This was not so for the outline figures, where form provided cues for good continuation. If these cues can be utilised by the visual system to determine veridical representations of depth/surface-structure then the stimuli of Holliday & Braddick may have had an advantage in this respect (see Ramachandran & Cavanagh, 1985, for evidence that form cues suggesting 'good continuation' are capable of influencing the resolution of the correspondence problem).

Motion parallax stimuli are more ambiguous

Motion parallax flows will produce representations that are potentially more ambiguous than those produced by disparity stimuli. In constructing representations of slanted flat surfaces from motion, an additional assumption has to be made to the ones that are made for surface representations derived from disparity: the patch of flow arises from a rigid body. It could conceivably have arisen from a non-rigid deforming flow and indeed, for stimuli containing large amounts of shear, impressions of non-rigidity are the most frequent (presumably because the motion parallax cue conflicts excessively with the pictorial cues to flatness) (Freeman & Fowler, 2000, made a similar observation). Disparity stimuli do not have the additional temporal aspect and are therefore not subject to this kind of ambiguity. Anecdotal reports suggested that such perceptual ambiguity was not a factor, but it is possible that through some means it could have had a detrimental effect upon the saliencies of slant.

The direction of patch motions may have a distracting effect

Randomisation of the motion parallax direction was introduced to prevent inter-patch cues from being utilised. It was felt, especially during the initial practice session, that the randomisation would sometimes produce displays that were particularly difficult to search. Take for example a target-present display containing 9 patches. If 8 of the patches happened by chance to be moving in the same general direction, including the target, the most salient aspect of the display was often not the patch with the odd speed gradient but the boundaries between patches with different general directions of motion. The cues from the contrasting motion directions could elicit involuntary attentional shifts towards them and so interfere with the detection of targets. Effects in which events like these capture attention involuntarily, even when the subject is trying to actively to suppress them, have been well studied and are thought to be indicative of a distinct attention orientating system (Remington *et al,* 1992). With practice it seemed that these involuntary tendencies diminished, and such practice effects have been documented previously with other stimuli (Warner *et al,* 1990). There is still a possibility that to some extent, contrasts in motion directions were hindering the searches for odd motion gradient signs.

The stimulus patches were too small

The patch stimuli that were used in this experiment were made reasonable large. However, it is possible that in order for efficient search to be obtained with RDK stimuli, larger patches are needed. Even for Holliday & Braddick's (1991) outline parallelogram stimuli, it was found that at small stimulus sizes search became inefficient.

5.6.2 (2) Specialised motion gradient mechanisms do not exist

An alternative explanation for the steep search slopes that have been found in the present experiment (and those of the *defl+trans* condition of the last experiment) is that there are no dedicated mechanisms selective for motion gradients. Indeed, this is the interpretation that would be given by theories such as FIT (section 3.2.1). In this case serial search would have resulted because attention was needed to determine the signs of the speed gradients in each patch (i.e. a *within-item-conjunction* search).

5.6.3 (3) Specialised motion gradient mechanisms do exist but they are not readily accessible

If specialised mechanisms selective for speed gradients do exist in the brain, classical FIT would have predicted a parallel search for this experiment. However, in previous chapters it has been acknowledged that while specialised mechanisms may exist, their presence is not sufficient to permit parallel search. An additional requirement is that the outputs of these mechanisms are simultaneously accessible to the observer (see section 3.3.7). In the subsection below, one hypothesis is discussed regarding why some mechanism arrays might be more readily accessible than others.

Are mechanisms latter in visual processing less accessible to the observer?

The idea that the feature dimensions of early (V1) mechanisms are the exclusive predictors of parallel search has long been refuted (section 3.3.4), but it is possible that there is a more general link between search efficiency and the stage at which the search stimuli are represented in the brain. For example, 3-D descriptions derived from stereoscopic gradients might be represented at earlier stages than 3-D descriptions derived from motion gradients, and as a consequence the former are searched more efficiently. This of course is highly speculative but still a possibility.

Braddick (1993) has previously speculated that parallel search (leading to 'pop-out') requires a certain type of detector organization (so as to produce difference signals necessary for the detection of an odd item) "and that such organization might be characteristic only of earlier levels of visual processing". Another possibility is that processing in early areas, such as V1 and V2, does not require attention, but processing in latter areas does require attention. Support for this might be garnered from instances in which efficiently searched stimuli have been found to elicit neural correlates in early areas. One example is that of 'inferred' contours. Prior to the 1990s, the processing of amodally and modally completed contours was often considered to be at a high, attentionally demanding and cognitive level. Visual search results first by Gurnsey *et al* (1992) and then by Davis & Driver (1994) did not fit this view; they found searches to be parallel. Single cell physiology carried out by Grosof *et al* (1993) with modally completed contours and latter by Sugita (1999) with amodally completed contours, found V1 cells whose activity correlated well with the presence of these 'inferred' contours[3]. The consistency between the physiological and behavioural results and the fact that they contradicted introspective accounts gives some validity to the argument that stimulus properties processed early are not attentionally demanding and support parallel search.

Why would stimuli processed earlier in the visual system be processed preattentively but those processed at latter stages require attention? The same rationale can be used as before, when V1 feature maps were being implicated: stimuli processed earlier will tend to be

---

[3] The offset abutting lines/gratings used to induce the illusory contours in the Gurnsey *et al,* 1999, and Grosof *et al,* 1993, studies have however been the subject of some criticism (see Davis & Driver, 1994)

processed by a finer grained, more detailed analysis (Nakayama, 1990) largely independent of attention and as such necessarily parallel.

Presumably, if this were true, then to explain Holliday & Braddick's result, the mechanisms responsible for the processing of disparity gradients would be occurring earlier in visual processing. A review of disparity processing was provided in section 1.4, within which it is clear that there is little evidence to suggest differential gradients defined by disparity are processed particularly early in the processing stream. There seems little reason therefore to believe that the parallel search observed by Holliday & Braddick (1991) is attributable to an early "preattentive" processing of disparity gradient.

Chapter Six: Experiment 3 - Grouping of 1-D speed gradients

6.1 Chapter Introduction

Experiments 1 and 2 of this thesis found no evidence of parallel search for stimulus items possessing odd motion gradients (slant signs), despite parallel search having been found for stimuli possessing odd disparity gradients (Holliday & Braddick, 1991). In the present chapter the following question is asked: if representations of slant (derived from motion parallax) can be grouped, do their individual representations become more accessible and so, for certain stimulus types, permit parallel search? One reason for suspecting that this might occur comes from studies suggesting that when display items can be grouped on the basis of some common attribute, they can be attended to more easily, and therefore searched more efficiently than displays in which stimulus items cannot group (He & Nakayama, 1995; Morita & Kumada, 2003). In the subsections 6.1.1 & 6.1.2, literature is reviewed that gives support to the notion that increasing the grouping between the stimuli of the display may promote parallel search.

6.1.1 Grouping by common surface

A number of studies have investigated the possibility that observers can restrict visual search to a subset of the display items (Wolfe *et* al, 1990; Egeth *et al*, 1984; Friedman-Hill & Wolfe, 1995; Pashler, 1987; He & Nakayama, 1995). For example, Egeth *et al* (1984) found that searches could be restricted to items possessing the target colour. Items possessing other colours could be successfully ignored and appeared not to demand consideration. In a series of studies by He and co-workers (He & Nakayama, 1995; McCarley & He, 2000; 2001) it was found that the attribute that an observer uses to restrict search to a subset of display items does not have be a classic "feature" (e.g. colour) it can be a higher-level surface property.

Evidence to show that search times are lower when the search stimuli form a common surface

In He et al's studies, the subset containing the target was defined either in terms of common average stereoscopic depth or in terms of being constituent parts of common planar surfaces. In the two main experiments of He & Nakayama (1995) the displays consisted of a matrix of 36 stimulus items (3-rows x 4-columns x 3-disparities [crossed, zero, uncrossed]; see fig. 6.1 for illustrations). Subjects were always required to search a subset of the matrix item for an odd colour. The first experiment required them to search the 12 items that were centred at zero disparity while ignoring the items at crossed and uncrossed disparities. Three conditions were examined; one in which each of the individual items were vertically orientated (see upper portion of fig. 6.1), one in which they had a slight positive slant (top slanting away) and one in which they had a slight negative slant (top slanting inwards). Only in the first condition were all the to-be-searched items coplanar and consistent with a common implicit surface. In the other two conditions they shared the same average stereoscopic depths but were not consistent with implicit surfaces. He & Nakayama found that the RTs for the coplanar surface condition were considerably shorter than those of the other two conditions.

A second experiment confirmed that the coplanarity in the first condition had been the key factor in reducing RTs. This experiment employed a condition in which coplanar surfaces were constructed from a range of disparities (see lower portion of fig. 6.1). Subjects were required to search the 12 items of the middle strata. As in the first experiment, there were two conditions in which the items were not consistent with coplanar surfaces (i.e. the items had either positive of negative oblique slants). RTs were again much shorter in the coplanar surface condition.

In section 3.3.3 of this thesis, a previous study by Nakayama & Silverman (1986) was discussed, in which they reported conjunction searches for [colour x disparity] and [motion x disparity] carried out in parallel. Their terminology was somewhat misleading however, in that the search task that they employed was unable to determine whether a true conjunction search is carried out in parallel. The reason for this is that they used the same number of depth planes (two) irrespective of the set-size. It is conceivable that subjects were able to carry out the task by selectively attending to one depth plane at a time. Indeed, the use of such a strategy was alluded to in the paper with the statement: "The observer had the distinct impression that each plane could be searched almost effortlessly, in turn". It is concluded therefore, that their experiments did not prove that colour and motion could be searched across numerous depth planes in parallel. More recent evidence supports this. Research by Moore *et al* (2001) has shown searches for particular binocular depth relations are actually very inefficient. In light of the more recent studies by He *et al* it seems more likely that the data of Nakayama & Silverman (1986) can be better explained in terms of a subset search based upon common surface rather than as conjunction searches. Research indicating that depth pre-cues are capable of decreasing discrimination times supports this view (Marrana & Moore, 2000).

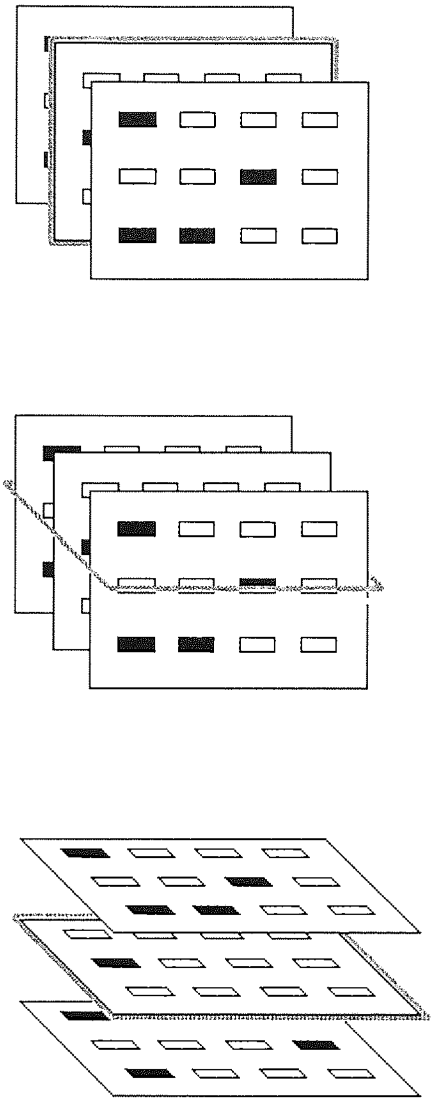


Fig. 6.1. Showing three of the displays used in the experiments of He *et al* (He & Nakayama, 1995; McCarley & He, 2000; 2001). 36 stimuli were allocated to elements of a three-dimensional (4x3x3) matrix. The positioning of the stimuli in terms of depth and stimulus slants was achieved stereoscopically. The opaque planar surfaces have been included for illustrative purposes only; they indicate the global surfaces implied by the stimulus items. In the actual displays, all stimulus items from all three depths were visible to the observer. Greyed outlines enclose the subset of items the subject was required to search. In the **upper** display it is intended that search be carried out on the 12 items of the middle depth plane. In the **middle** and **lower** displays the search is to be carried out on the 12 items of the middle strata (also note that both are arranged in ground-like configurations). Items possessing the target colour are denoted here by the blacked out parallelograms, but in reality they were coloured green or red. In all three shown here therefore, a target is present (located in the leftmost column of the mid disparity on the second row).

Why do search times decrease when the stimulus items appear to belong to a common surface? One possibility is that the larger RTs of the non-coplanar conditions were due to an increased difficulty in restricting search to the designated subsets. If this had been true the task would be akin to a conjunction search. In other words, the 3-D position of a target-coloured item would need to have been confirmed as a 'target location' before the item was accepted as a target. He & Nakayama (1995) found no evidence in favour of this explanation. Although they did not vary set-size directly they did vary the number of target coloured items that appeared in the to-be-ignored depth planes. Increasing the number of these target-coloured distractors did not seem to cause any systematic increases in RTs in any of the conditions studied. Thus it is suggested that the longer times taken to respond to the non-coplanar displays were due to the initial difficulty of grouping the items into subsets and not because of degraded search performances. Subsequent findings by McCarley & He (2000; 2001) have also gone on to show that odd-item detection is quicker when the items are to be grouped into coplanar floor-like surfaces as opposed to ceiling-like surfaces. This suggests that the slant sign to which the search stimuli are to be grouped to does affect the ease of the grouping.

An obligatory spread of attention over implicit surfaces

When attention is directed to an implicit surface constructed from a subset of display stimuli, attentional deployment over that surface seems to some extent to be mandatory. He & Nakayama (1995) demonstrated this in their third experiment. The observers' task was to detect an odd-coloured item, which could have occurred in one of two rows (3 stimulus items were in each row). One row appeared in a crossed disparity and the other in an uncrossed disparity. A pre-cue indicated which row the target was most likely to appear in (correct 80% of the time). They tested two types of configuration; one in which the items of both rows were implied as belonging to the same surface, or ones in which the items of the two rows were perceived as forming distinct surfaces. The disparity between the two rows was manipulated, and, as before, RTs were measured. The critical finding comes from data for the invalid cueing (in which the incorrect row was indicated). Increasing the disparities between the rows was found to have little effect upon the RTs for configurations in which the rows were implied as belonging to the same surface. However, in the alternative condition, in which the two rows were perceived as distinct surfaces, RTs could be seen to increase, sometimes precipitously, as the disparity was increased. This could imply that shifting attention between surfaces takes more time. An alternative explanation is that surface selection leads to attention being spread over it in a compulsory manner. Results from McCarley & He (2000) favour this latter idea. They found that for certain conditions the time taken to search for an odd colour within a subset of display items increased as the number of target-coloured distractors increased. Previous experiments had failed to report such systematic 'interference' effects (He & Nakayama, 1995) however, none of these had used stimulus items that formed implicit surfaces involving both the target and to-be-ignored items. In the two conditions of interest from McCarley & He's study the items were vertically orientated and arranged in the usual matrix configuration (see the middle display of fig. 6.1). As a result, the arrangements were consistent with three implicit frontoparallel surfaces. However, the observer's task was to search the middle horizontal strata and not one of the frontoparallel surfaces. Although not explicitly mentioned in their paper, it seems likely that the cause of the interference effects was due to an inability to effectively ignore the to-be-ignored items of these implicit frontoparallel surfaces.

Selectively allocating attention to a subset of items

Some concern might be expressed at this point, relating to the concept of attention being directed at subsets of items and that search is performed only on these subsets. The result of Nakayama & Silverman (1986) for example, showed that search (for on odd coloured items or odd directions of motion) amongst a subset of items, defined by their common depth, can be highly efficient. Under standard FIT however, efficient search is considered reflective of preattentive processing, not attentive processing. This apparent discrepancy can be resolved if it is accepted that, for the most part, top-down (selective) attention acts only to filter out preattentive information that is not relevant to the task - even when the ignored stimuli can be processed in their entirety by preattentive processing. It remains to be confirmed whether there are instances in which grouping can proceed in the absence of attention. Behavioural data from humans have suggested this might happen (Moore & Egeth, 1997), as has a neuropsychological study involving neglect patients (Mattingley *et al,* 1997); surface completion was found even when processing in the neglected region was required[1].

Evidence to show that search slopes are shallower when the search stimuli form a common surface

The three studies of He and colleagues that have been reviewed so far did not manipulate set-size and therefore could not confirm that the items forming implicit surfaces were being searched in parallel. They did manipulate the number of to-be-ignored stimuli possessing the target colour but this manipulation could only measure interference effects, of which they found none.

---

[1] The issue of whether neglect arises from an attentional deficit is, however, contentious.

Morita & Kumada (2003) have also investigated search among items adhering to implicit surfaces, but they also manipulated set-size. Pictorial depth cues, rather than stereoscopic cues were used. The researchers used a perspective cue and, because the stimulus items were figural (top-viewed cubes), there was a concomitant depth cue from their compression/size. The primary interest of the research was set-size differences between searches in which the relevant cube faces were coplanar and those in which they were not. From the data obtained it was concluded that performance is better in the former case, where shallow search slopes were observed. An additional issue that was addressed was whether an asymmetry exists for the slant sign of the search surface (prompted by the RT asymmetries observed by McCarley & He, 2000; 2001). A significant asymmetry was noted, in which implicit floor-like organisations were found to be searched more efficiently than ceiling-like organisations.

6.1.2 Distractor grouping

In the studies that have been reviewed so far it has been shown that search becomes faster or more efficient when the to-be-searched items can be readily grouped. For example, in the case of He & Nakayama's (1995) study this grouping was by common surface, whereas in the study of Egeth *et al* (1984) it was by common colour. Here, I further suggest that the advantages of such grouping probably continue to be observed if it is just the distractors that group. The idea that distractor grouping can improve search efficiency is well established. For example, Duncan & Humphreys (1989) have shown that the more homogenous the distractors, the more efficient search tends to be. This is consistent with the idea that distractor grouping by similarity is occurring. Donnelly *et al* (1991) also showed that inefficient searches could be made efficient when the distractor stimuli are designed so that they can be grouped by common shape. This is probably also true for grouping by common motion. By using local direction difference threshold measures instead of set-size effect, Bravo (1998) has shown that the type of common motion formed from a set of distinct motions has an affect on target detection. Bravo found target detection to be better when the common global motion was consistent with the generic 2-D complex motion group of 'spirals' rather than 'deformations'[2].

In this chapter, three distractor groupings, each derived from motion parallax stimuli, were conceived of. (1) Distractors group on a row-wise basis. (2) Distractors group on the basis

---

[2] The experimental design of the study did ensure that the difference could not be attributed to local motion differences.

of a single common (approximately) corrugated surface. (3) Distractors group on the basis of a single common planar surface.

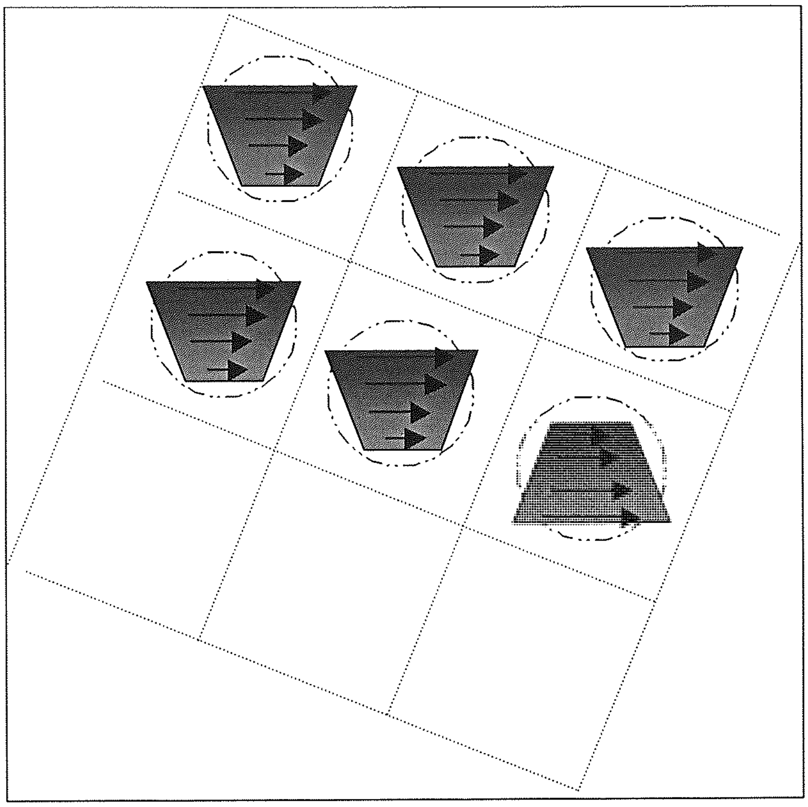
6.2 Experiment 3a: Distractors group on a row-wise basis

In this experiment the distractors were arranged so as to promote distractor grouping on a row-wise basis. The stimulus design was similar to that of experiment 2. Each of the display stimuli were constructed from an RDK patch and the dots within the patches would adhere to motion transforms constructed from a homogeneous translation summed with a 1-D vertical speed gradient. This resulted in patch-wide motion parallax and distinct impressions of slant from motion. However, unlike the experiment 2, where measures were taken to ensure the independence of the stimulus patches, in this experiment patches were not spatially jittered and in one condition the notional patch allocation grid was not rotated through 22.5°. In this condition (0° rotation) the distractor patches on each row of the allocation grid were therefore free to group by virtue of their common slant (see the lower display of fig. 6.2). In the former (22.5° rotation) condition this would have been prevented. Given the points made in section 6.1, if row-wise distractor grouping does occur then superior search performance would be expected from the 0° condition.

6.2.1 Stimuli and procedure

Stimulus, task and viewing conditions were almost identical to those applied in the search experiment 2 with the following exceptions: The RDK patches were not spatially jittered and trials were carried out only for the set-sizes of 3, 4, 6, and 9. Trials for the two different targets (i.e. floor-like and ceiling-like slants) were intermixed. Although intermixing trials is known to affect RTs in general, by increasing them by a fixed amount, any effect upon the search slopes is unlikely (Bravo & Nakayama, 1992). Motions in the RDK patches were again constructed from horizontal homogeneous translations (with speeds of $\pm 1.15°\text{sec}^{-1}$) summed with shear components of $\pm 1.16\text{sec}^{-1}$ (vertical speed gradient). These values match those used in the ‘slow’ condition of experiment 2 (section 5.3). Data were collected for two conditions (see fig 6.2). In the 22.5° condition the notional allocation grid was rotated by +22.5°. In the 0° condition the allocation grid was not rotated at all. In contrast to the experiment 2 no spatial jitter was added to the patches. 20 sessions were carried out in a pseudorandom order (10 for each condition). Each session contained 10 trials per set-size, for both target-present and target-absent conditions. A total of 80 experimental trials were presented in random order (per session). An additional 15 practice trials were carried out at the beginning of each of these sessions. Data were collected from a single observer (PAA).

22.5° condition



0° condition

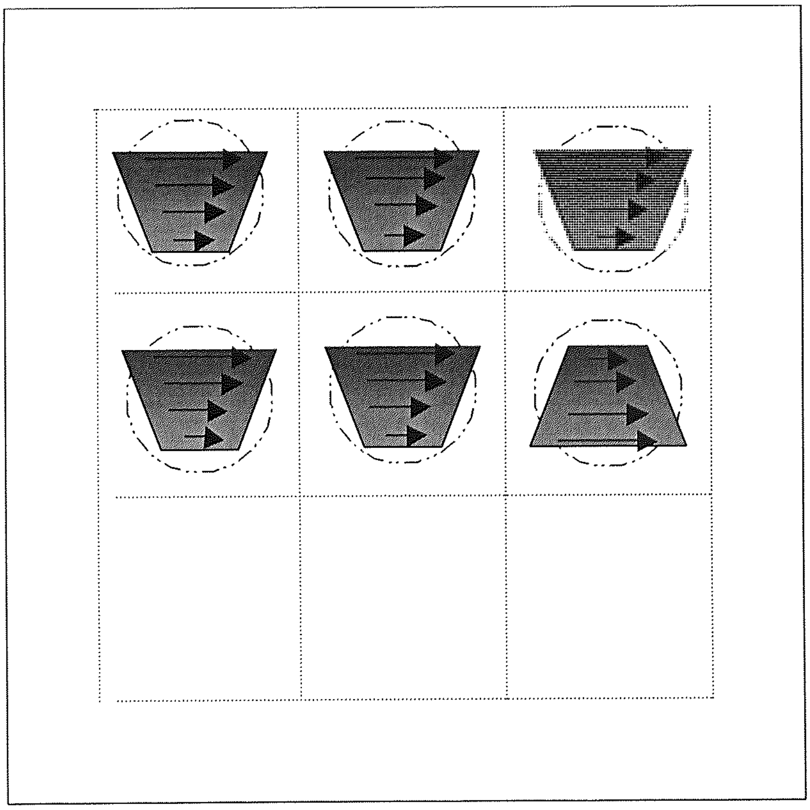


Fig. 6.2. Schematics of the two display conditions are shown: **22.5° condition** = 22.5° display grid rotation; **0° condition** = 0° display grid rotation. The dashed lines have been included to show aperture delineation and the notional grid. Black arrows have been used to illustrate the speed gradients and their direction. Shaded trapeziums illustrate the perceived slant that is elicited from the RDK patches. Both displays show a trial in which the set-size is 6 and a target is present (middle row, right column).

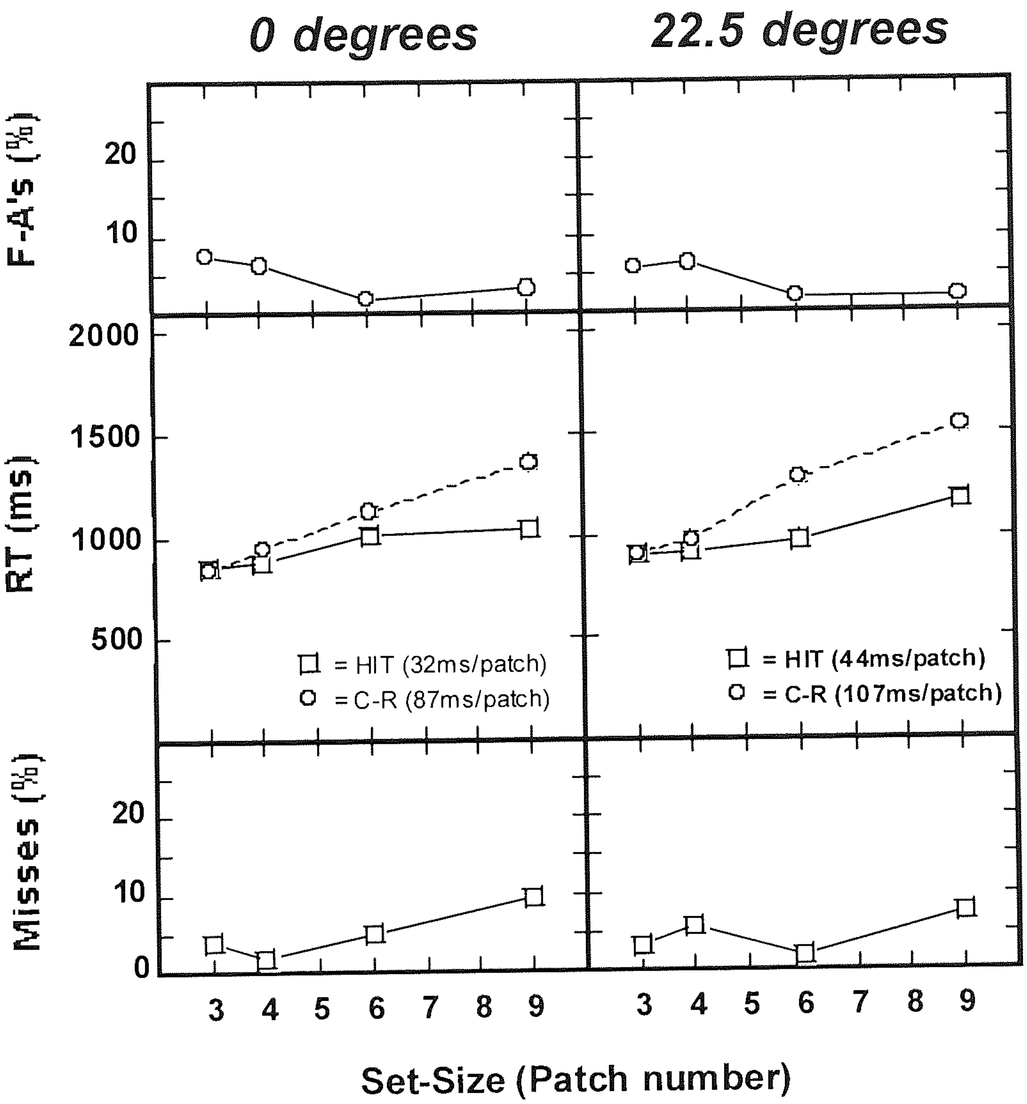


Fig. 6.3. Showing data for the 0° and 22.5° conditions from observer PAA. The left-hand column shows data for searches in which there was no rotating of the display grid. The right-hand column shows data for searches in which there was a 22.5° rotating of the display grid. Corresponding error rates are plotted in terms of error type. Misses are shown on the upper portion of the graph and F-As are shown on the lower portion. Linear-regressions are plotted for the RT data for Hits & C-Rs. Slope estimates are inset in plots (bracketed values).

6.2.2 Results

The mean RTs for correct responses are plotted in fig. 6.3, along with corresponding error rates. Both Miss and F-A rates remained reasonably low (below 10%). The 22.5° condition yielded set-size slopes of 44ms/patch (Hit) and 107ms/patch (C-R). Comparing these with the set-size effects from the experiment of Chapter 5 (55 & 62ms/patch for Hit slopes; 139 & 135ms/patch for C-R slopes) seems to confirm the subjective impressions that jitter removal leads to better search performance. The 0° condition yielded search slopes of 32ms/patch (Hit) and 87ms/patch (C-R). The removal of the 22.5° display rotation seems to have led to further improvements in search performance.

6.2.3 Discussion

The superior search found with the 0° condition is consistent with the hypothesis that the distractor patches group together on a row-by-row basis and, for reasons discussed in sections 6.1.1-6.1.2, this grouping leads to an enhanced search performance. This finding opens up the possibility that the failure to find parallel search in the experiment of chapter 5 was due to the fact that the global configuration of the display had prevented observers from accessing surface descriptions for the individual patches in any effective manner. In other words, the grouping of individual display stimuli might be a prerequisite for enabling a simultaneous accessing of their representations. Therefore, the more incongruous the stimuli are with global organising principles, the more inefficient the access to their representations.

6.3 Experiment 3b: Distractors group on the basis of a single common corrugated surface

In the previous chapter it was speculated that inefficient search performance might be attributed to the fact that the parallax stimuli in those experiments did not evoke sufficiently strong impressions of slant. Under the working hypothesis this would diminish the accessibility of the representations at the decision stage. Experiment 3a (section. 6.2) showed that search performance could be increased if the display configurations are constructed so as to favour a grouping of the stimulus patches. This is consistent with the work of He and colleagues and Morita & Kumada (2003) who showed that grouping by common 3-dimensional surface can have particularly profound effects on search performance. It is hypothesised here, that in these cases, the individual search items are processed preattentively, but to ensure the decision stage has full access to the individual stimulus representations, they must conform to a single global representation (e.g. a surface), to which the observer can apprehend. If grouping does improve the access to individual stimulus representations then it is possible that this would permit toleration of poor slant saliencies.

The hypothesis that grouping by common 3-D surface can enhance search performance, was tested using motion parallax stimuli. The same discrete RDK patches were used as in experiment 3a, but here they were arranged into one of two types of global configuration. In the first type, the distractors could simulate identical slants and as usual the target would simulate a slant of the same magnitude but opposite sign (as in the 0° condition of experiment 3a) (see fig. 6.4). Displays adhering to these configurations were used in the first 'non-contiguous' condition. In the second type, the slant impressions from the distractors were made so as to be consistent with a single surface, which to a first approximation was corrugated (see fig. 6.5). The target would simulate a sign of slant incongruous with the implicit (corrugated) distractor surface. Displays adhering these configurations were used in the second, 'contiguous' condition. If the working hypothesis were correct then search performance should be best in the 'contiguous' condition.

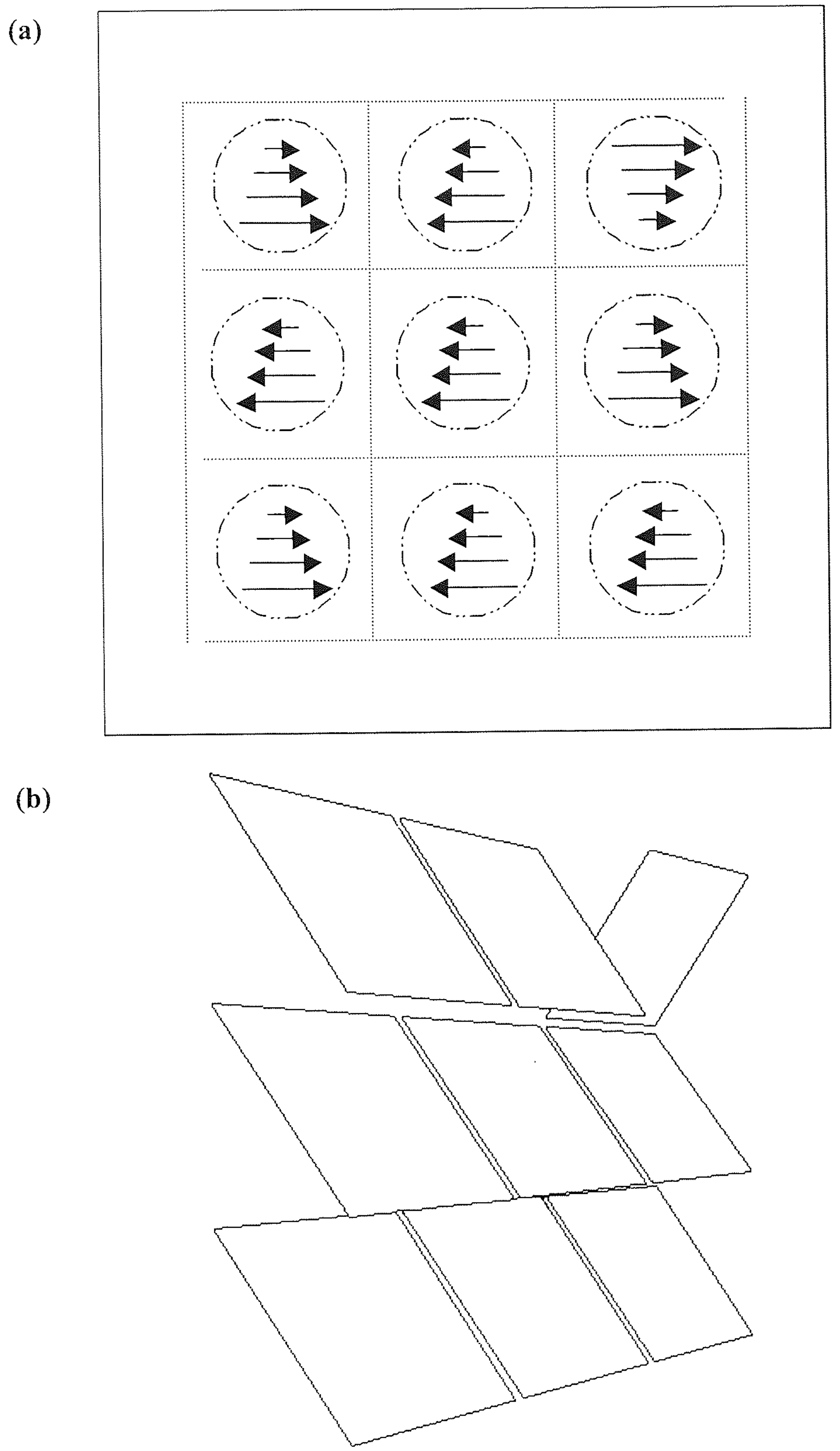
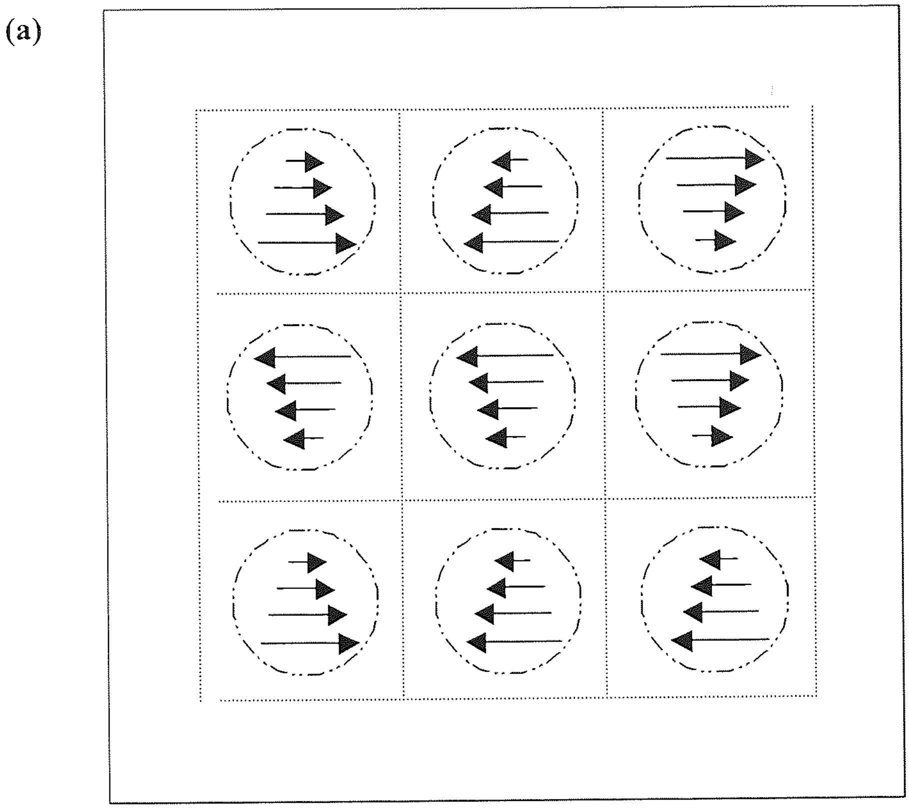


Fig.6.4. (a) Schematic of an example display from **non-contiguous** condition (the same display configuration as 0° condition of the experiment 2). It has a set-size of 9 and a target patch is located in the top-right grid element. The dashed lines have been included only to illustrate the notional grid used for patch allocation. Patch demarcation is also shown. (b) A coarse representation of the impressions of slant derived from the display. The parallelograms represent the surface slants that were implied by the speed gradients of the stimuli and have been used so as to enable potential groupings by common surface to be more easily apparent. In this particular example, it is clear that a grouping might occur on a row-wise basis; the top-right (target) item would not group with the other two items on its row, however.



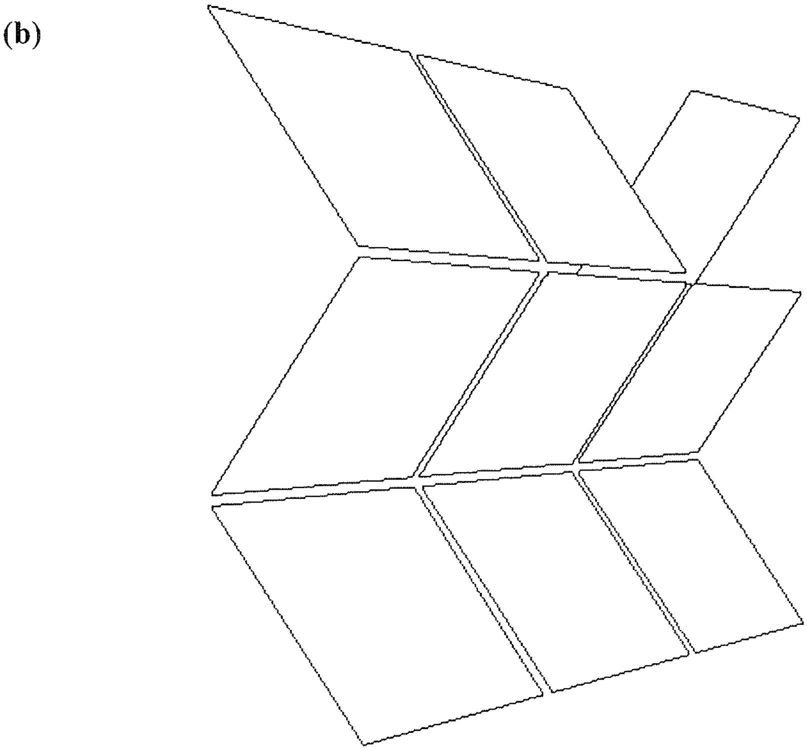


Fig. 6.5. (a) Schematic of an example display from a **contiguous** condition. As in fig. 6.4, the display has a set-size of 9 and an odd patch is located in the top-right grid element. (b) A coarse representation of the impressions of slant derived from the display. The parallelograms represent the surface slants that were implied by the speed gradients of the stimuli. In this example, the grouping might occur not just on a row-by-row basis, but also across rows, consistent with a single corrugated surface. As before however, the top-right (target) item would not be able group by common surface with the surrounding stimulus items.

6.3.1 Stimuli and Procedure

In the first, non-contiguous condition, patches were arranged obeying the same rules as those of the 0° condition of the previous experiment (see fig. 6.4 for an illustration of an example display). That is to say, the allocation grid was not rotated and there was no spatial jitter added to the patches. In the second, contiguous condition, patches were again arranged in a non-rotated allocation grid and no spatial jitter was added to the patches. In contrast to the non-contiguous condition though, the distractor patches were arranged so that their slant impressions were consistent with coarse approximations of a common corrugated surface (see fig. 6.5). Impressions of continuity between distractor patches seemed apparent on casual inspection of these displays. Other than these differences, the stimuli were identical to those used in the previous experiment, including the shear and translation magnitudes of the RDK patches (i.e. $\pm 1.61 \mathrm{sec}^{-1}$, $\pm 1.15^{\circ}\mathrm{sec}^{-1}$ respectively). One minor exception was with the set-sizes used. Only set-sizes that would form rectangular blocks of patches were employed: 2, 4, 6 and 9. Data for the two conditions were collected concurrently in pseudorandom order within 20 sessions (10 sessions per condition). Two practice sessions were carried out to familiarise the observer with the task and stimuli of the novel condition prior to recorded sessions. Each session contained 10 trials per set-size, for both target-present and target-absent conditions. A total of 80 experimental trials were presented in random order per session (15 practice trials were carried out at the beginning of each session). Data was collected from a single observer (PAA).

6.3.2 Results

Mean RTs for correct responses are plotted in fig. 6.6 with corresponding error rates. F-A rates remained characteristically low. Miss rates tended to be slightly higher than usual for the corrugated condition, with three out of the four set-sizes exceeding 10%. The non-contiguous condition yielded set-size slopes of 29ms/patch (Hit) and 83ms/patch (C-R). The C-R slope is therefore x2.7 that of the Hit slope. These values are comparable to the results obtained for the same condition in experiment 2 (32ms/patch (Hit) and 87ms/patch (C-R)). The contiguous condition yielded higher set-sizes of 70ms/patch (Hit) and 109ms/patch (C-R). The C-R slope was therefore x1.6 that of the Hit slope.

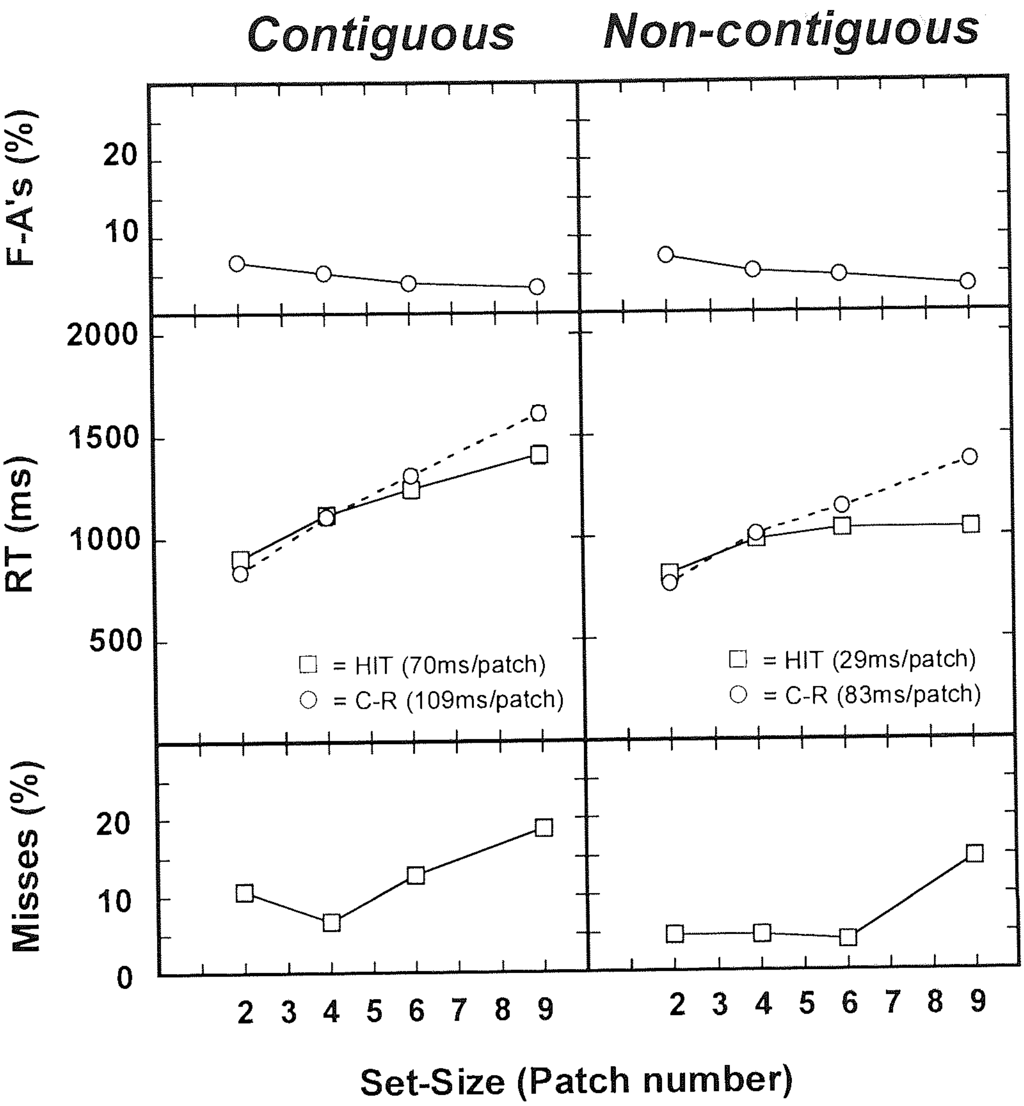


Fig. 6.6. Data for the contiguous and non-contiguous conditions from observer PAA. The left-hand column shows data for searches in which the distractor patches were consistent with a contiguous corrugated surface. The right-hand column shows data for searches in which the distractor patches possessed the same magnitudes of translation and shear. Corresponding error rates are plotted in terms of error type (Misses in the lower plots and F-As in the upper plots). Linear-regressions were carried out on the reaction time data for Hits & C-Rs. The slope estimates are inset in plots (bracketed values).

6.3.3 Discussion

This experiment yielded no evidence in favour of the working hypothesis: search was highly inefficient in both the contiguous and non-contiguous conditions. While one possibility is that the working hypothesis is incorrect - at least for when the common surfaces are defined by motion cues - another possibility is that the grouping of the patch motions into common corrugated surface impressions was simply not salient enough. This could have arisen because the continuity between distractor slants was not sufficiently smooth between rows or because the complexity of the surface was too great to be described by perspective motion alone. This second point seems unlikely however, as the detection of sinusoidal modulations, to which the stimuli approximates, is known to be robust over a wide range of frequencies and amplitudes (Andersen & Atchley, 1997).

For the single observer that participated, search performance in the contiguous condition had a tendency to be slightly worse than the non-contiguous condition. If this difference is genuine, there are two possible explanations. (1) Grouping for patches that exhibit common signs of slant is greater than that for patches adhering to a common surface (i.e. a grouping by similarity was greater than grouping by good surface continuation). (2) In order to determine the presence or absence of a target patch in the contiguous condition, the observer had the additional burden of comprehending the relative locations of the distractor patches, in addition to their slant signs.

6.4 Experiment 3c: Distractor grouping on the basis of a single common planar surface

If the corrugated configuration of experiment 3b did not elicit sufficient impressions of good continuity between the distractor stimuli, the hypothesis that grouping by common surface promotes the effective monitoring of the constituent stimuli of the surface would not have been disproved. In one last attempt to identify efficient parallel search with 1-D motion gradients, a display configuration promoting better impressions of good continuity was considered, namely, global planar surfaces. He & Nakayama (1995), McCarley & He (2000) and Morita & Kumada (2001) seem to have achieved good continuity with display items that adhered to these types of configuration. He *et al* used stereoscopic cues to create impressions of slant in global planar surfaces while Morita & Kumada used pictorical depth cues. Morita & Kumada also observed superior search performance when their stimulus items adhered to coplanar configurations. Here the slant cue would be motion parallax and the task envisaged would be similar to the previous: detection of the presence or absence of a patch containing an incongruous motion gradient (slant sign).

The issue of grouping is, however, complicated when discrete patches of parallax motion are being considered for coplanar configurations, for the following reason. Slant estimates are dependent upon estimates of translation as well as shear (*Def*). This has been pointed out previously in equation 1.10 and is restated below for convenience[3]:

$$S_{\psi} = m_{\psi} \arctan\left(t_{\psi} \frac{Def}{a_T} \right) \tag{6.1}$$

For a particular surface translating relative to an observer, the estimate of translation will be dependent upon the size and location of the sampled region. This is most clearly illustrated by envisaging a large planar surface (slanted so that the top is furthest away from the observer) translating horizontally relative to the viewer's line of sight. If the motion of this surface is viewed through a small aperture, the vertical positioning of the aperture would be critical to the average amount of translation observed (a_T) but not to the amount of shear (*Def*) observed: the average translation would increase with the lowering of the aperture. Because translation estimates are dependent upon the area being sampled, given equation 6.1 so, therefore, should the estimate of slant, S_{ψ} (see fig. 6.7).

This prompts the question as to how the visual system obtains estimates of translation when a single simulated surface is viewed through multiple apertures placed at different positions. Does it (a) carry out a 'local analysis' considering only the average translation within each aperture, or does it (b) carry out a 'global analysis' considering the average display-wide (pan-aperture) translation? If the former were true then the slant impressions derived for each aperture would be expected to vary depending upon their vertical position in the display. If the latter were true then the slant from each aperture would be correctly 'captured' within a single coplanar representation of slant. This issue is of direct relevance to the visual search experiment that has been proposed for the present experiment. Here the intention is to have distractors group on the basis of coplanar slant (as illustrated in the 'Global Analysis' situation of fig. 6.7). This will only happen if the visual system obtains pan-patch translation estimates. If, however, estimates are restricted to individual patches, the slant angle would end up varying systematically through the rows of the display, and the coplanarity achieved in the displays of He *et al* (1995; 2000) and Morita & Kumada (2001) would be deemed unobtainable with a motion parallax cue (as illustrated in the 'Local Analysis' situation of fig. 6.7). In the 'corrugated' condition of experiment 3b the

---

[3] The convention of expressing shear components in terms of average local deformation magnitude is used here to be consistent with the mathematical analyses that were discussed in chapter 1 (e.g. Koenderink, 1986; Freeman *et al*, 1996a). In no way, however, is it intended to allude to the nature of the actual mechanisms involved in the biological extraction of slant.

issue of the spatial extent of translation analysis was not important. The estimates would have been the same irrespective of whether they had been derived in a patch-wide or display-wide (pan-patch) manner.

To address the empirical issue of whether the spatial extent of translation estimation is local or global, a preliminary slant matching experiment was carried out. This involved the matching of pictorial representations of slant to those generated from RDKs with 1-D speed gradients. The conditions of most interest involved configurations of just two RDK patches. In a typical search task, displays would contain larger patch numbers, but here the intention was to confirm whether or not global estimates of translation are derived for when the lowest possible set-size is used (for when a target is present in a display with a set-size of 3). If it does occur for this, the most parsimonious condition, then it seems reasonable to suppose that the same would be true for configurations involving even greater patch numbers, so long as they remain clustered at the same density irrespective of the set-size (as has been the case for all the previous search experiments of this thesis, with the exception of experiment 1). If a global analysis does not occur then there would be little reason to proceed with a search experiment involving these stimuli, as the good continuity that we seek would not be present. There is also the possibility that the saliency of the target might become dependent upon its vertical positioning within the display.

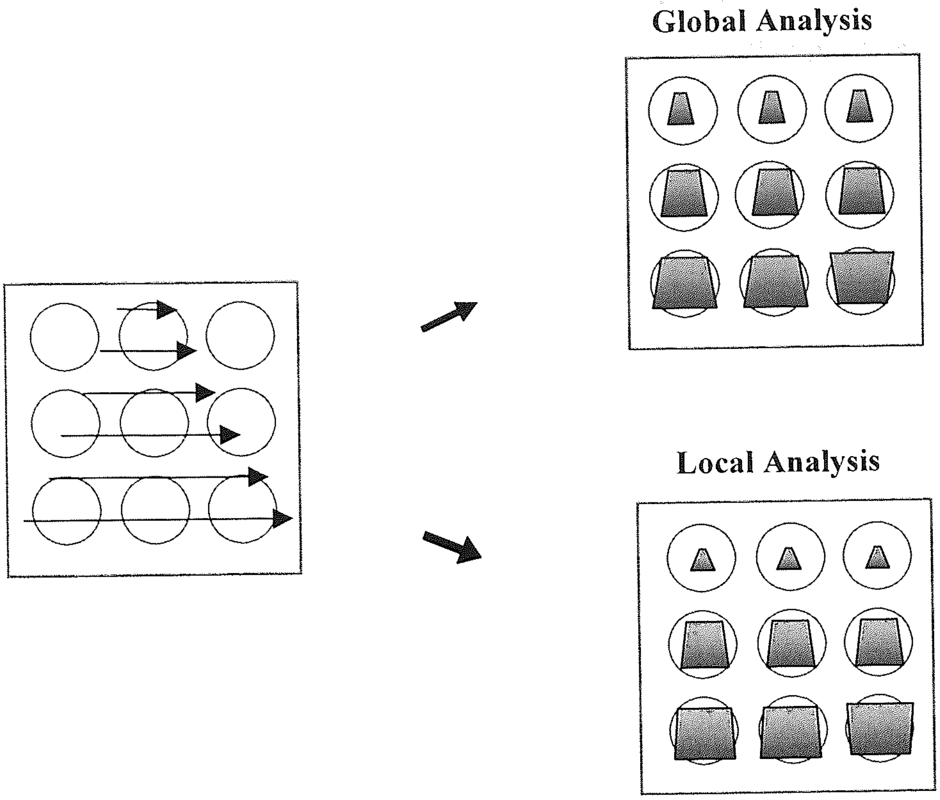


Fig. 6.7. Showing the possible impressions of slant elicited for a typical search display in which the dot motions of each distractor patch adheres to a global motion transform containing a positive uniform speed gradient. This example display has a set-size of 9 with the bottom-right patch designated a target. In the upper-right display a 'global analysis' of translation is assumed. Consequently, each distractor patch adheres to a common planar surface. In the lower-right display a 'local analysis' of translation is assumed. Consequently, the distractor patches on each row appear to have slightly different slants, becoming progressively shallower as the rows are ascended.

6.4.1 Stimulus and procedure: Slant-matching experiment

The displays in this preliminary slant matching experiment always consisted of a main aperture, 9.0° in diameter, two inset sub-apertures of 3.0° diameter, and a central fixation cross located at the origin of the main aperture and set to eye level (see fig. 6.8). Boundaries for the three apertures were permanently delineated by red outlines. The horizontal positioning of both sub-apertures was centred in the main aperture, having a vertical separation of 0.30°, 3.02° or 5.72°. This separation was again centred on the origin of the main aperture, as is made clear in fig. 6.8. Three configurations were tested: whole, double and single. In the whole condition, dots were randomly allocated within the area of the main aperture, including the sub-aperture areas. In the double aperture condition, all dot positions, other than those within the two sub-apertures, were hidden. Similarly, in the single aperture condition all dot positions, other than those occurring in just one of the two sub-apertures, were hidden. Each dot underwent horizontal translations, with their motion sign and magnitudes described in terms of a global transformation centred at the origin of the main aperture (fig. 6.9 depicts a global motion transform with a positive vertical speed gradient).

The subject's task was to make slant matches based exclusively on a slant impression ascertained from a 'target' sub-aperture, irrespective of whether dots were appearing elsewhere in the display. As with the previous experiments, viewing was monocular (using the dominant eye). At the beginning of each trial the sub-aperture to be examined was indicated by a pre-cue. This entailed its outline briefly changing colour to green. A 400ms period followed before the onset of the RDK. The RDK movies were constructed from limited lifetime dots whose births and deaths were staggered evenly throughout 64 frame movie cycles. Movie cycles repeated continuously until a response was made. Long movie lengths were opted for (lasting a total of 1.6sec) because a single movie might be presented for a prolonged period before a response is made. Repeating movies of shorter lengths often led to impressions of unintended cyclical patterns (due to correspondence noise). An average dot density of 0.6dots/degree was maintained, with the average inflow of dots to the apertures matched to that of the outflow. Dot velocities were dictated by the global transformation that was applied over the main aperture. Six translation components were used: ±1.69; ±2.07; $\pm 2.46°\text{sec}^{-1}$, and shear components were selected from the range -0.94sec^{-1} - $+0.94\text{sec}^{-1}$. This choice of values ensured that all dots of a single display moved in the same direction (i.e. there are no stationary points) even for the 'whole' configurations.

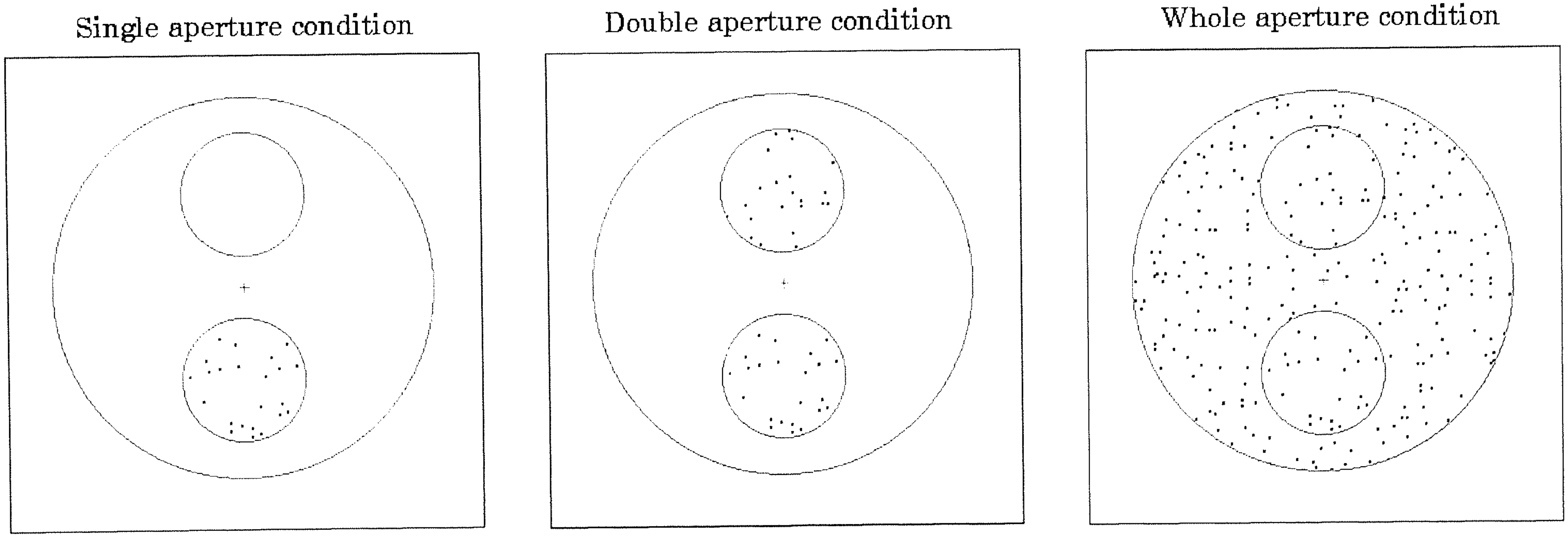


Fig. 6.8. Showing the three configuration conditions of Experiment 3c. Random dots could fill either one sub-aperture ('single aperture condition'), both sub-apertures ('double aperture condition) or the main aperture including both sub-apertures ('whole aperture condition'). In all conditions the dots would adhere to motions dictated by a global motion transform (containing a uniform vertical speed gradient). All three example displays show sub-apertures separated by the same distance. In the experimental displays, one of three separation distances was used. Note that the red outlines delineating the apertures were present in the displays and persisted throughout the experimental procedure. For printing economy, dots have been illustrated here in black, whereas in the actual displays, dots were white and presented on a black background.

Slant matches were made with a polar projected wire-frame cube on a neighbouring monitor, positioned so that the centre of the cube was vertically aligned with the centre of the main aperture. The cube could be rotated around one axis, enabling it to be pitched at points between ±90°. A white 'X' extended over the front face of the cube to help indicate its pose. The cube was orientated so that when the front face was pitched at 0° it appeared frontoparallel to the observer. Pitching was controlled from the y-coordinates of a PC mouse in steps of 1°. The observer was free to attempt to make as near perfect a match as they could by switching fixation between the green fixation cross of the main display monitor and the cube representation of the neighbouring monitor. It was found that, with practice, fixation was shifted between monitors about 5-7 times in a trial. When satisfied with a match, this could be accepted with a left mouse button press. In the rare instances when it was felt that the RDK stimulus was not eliciting a salient slant impression - usually because the random allocation of dots did not lead to a sufficiently scattered dot distribution - a right mouse button press could be used to regenerate the display and start the trail anew.

Thirty slant estimates were obtained for each configuration (whole; double; single), separation (0.30°; 3.02°; 5.72°), target (upper; lower), sign of translation, and sign of shear. Within this block of 30 trials there were 3 magnitudes of translation. The magnitudes of the shear components for these were systematically allocated from the range of discretely sampled shear magnitudes. This resulted in data collection from a total of 2160 trials. The conditions for each of these trials were pre-computed prior to the experimental runs, randomly shuffled and divided up into 30 blocks of 72 trials each. Each block was carried out as an individual experimental session.

6.4.2 Results: Slant-matching experiment

Before carrying out a tabulation of the results it is first necessary to consider what the effects of a 'global analysis' of translation would be. Comparisons between the single aperture conditions and the double and whole conditions are key to determining whether or not a global estimation occurs. If a global estimation can be obtained by taking into consideration translation in areas peripheral to those being attended, then the slant estimations from the single and whole conditions would be expected to differ. If, additionally, analysis can also operate over discrete patches of stimuli – as would be required for the search experiment for which this is antecedent - then slant estimations should differ in a similar way between the double- and single-aperture conditions. Fig. 6.9

illustrates the effect of 'global analysis' for a specific 1-D speed gradient. Here the def/trans value is negative and therefore physically consistent with a floor-like stimulus. The perceived slant magnitudes clearly depend on whether the visual system bases slant judgements upon local translation estimates or on a global translation estimate (derived from the parallax motion of either the main aperture or both sub-apertures). Global translation estimates would lead to collinear slant impressions, while local translation estimates, arising from the individual sub-apertures, would yield differing slant impressions. From the figure it is clear that if the global analysis does occur, the direction with which the slant magnitudes differ would depend on which of the two sub-apertures is pre-cued for report. In the example given in fig.6.9 a target in the upper region would lead to steeper slant impressions under a global analysis, however, a target in the lower region would lead to shallower impressions. Of course the converse would be true for positive values of def/trans.

Fig. 6.10 divides each condition into a further two sub-groups based upon the direction that the slant would be pushed in the double and whole conditions, were they based upon a global analysis. A steepening of the slant would be expected if the global translation were greater than the local translation. These conditions will be denoted from hereon with the expression 'Global>Local'. A shallowing of the slant would be expected if the local translation were greater than the global translation. These conditions will be denoted the expression 'Local>Global'.

For each of these sub-divisions 120 estimates were made. In figures 6.11-6.13 the data are plotted in terms of their 'local def'/'local trans' values (denoted from hereon with the expression 'def/trans'). The inset values show that sample-size is always below 120. There are two reasons why some estimates were removed from the analysis. The most common estimates to be removed were zero estimates. Frequently they were thought to reflect responses in which the observer was uncertain as to the sign of slant (typically occurring at low def/trans magnitudes and prima facie inspections, suggested no bias for one sign of def/trans over another). A second group of discarded estimates were those in which the stimuli were perceived as orthographic rotations. These interpretations tended to occur in only the whole aperture trials in which the def/trans values were negative and of a large magnitude. In such trials an interpretation of motion parallax would elicit a ceiling-like impression whereas an orthographic (structure-from-motion) interpretation would elicit a floor-like impression.

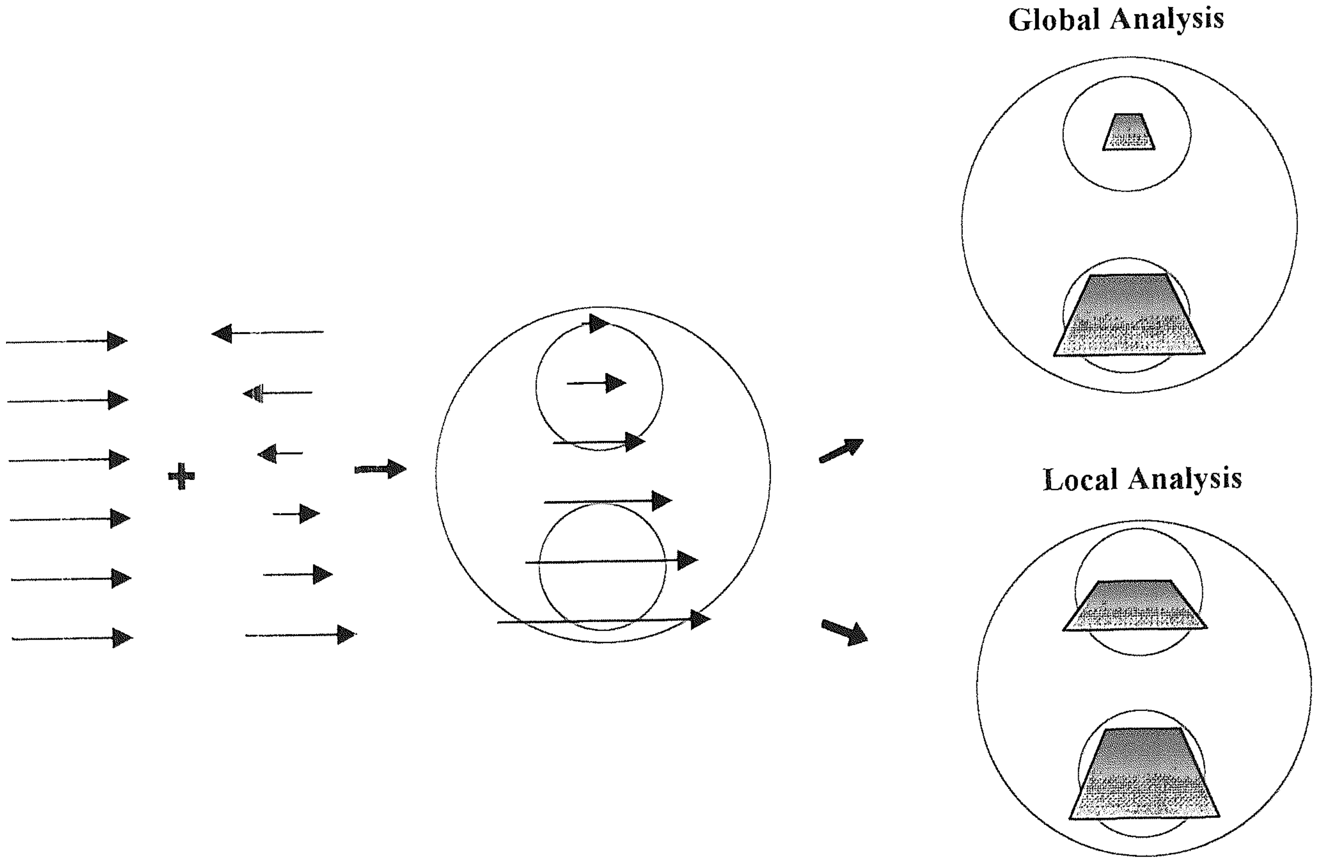


Fig. 6.9. Two possible surface interpretations for the motion parallax stimuli in the double aperture configuration. The two motion flows on the left illustrate that translation and shear components can be used to construct the 1-D global speed gradients used in experiment 3c. A resultant speed gradient that is consistent with a floor-like slant is illustrated. The two interpretations of slant are depicted in the two displays on the right. The shaded trapezium shapes overlaid upon the aperture outlines represent slant impressions. If a global estimation of translation is used in slant estimation then the slant angles from both sub-apertures would be identical and the two would appear coplanar (upper diagram). If a local estimation of translation is used then the slant angles would appear different and no coplanarity would be apparent (lower diagram).

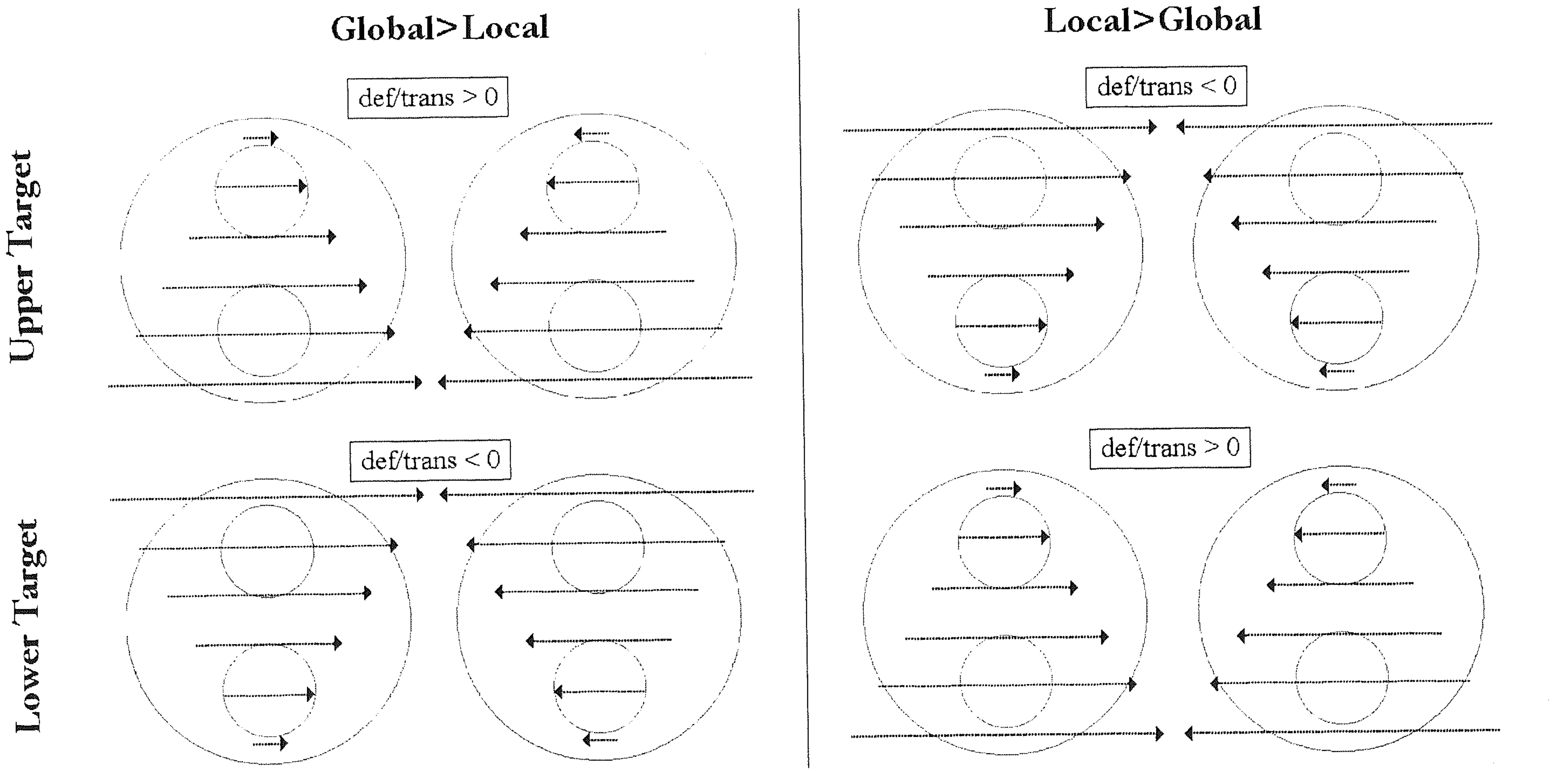


Fig. 6.10. Sub-apertures with a green outline denote the location where slant estimation is based (target). The 8 display depictions illustrate how the stimulus conditions can be reduced to 4 conditions: (1) 'Global>Local' &.'def/trans>0', (2) 'Local>Global' &.'def/trans>0', (3) 'Global>Local' &.'def/trans<0', (4) 'Local>Global' &.'def/trans>0'. The two directions of global translation are always pooled. See main text for further descriptions.

Figures 6.11-6.13 also include curves of best-fit to the Freeman *et al* (1996a) model (eqn.6.1). Estimations of the m_Ψ and t_Ψ parameters were achieved using the in-house curve-fitting software (based upon the Levenberg-Marquardt minimisation algorithm) of KaleidaGraph 3.09. The plots show there is degree of variability within the slant estimations for the local def/trans values, however, the sample-sizes used enabled all model fits to achieve R^2 values greater than 0.95.

One factor that should be borne in mind is that the range of available def/trans values depends upon target location. The Local>Global conditions at the greatest eccentricities covered the largest range of values, while the Global>Local conditions at the lowest eccentricities covered the smallest range. If the model, given by equation.6.1, sufficiently describes the data, this confound should have no systematic effect upon the two free parameters.

Parameter values obtained by fitting the data to the Freeman *et al* (1996) model are plotted in figures 6.14-6.15. Figure 6.14 plots the values from the double aperture conditions. Figure 6.15 plots the values from the whole aperture conditions. Both figures include the values from the single aperture conditions for comparison. If the model were capable of fully describing observer data then the m_Ψ values (the saturation level of the arctan function / gain) would be expected to remain around the same irrespective of condition. In the single aperture conditions the t_Ψ values (the steepness of the arctan function) should also remain the same. This would also be true for the double and whole aperture conditions if a purely 'local analysis' of translation were occurring. If, however, a 'global analysis' were occurring then the t_Ψ values would be expected to differ, with the 'Global>Local' conditions producing larger values than those of the corresponding single aperture conditions, and the 'Local>Global' conditions producing lower values than those of the corresponding single aperture conditions. All m_Ψ values should remain the same however.

Double aperture condition versus single aperture condition

In figure 6.14, comparison of the single and double aperture configurations for the 'Local>Global' condition shows only small and inconsistent estimates of m_Ψ and t_Ψ values. The m_Ψ values lie within a range of 0.4-0.5. Veridical judgements based exclusively upon the motion parallax cues would be expected to yield values of 1.0. The discrepancy between this and the observed values probably reflects general slant underestimation arising from the conflicting pictorial cues to flatness (Braunstein, 1968).

A comparison of the single and double aperture configurations for the 'Global>Local' conditions does show consistent differences. There is an increasing difference in the t_Ψ values as the target eccentricity, and hence the separation of the sub-apertures decreases. Furthermore, the direction of this difference is consistent with that predicted from the 'global analysis' hypothesis, in which the stimuli of the unattended aperture in effect acts to steepen the slant estimates of the attended aperture. Note, however, that the m_Ψ values also diverge in a similar monotonic fashion. To establish whether the t_Ψ differences can be explained by the corresponding changes in m_Ψ, and not 'global analysis' *per se*, t_Ψ values for the double aperture data were determined for when the m_Ψ parameter was set to be the same as those in the corresponding single aperture conditions. These data are plotted in Figure.6.16. Clamping of the m_Ψ parameter to those of the single aperture conditions eliminated the differences, suggesting the previous effect was not due to a simple 'global analysis' of translation.

Whole aperture condition versus single aperture condition

In figure 6.15 comparisons of the single and whole aperture configurations showed no signs of systematic differences when both t_Ψ and m_Ψ were set as free parameters, either in the 'Global>Local' or the 'Local>Global' conditions. When m_Ψ was clamped to those derived from the single aperture conditions a suggestion of 'global analysis' did emerge (fig. 6.17). At all three sub-aperture separations of the 'Global>Local' condition, the whole aperture configurations generated fits with higher t_Ψ values than did the single aperture configurations. Differences in the 'Local>Global' conditions were small and non-systematic however.

A slant sign asymmetry for m_Ψ and t_Ψ parameter fits

It was previously stated that if the Freeman *et al* (1996a) model is sufficient to describe slant matching data then the m_Ψ value would be expected to remain the same irrespective of the condition. While it is true that the values do not vary drastically, remaining around 0.5, comparison between the 'Global>Local' and 'Local>Global' conditions is suggestive of a possible difference, with the former tending to provide fits in which m_Ψ values are slightly larger than the latter. Considering just the single aperture configurations, there is an average difference of around 10%. A systematic trend for m_Ψ may also present itself in the ‘Global>Local’ condition of the single aperture configuration. In this case values seem to increase linearly with an increasing eccentricity. How might these effects be explained? One possibility is that the proposed model is insufficient for describing slant-matching data. Another possibility is that the differences reflect anisotropy between the two signs of slant and that this is not symmetrical around fixation (i.e. the anisotropy is only present for target eccentricities below fixation). This speculation was borne out of observations of the raw data (figures 6.11-6.13). In some plots, the data points from the two limbs of the arctan functions tended to be asymmetrical (this is particularly clear when comparisons are made between the data points of the upper and lower limbs of the ‘Local>Global’ plots in the single aperture conditions).

To investigate these asymmetries further, model fits for the single aperture data were carried out on the individual limbs of the arctan functions. In terms of figure 6.10 this involves fitting the model to each of the four prescribed sub-conditions. Because the resulting fits were fairly unconstrained, R^2 values remained low at around 0.5. Despite this, some consistent effects could be discerned from the analysis (see fig.6.18). The first thing to note is that t_Ψ values for the floor-like slants (positive def/trans conditions) are consistently higher than those for the ceiling-like stimuli (negative def/trans conditions). Of primary interest here though is the observation that targets appearing in the lower visual field show an asymmetry for slant sign with respect to m_Ψ value. m_Ψ values derived from data for floor-like matches tended to decrease with increasing target eccentricity. m_Ψ values derived from data for ceiling-like matches tended to increase with increasing target eccentricity.

Single aperture condition: t_Ψ and m_Ψ are free parameters

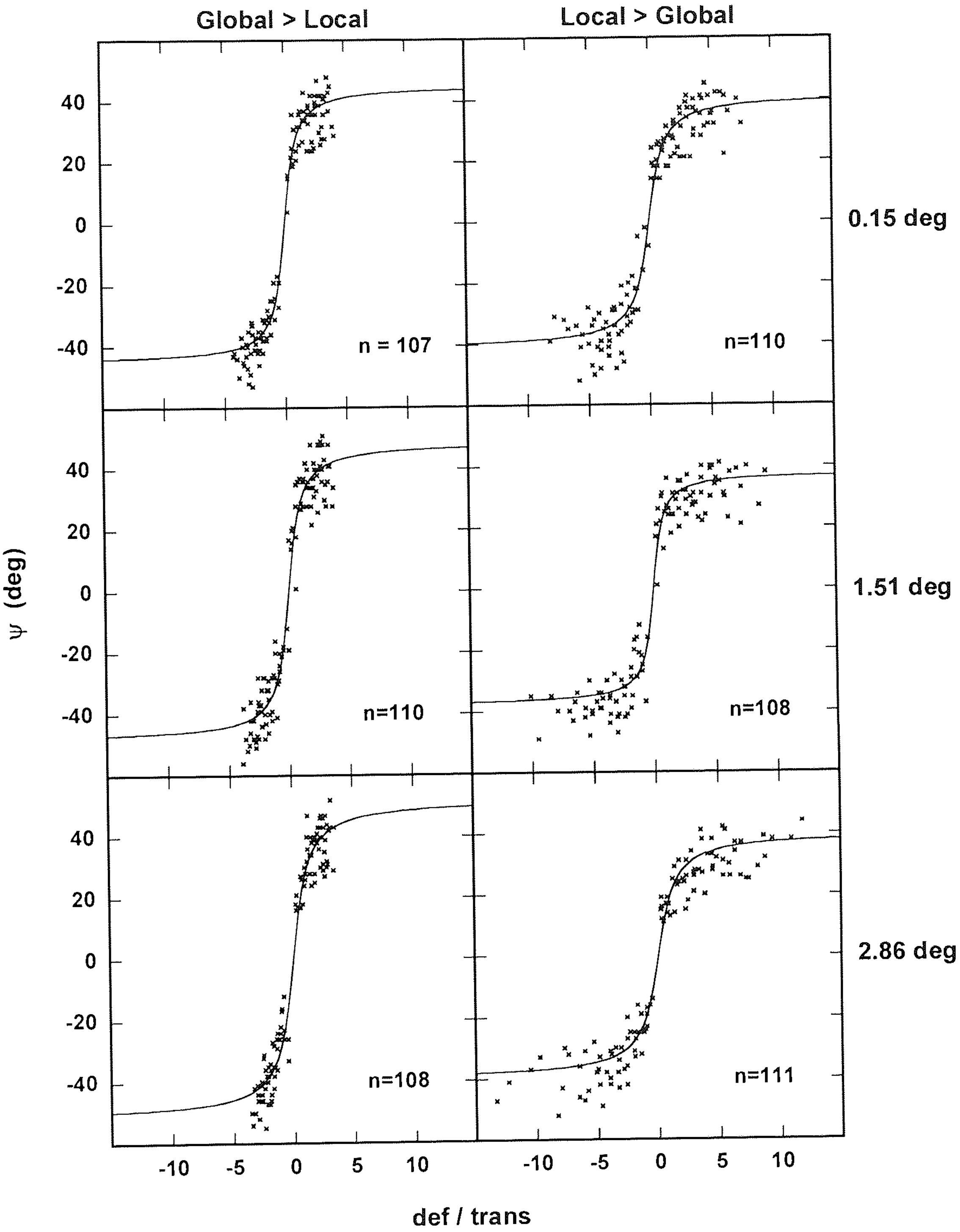


Fig.6.11. Slant estimation data for single-aperture conditions. Estimated slant is plotted as a function of def/trans, where def denotes the average local def component; and trans denotes the average local trans component. The left and right column plots show the 'Global>Local' and 'Local>Global' conditions respectively. Each row of plots shows a different amount of sub-aperture eccentricity. The inset n values denote the number of estimates that have been included. Curves show best-fits for the model described by eqn.6.1, in which both parameters were free to vary.

Double aperture condition: t_Ψ and m_Ψ are free parameters

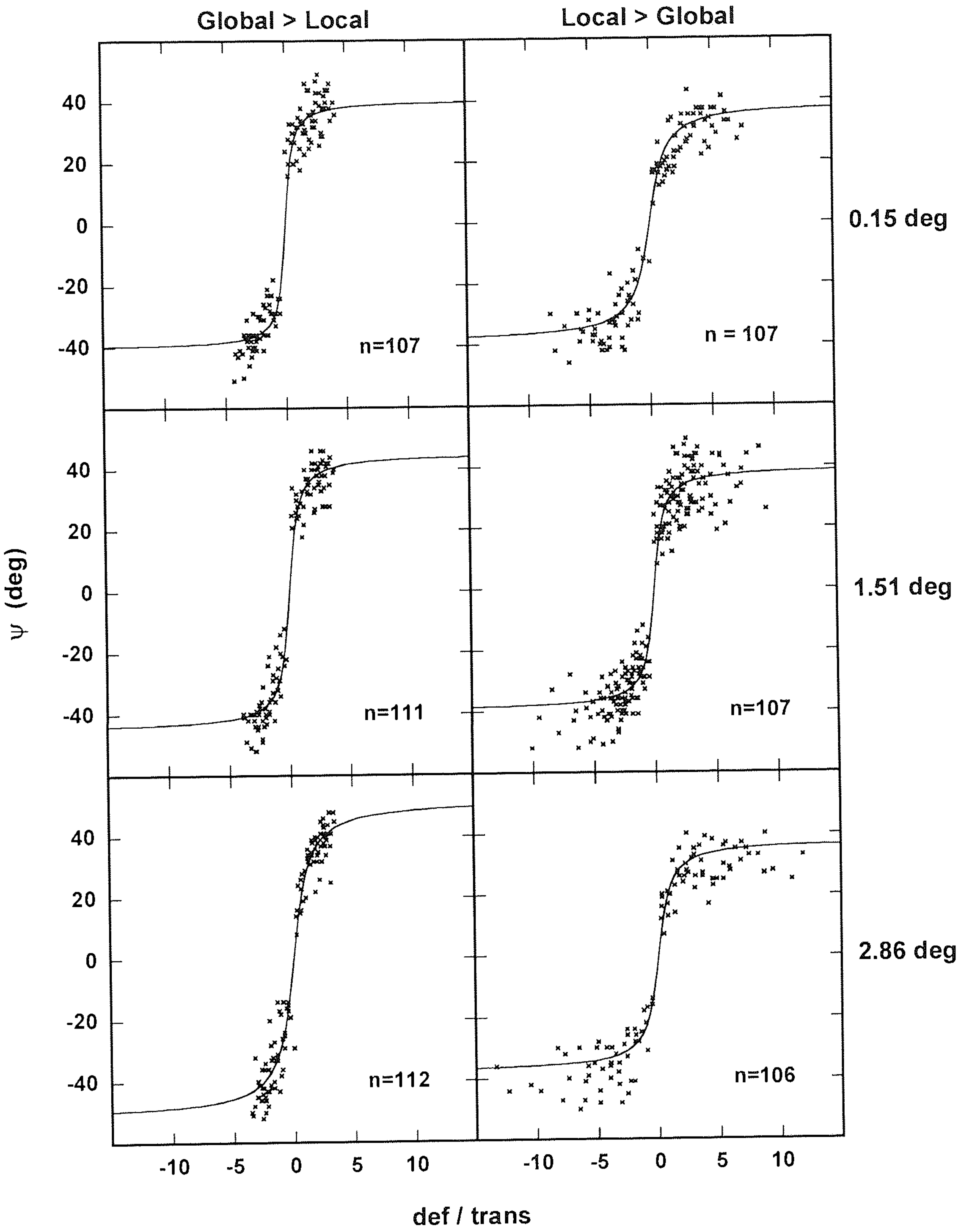


Fig.6.12. Slant estimation data for double-aperture conditions. Estimated slant is plotted as a function of def/trans, where def denotes the average local def component; and trans denotes the average local trans component. The left and right column plots show the 'Global>Local' and 'Local>Global' conditions respectively. Each row of plots shows a different amount of sub-aperture eccentricity. The inset n values denote the number of estimates that have been included. Curves show best-fits for the model described by eqn.6.1, in which both parameters were free to vary.

Whole aperture condition: t_Ψ and m_Ψ are free parameters

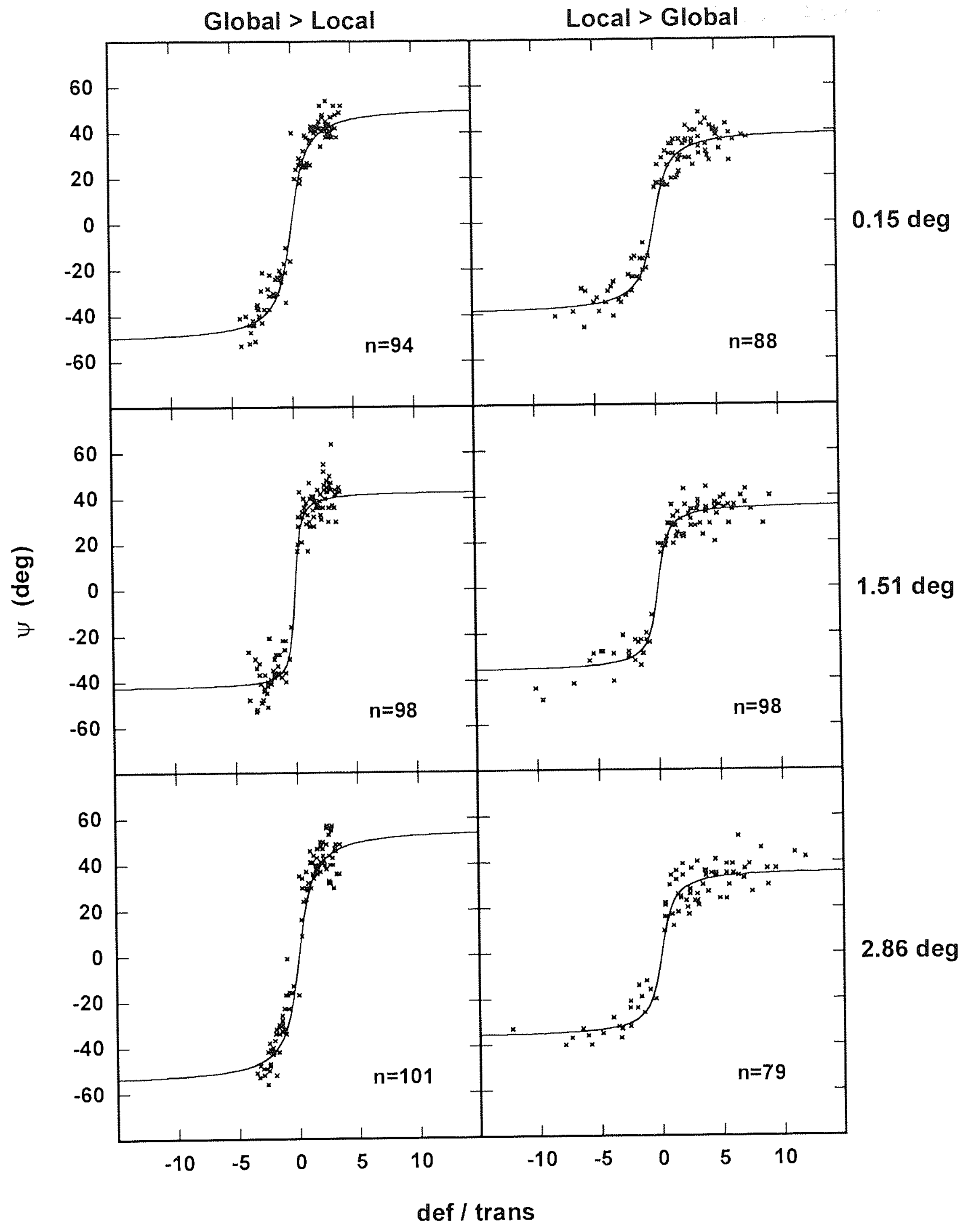


Fig.6.13. Slant estimation data for whole-aperture conditions. Estimated slant is plotted as a function of def/trans, where def denotes the average local def component; and trans denotes the average local trans component. The left and right column plots show the 'Global>Local' and 'Local>Global' conditions respectively. Each row of plots shows a different amount of sub-aperture eccentricity. The inset n values denote the number of estimates that have been included. Curves show best-fits for the model described by eqn.6.1, in which both parameters were free to vary.

Double aperture condition: t_Ψ and m_Ψ as free parameters

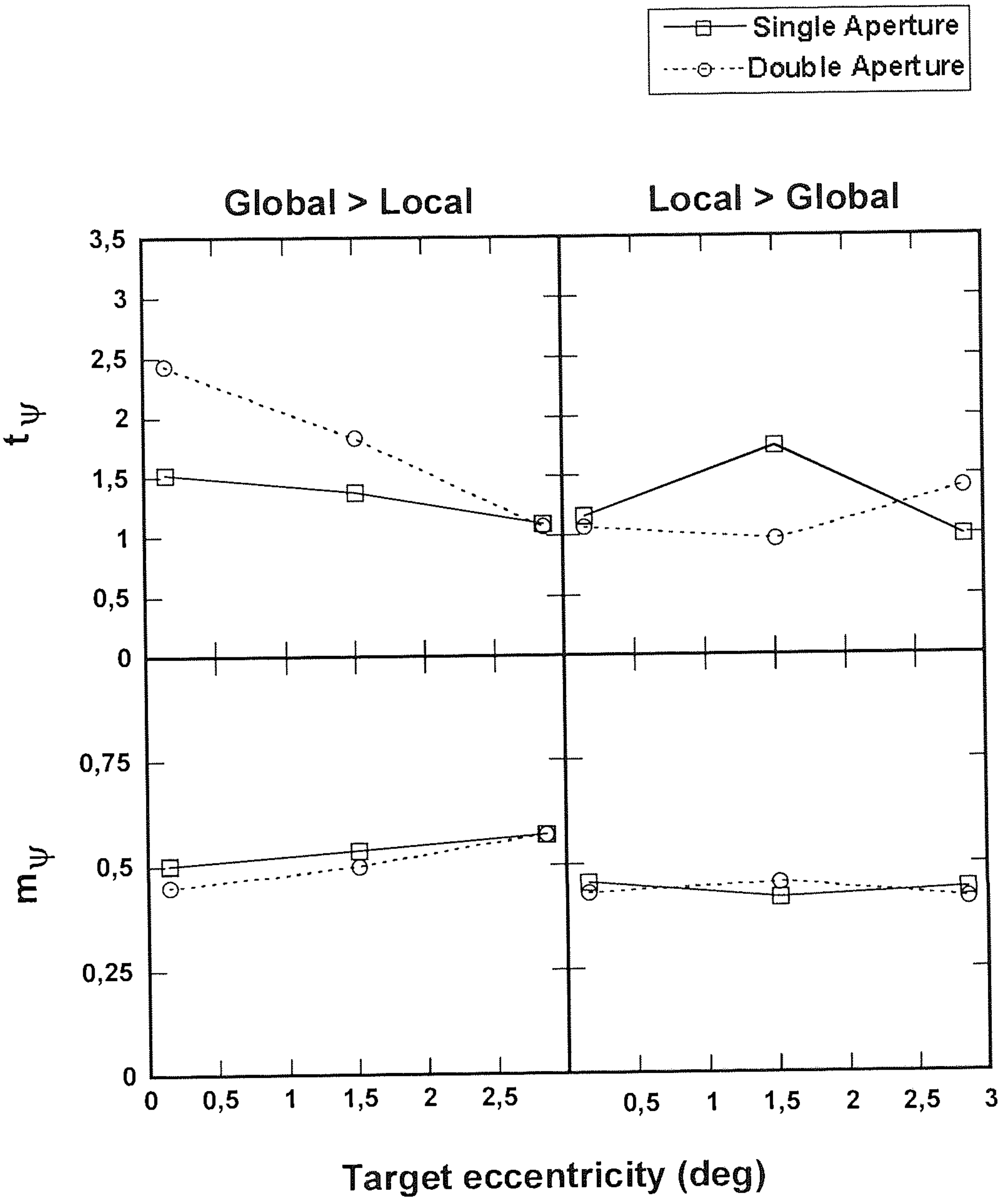


Fig.6.14. Parameter values for the model described by eqn.6.1 are shown both for the single-aperture and double-aperture conditions. t_ψ values are shown in the upper two plots and m_ψ values in the lower two plots.

Whole aperture condition: t_Ψ and m_Ψ as free parameters

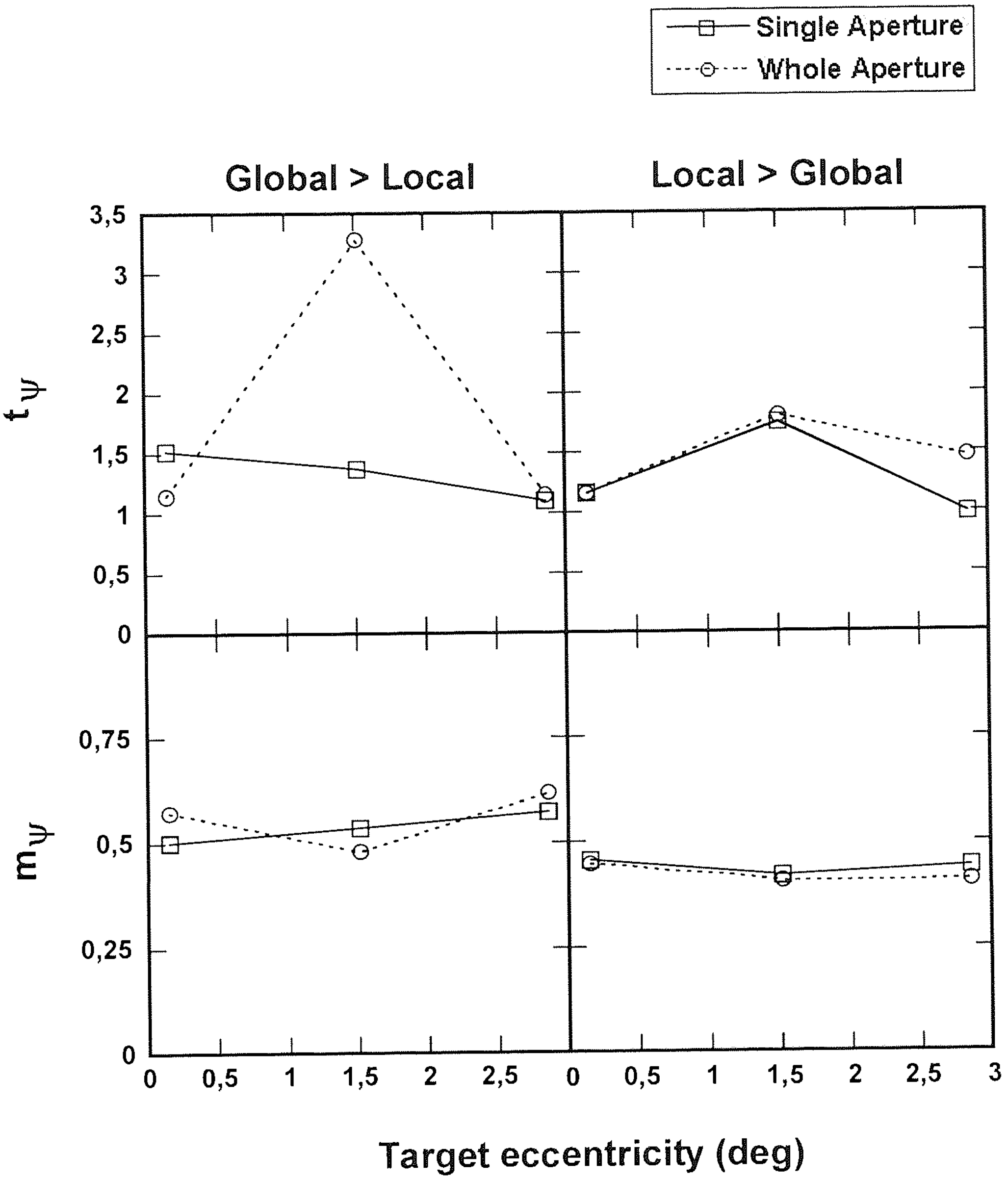


Fig.6.15. Parameter values for the model described by eqn.6.1 are shown both for the single-aperture and whole-aperture conditions. t_ψ values are shown in the upper two plots and m_ψ values in the lower two plots.

Double aperture condition: t_Ψ as a free parameter

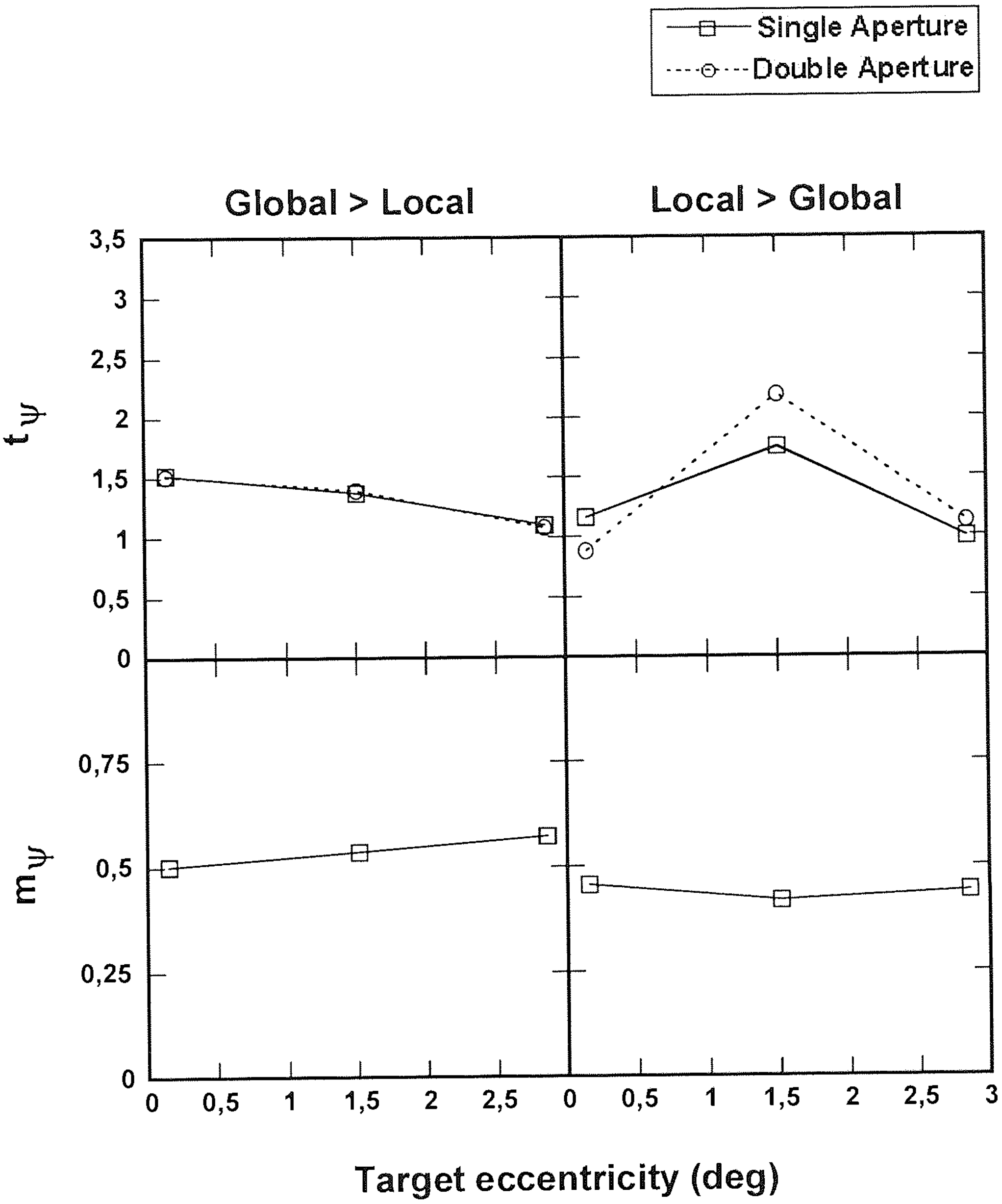


Fig.6.16. Parameter values for the model described by eqn.6.1 are shown both for the single-aperture and double-aperture conditions. t_ψ values are shown in the upper two plots and m_ψ values in the lower two plots. t_ψ values for the double-aperture condition were determined for when the m_Ψ values were fixed to those of the corresponding single aperture conditions.

Whole aperture condition: t_Ψ as a free parameter

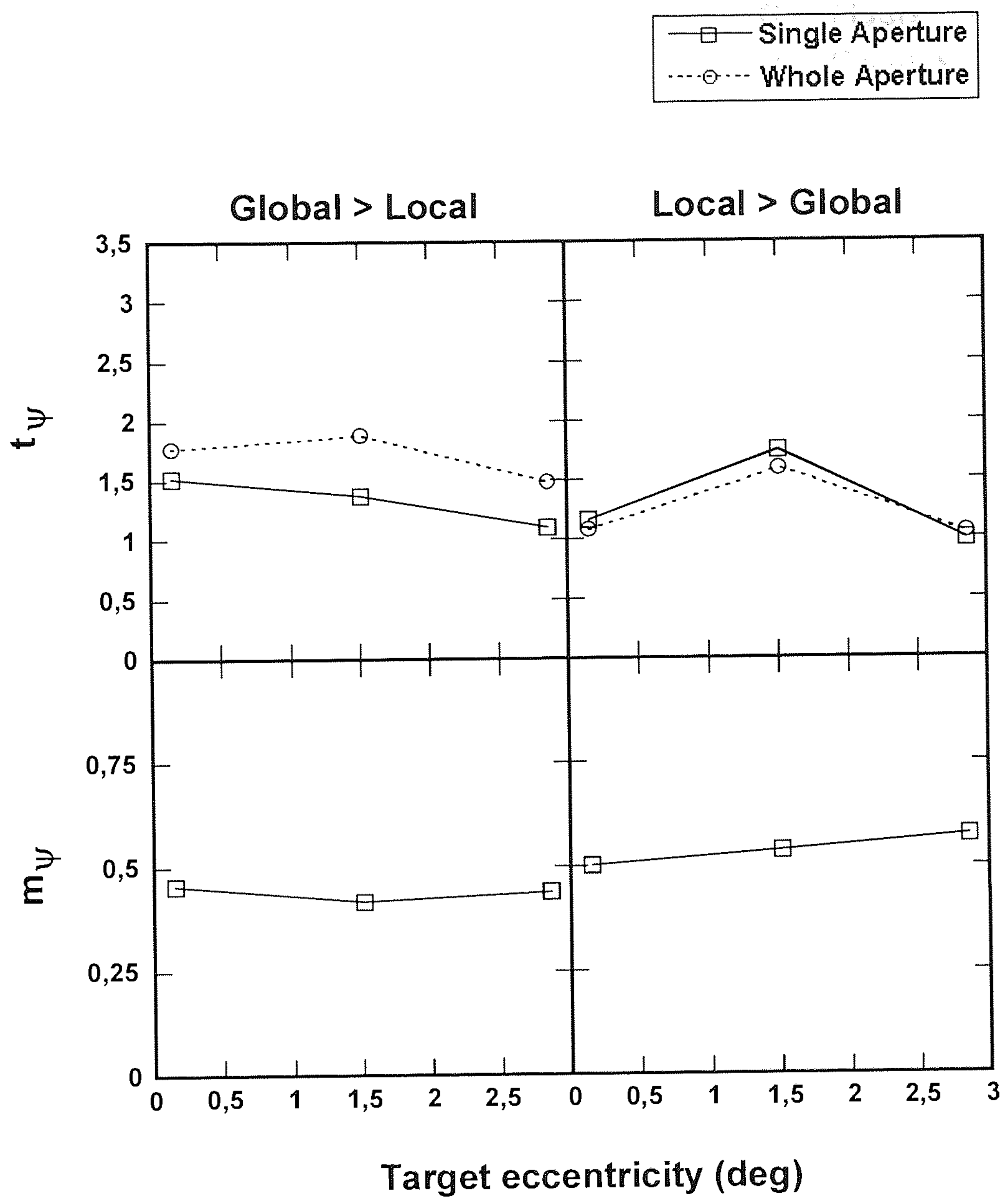


Fig.6.17. As in fig.6.16 except that the t_Ψ values have been determined for the whole-aperture conditions.

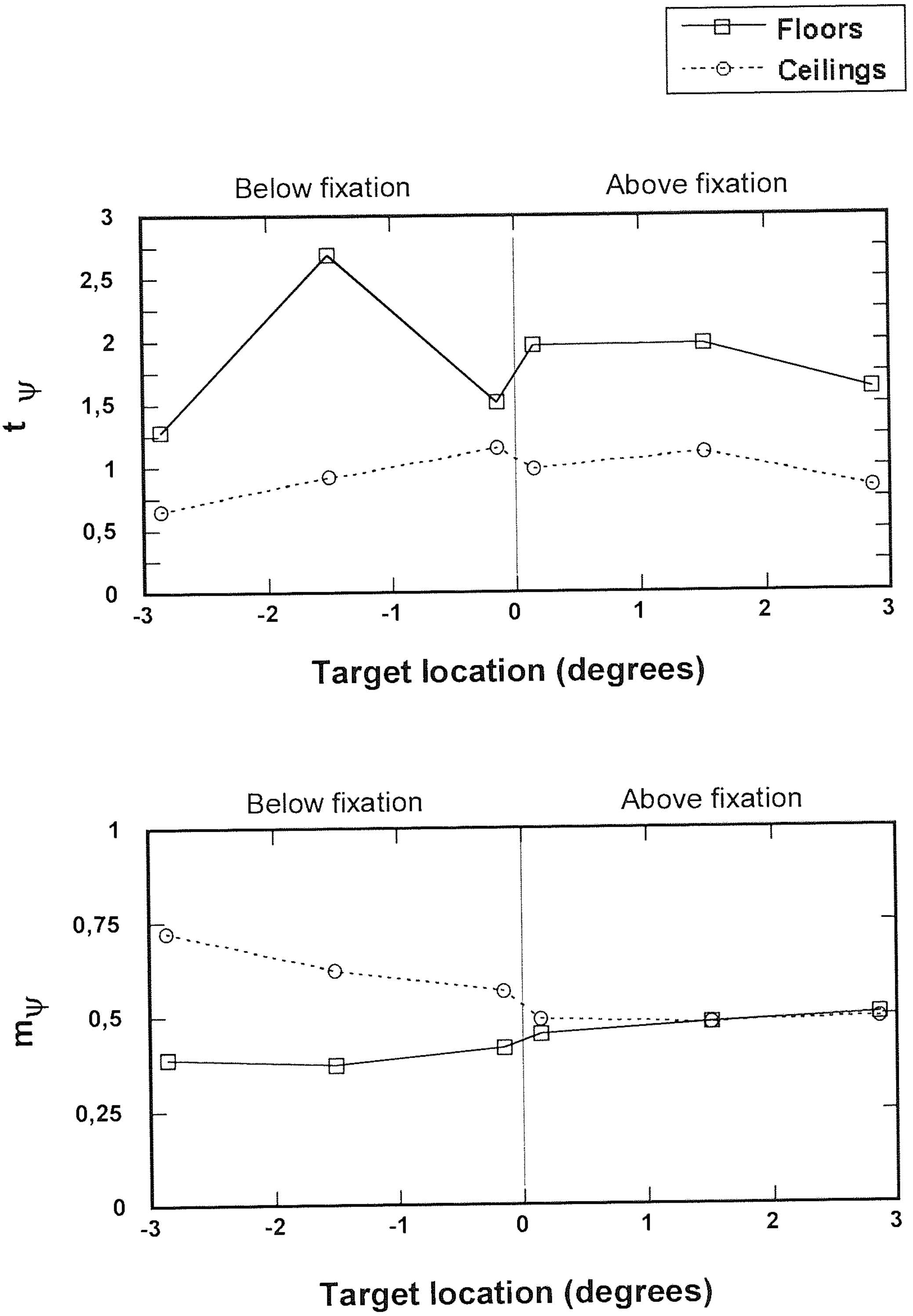


Fig.6.18. Parameter values, t_ψ and m_ψ, are shown for the single-aperture conditions only. Values have been determined for 'Local>Global' and 'Global>Local' conditions for data gathered at each of the six target (sub-aperture) eccentricities.

6.4.3 *Summary and conclusions: Slant matching experiment*

No evidence for a 'global analysis' of translation

The aim of this preliminary slant matching experiment was to establish the viability of using individual patches of motion parallax (each containing the same component of shear) to generate impressions of common coplanar surface. If this were possible, search displays could be constructed in such a way that the distractor patches group by virtue of their membership to a common planar surface.

It was noted that, in order to achieve this grouping, the visual system would need to carry out a global (pan-patch) analysis of translation. If a purely local (patch-wide) analysis was carried out, the individual slant impressions might not conform to an implicit planar surface. In this experiment no evidence was found to suggest a 'global analysis' of translation could occur over discrete RDK patches. If a 'global analysis' had taken place then the t_ψ parameter values should have taken on consistently higher values in the 'double-aperture condition' relative to the 'single-aperture condition'. The plots in figure 6.16 show that there is no clear evidence for such as trend. It remains a possibility that the permanent presence of the aperture outlines may have helped the subject restrict their analysis of translation to the local 'target' sub-aperture. This may have been the case, but such demarcation is also likely to be present with the stimulus patches of a search display, where the presence of randomly allocated dots can act to clearly specify the spatial extent of each of the patches.

A slant sign asymmetry for estimates of m_Ψ but only when viewed below fixation

An additional finding of passing interest did emerge from a reanalysis of the data for the 'single aperture conditions'. In this analysis, the model (eqn.6.1) was fitted separately to the data for the two signs of slant. Figure 6.18 shows that m_ψ values were higher for ceiling-like slants than for floor-like slants. Moreover, this anisotropy was apparent only when the motion parallax stimuli appeared below fixation. The reason for this anisotropy remains unclear.

Chapter Seven:

Visual Search - Further discussion

7.1 Chapter Overview

In this chapter the serial/parallel processing dichotomy, which is implied by models such as FIT, is given further consideration. Of particular importance is research suggesting that visual searches can sometimes be carried out in parallel even when the search slopes suggest a serial analysis. Three reasons for why this might occur are discussed. The first is that search performance may be limited solely by the attentional demands of the display stimuli and not the availability of specialised mechanisms to process them. A second reason may be that the decision criterion is relaxed as set-size is increased. The third is that a HTT model of visual search may not be appropriate, and therefore, the addition of distractors may lead to an increase in the noise at the decision stage.

A study by Thornton & Gilden (2001) is also reviewed in which singleton ('odd-item-out') searches were carried out for contraction and expansion stimuli. Consistent with the results of experiment 1 of this thesis and Braddick & Holliday's (1991) study, positive search slopes were found. However, using an alternative search paradigm, Thornton & Gilden found evidence of parallel processing for these two radial motions.

In the preceding chapters of this thesis the possibility that search was being carried out in parallel, even when steep search slopes were being obtained, was not ruled out. It was acknowledged that multiple representations of the search stimuli might have been processed simultaneously but access to them was limited. What was not explicitly discussed in the previous chapters was the nature of this limited access. One possibility is that representations can only be accessed serially. Another is that they can be accessed in parallel, but the efficacy with which they are accessed decreases as the number of monitored representations increases. In this latter situation search can be considered as parallel but positive search slopes would be observed.

7.2 Discrepancies with Feature Integration Theory

7.2.1 There is no dichotomy in search slopes; they lie on a continuum

FIT emerged in part from the initial empirical findings of Triesman & Gelade (1980), in which search slopes appeared to be either near flat (<5ms/item) or steep (>20ms/item). In the years that followed this seminal study it became evident that no clear-cut dichotomy actually existed. Rather set-size slopes lie on a continuum, including the range between 5-20ms/item. It is not only the stimulus dimensions in which the target-distractor difference occurs that determine slope magnitude; as has been discussed in previous chapters, manipulating distractor heterogeneity (section. 6.1.2) and target-distractor saliencies (section 4.4.6) can also have an affect. Moreover, these manipulations affect slopes in a gradual manner. For instance, feature searches with large target-distractor differences will tend to yield near-zero search slopes but when these differences are decreased not only do general performance decrements result but search slopes also increase (Foster & Ward, 1991; Duncan & Humphreys, 1989)[1].

The above findings imply that the minimum criterion of 20ms/item for serial search is somewhat arbitrary, as is the 5ms/item maximum criterion for parallel search. As a result different studies might set different criteria. Davis & Driver (1994) for example classed searches with effects up to 10ms/item as evidence for parallelism. In terms of the SSTS strategy of FIT a 10ms/item Hit slope would imply attentional shifts every 20ms. It may seem unlikely that attentional shifts could be carried out so rapidly but this is still open for debate (Wolfe, 1998. pp.51).

7.2.2 Is the 'beam' metaphor more appropriate than the 'glue' metaphor?

Triesman & Gelade's FIT (1980) proposed the function of attention to be like a 'glue', enabling two or more features from approximately the same region to be perceived as a single entity or object. This conception differs somewhat from the views that have emerged from cueing studies. Instead, these have attributed focal attention the metaphor of an 'attentional beam', acting to enhance the saliency of any stimuli falling within that beam. If the beam metaphor is more appropriate then this might raise doubts about the concomitant preattentive/attentive dichotomy espoused by FIT (table 1.1). This chapter will briefly examine some of the evidence that favours the 'beam' metaphor, and latter the potential ramifications for FIT.

Attention as a 'beam'

There is a vast literature that can be drawn upon to support the view that focal attention acts to enhance stimulus saliencies. A key study in establishing this view has been that of

---

[1] It should be noted that Wolfe's Guided search model (Wolfe *et al.*, 1989) does maintain the idea of a clear dichotomy between parallel and serial processing while attempting to account for the continuum in slopes.

Posner *et al* (1980), in which the effects of pre-cuing on stimulus detectability were examined. The observer's task was to make a speeded response when they detected a luminance onset. The validity of cues and the type of cue were varied in the experiments. On the majority of trials the cues were signalling the correct target location. On some trials however the cue was invalid. Trials employed two types of pre-cues: 'peripheral' and 'central'. On the valid trials of the 'peripheral' cues, the location of the target would be indicated directly: an outline at the location would briefly appear. On valid trials of the 'central' cues, a central arrow would indicate the location of the imminent target. Both cue types were found to improve RT's by an average of 50ms. Because such a task requires the detection of a feature-like stimulus, FIT would consider the required processing to be automatic and preattentive, entirely independent of attention. Clearly the reduction in the spatial uncertainty of the target did have an effect. The investigators credited these improvements to perceptual enhancement and speculated that stimuli within the beam of focal attention had become more salient.

Can attention modulate perceptual processes?

If Posner's assertion that focused attention leads to perceptual enhancements is correct then how might this occur? One possibility is that attention can increase the overall gains of specific mechanisms. The selective enhancement of perceptual mechanisms has been suggested to occur in MT (Huk & Heeger, 2000) and V1. In terms of a Signal Detection Theory (SDT) framework, these effects might be considered as increases in d', through the positive lateral shifting of the S+N (target) distributions (assumed additive noise). It is also possible that selective attention could operate by attenuating unattended stimuli (e.g. Moran & Desimone, 1985). In a SDT framework this might be conceived as a decrease in d' through the negative lateral shifting of the N (distractor) distributions.

There is a certain amount of evidence, in addition to the pre-cueing studies, to suggest that attention can modulate perceptual processes. The selective enhancement of activity in early perceptual mechanisms has even been implied for local motion processing (Watanabe & Miyauchi, 1998; Watanabe *et al*, 1998). Lee, Itti & Braun (1999) have also claimed that data for detection and discrimination with high and low attentional loads are well accounted for by a model in which attention enhances signals from V1 mechanisms most selective for the stimuli and suppresses signals from mechanisms with a lesser selectivity; the net effect being an increase in orientation tuning for attentive vision.

FIT proposes feature processing to be preattentive, but the studies documented above suggest that attention may even be influencing early (and presumably feature-like) processing. If this is true, then it raises the following questions: Is attention needed even for so-called feature search? And, with sufficient attentional effort could all conjunction searches also yield near-zero search slopes? The first question has been addressed by Joseph *et al* (1997) who tested the attentional demands of a feature (orientation) search task by introducing a second, concurrent and attentionally demanding task. If feature search is genuinely preattentive, search should have remained highly efficient irrespective of the attentional demands of the secondary task. Instead, steep positive set-size slopes were obtained. Nakayama & Joseph (1998) have subsequently argued that all searches demand attention but only when perceptual load is high does search performance become limited by attention.

7.3 Parallel searches may also be inefficient

There are a number of reasons for suspecting some inefficient searches to be caused by something other than the strict Serial Self-Terminating Search (SSTS) advocated by FIT. The 2:1 ratios between Hit ad C-R slope magnitudes that are predicted are frequently violated as demonstrated from the slope magnitude in table 4.1. (Section 4.3). Furthermore, a study by Ward & McClelland (1989) has also reported RT variances to differ markedly from those predicted by SSTS.

7.3.1 Search performance can be limited by attention alone

In the previous section, 7.2, it was speculated that all searches demand attention, but only when the perceptual load is high (i.e. each stimulus item is complex) will search performance become limited by attention (Nakayama & Joseph, 1998). Upon reaching this attentional limit, search might be executed in one of two ways as a result. It could be carried out sequentially or it could be carried out in parallel, albeit at a sub-optimal efficiency. Models assuming the latter have become known as parallel limited-capacity models and assume that increases in set-size lead to the increasingly parsimonious division of the fixed attentional resource. A trade-off occurs therefore between the number of stimuli to be searched and the time taken to search each of them, resulting in positive

RT/set-size slopes[2]. Fig.7.1 shows the potential loci of the capacity limits to be either at the perceptual stages or the stage of decision.

In FIT the limiting factor is only ever assumed to be the amount of neural substrate available, and so, if there were special-purpose mechanisms present, search would be carried out in a very efficient manner. Parallel limited-capacity models on the other hand do not attribute inefficient search to a limit on neural resource but to a limit on attentional resource.

Although the search slopes that have been observed for the radial, deformation and motion parallax stimuli in this thesis (chapter 4-6) were steep, this section argues that the cause of these search inefficiencies may not necessarily have been due to serial search.

---

[2] There are other reasons as to why increasing the perceptual load might impair search performance. These include lateral masking effects or the effects of multiplexing perceptual signals.

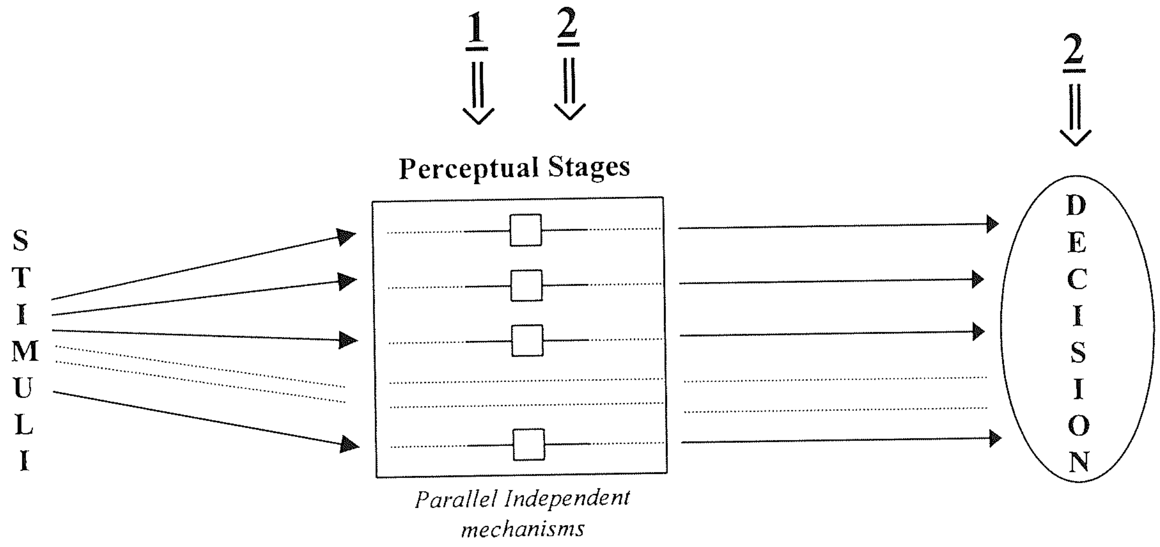


Fig. 7.1.

1. **Loci of limiting factor in FIT**. Inefficient search arises because specialised mechanisms are not available for the preattentive processing of the display stimuli. Processing therefore requires attention, but because attention is thought to be a limited resource, FIT proposes a processing bottleneck that operates spatially. Consequently the processing of multiple stimuli is thought to be carried out sequentially, on an item-by-item basis.

2. **Loci of limiting factor in Parallel Limited-Capacity models**. Inefficient search is not attributed to limited processing substrate but to a limited attentional resource at either the perceptual or decision stages. Instead of resulting in a processing bottleneck the limit is thought to lead to a graded division of resources between the multiple stimuli. This attentional resource may reflect the decreasing efficiency with which the decision stage is able to monitor individual analysers as the number of analysers monitored increases. Alternatively it could reflect the limits at the perceptual stages, possibly arising from metabolic constraints.

7.3.2 Search accuracy as a function of set-size

The results section of Chapter 4 refers to a tendency for Miss rates to increase with set-size. This occurred despite the fact that observers were attempting to maintain a constant accuracy irrespective of set-size. If, for whatever reason, accuracy is progressively relaxed as set-size increases, then a speed-accuracy trade-off may result, a corollary being that RT slopes are underestimated. This did not become a key issue for the search experiments reported in this thesis as none of the slopes approached zero and so parallel unlimited-capacity processing was never a viable explanation. Had the Miss rates showed the opposite trend however, this point might have become important.

In another search experiment (not reported in this thesis) another paradigm was used to test an alternative approach to the measurement of set-size effects (this was carried out only on the complex motion stimuli that were tested in experiment 1). Its aim was to factor out the potential effects of the involuntary shifts in decision criterion. Here visual search was considered within a signal detection theory (SDT) framework and search displays were presented for limited durations. A single interval (1IFC) YES/NO paradigm was used, in which either a target-present or target-absent display could appear on a single trial. By obtaining percentages of Hit and F-A rates, *d'* values could be obtained. For these stimuli, studied with display times of 240ms, 430ms, 840ms, 1250ms, the data was noisy, making direct comparisons between conditions (such as between *defl* and *defl+trans*) difficult. Such a SDT approach to visual search was therefore abandoned in favour of RT measures.

7.3.3 Decision noise may account for small search slope magnitudes

While the above experimental approach described in section 7.3.2 proved unsuccessful, there is a growing body of work in which a SDT framework has been used to probe visual search performance (e.g. Palmer *et al*, 1993; Palmer, 1994; Bennett & Jaye, 1995). This novel approach has similarly enabled the issue of speed-accuracy trade-offs to be circumvented. However, instead of estimating a single *d'* for displays containing odd-item-out stimuli (always defined by a polarity switch in one of the components) these experiments establish just-noticeable-difference (JND) thresholds for a target item embedded in displays of distractor items. For example, the target might be defined by an orientation difference and the JND for the target orientation established at various set-sizes.

Implicit in these SDT approaches is an assumption that visual search is best modelled by a low-threshold theory. Previously, this has not been the case: most visual search theories have tacitly assumed a high-threshold theory (HTT) for target detection. HTT will be considered in more detail in Chapter 9 of this thesis, but a brief overview is provided here. Observer decision is based upon a decision rule akin to the operation of an inclusive OR/Maximum-of-outputs (MAX) operator. Observers respond 'yes' if, and only if, one or more analysers have exceeded some criterion activation. Moreover, this criterion is high enough to prevent non-target stimuli from being confused with target stimuli, implying that the monitored analysers are incapable of exceeding criterion activity in response to non-target activity.

If a feature search is considered in terms of FIT it becomes clear that HTT is implied. FIT makes the prediction that if observers maintain constant accuracy, an increase in the number of distractors would not affect performance (RTs); this would imply perfect monitoring of the feature detectors but also that the additional distractors are not a source of F-As.

Uncertainty effects have argued against a high-threshold theory

HTT is unable to account for many aspects of the data obtained from threshold psychophysics. One of the main criticisms it has faced stems from the fact that uncertainty about the stimuli to be detected frequently results in higher thresholds, whether they be discrimination thresholds (Cohn & Lasley, 1974; Yager *et al*, 1984) or detection thresholds (Davis, Kramer & Graham, 1983). In many cases, a large part of the effect can be ascribed to some kind of attentional limitation (e.g. Ball & Sekuler, 1981) but in other cases, where the effect is less acute, the uncertainty effects can also be well accounted for by appealing to a low-threshold theory.

Low threshold theories accept that analyser activity is noisy and if, as in HTT, a maximum-of-outputs decision rule is assumed then the noise level alone might sometimes be high enough to exceed the requisite criterion for observer responses. This means that a lone unstimulated analyser would be capable of eliciting a F-A response. Similarly, in a 2IFC experiment, every so often, the noise interval may, by chance, produce a higher response than the signal interval. In the visual search literature this analyser noise, which can lead to F-As, is often referred to as decision noise. Increasing stimulus uncertainty is presumed to increase the number of mechanisms whose activity needs to be monitored, therefore increasing the amount of decision noise. In order to maintain a certain level of behavioural performance the deleterious effects of decision noise need to be compensated by increasing signal levels.

The study of Davis *et al* (1983) looked at uncertainty effects for sinusoidal gratings and found that uncertainty about the location and spatial frequency of the gratings tended to raise detection thresholds. However, it was deduced that these threshold increases could be wholly accounted for by a decision noise hypothesis, and so there was no need to invoke explanations based upon attentional limitations. Uncertainty about the contrast had no effect upon detection thresholds. This is consistent with a typical multiple analyser model, in which a decision is based upon a simultaneous monitoring of perceptual mechanisms (Graham, 1989); in this case selective for spatial frequency and spatial position but not for contrast.

Spatial uncertainty can account for small set-size increases

Of most relevance to visual search are the potential effects of decision noise arising from uncertainty about spatial location. If the assumption is made that for every distractor added there is a corresponding increase in the spatial uncertainty of the target, it becomes conceivable that under a low threshold theory framework, decision noise might be able to account for certain magnitudes of set-size effect. Indeed, Palmer and colleagues have confirmed that for odd items defined by a simple ("feature-like") difference (for example, line orientation, line length, colour or size), log-JND/log-set-size slope are entirely consistent with decision noise models (Palmer *et al*, 1993; Palmer, 1994). Furthermore, these slopes are all of a very similar magnitude. As expected, more complex searches, such as the finding of a 'T' junction among 'L' junctions, yield higher slope magnitudes[3] and, consequently, require some additional explanation, possibly in the form of attentional limitations.

When cast in terms of SDT, a threshold approach to visual search has at least one clear advantage over RT approaches; clear predictions can be made about the upper slope limit for the effects of decision noise. Slope magnitudes exceeding this limit must attribute at least some of the effect to capacity limitations. The maximum slopes used to diagnose the presence or absence of processing limitations from RT data is chosen more arbitrarily. Even for RT measures in which F-A and Miss rates appear constant, an effect of decision

---

[3] See Bennett & Jaye (1995) however. In their study they report that decision noise can also account for the set-size magnitudes for complex stimulus items (letters).

noise cannot be ruled out. It may be that the subjects' decision-making behaviour is sophisticated enough to be able to compensate for the deleterious effects of decision noise by selecting appropriate criterion levels for each set-size. In summary, decision noise explanations may need to be considered before other alternative explanations.

The experimental conditions of experiments 1-3b of this thesis that showed inefficient search all possessed search slopes with magnitudes greater than 20ms/patch and positive error rate trends. It seems unlikely therefore that a decision noise hypothesis alone could account for these inefficiencies of search.

7.4 Multiple-target search for complex motions

7.4.1 Redundancy gains and the 'pure target' conditions

An alternative RT approach for determining whether an inefficient search is in fact parallel has been developed by Thornton & Gilden (2001). This study was mentioned in Chapter 2 with respect to the singleton searches that were carried out by the researchers. For these searches positive set-size effects were found for rotation (56ms/item), expansion (16ms/item), contraction (13ms/item) and translation (20ms/item). Thornton & Gilden employed another two techniques in an attempt to further characterise the nature of the searches. Their primary interest was concerned with the associated attentional demands for the processing of these motion types. Acknowledging the fact that parallel search can sometimes masquerade as serial when diagnosed with a traditional singleton search, they proposed that a slightly modified paradigm could act as a more decisive test for parallelism. These modifications were in the form of additional conditions; ones in which additional target motions could take the place of distractors. The conditions particularly pertinent to their diagnosis of parallelism were ones in which no distractors were present. These were referred to as 'pure target' conditions. Decrements in RT with 'pure target' set-size ('redundancy gains') are presumed to be indicative of parallelism, if and only if RT variances do not decrease concomitantly. If the variances do decrease then the 'redundancy gains' could just as well be attributed to the fact that increasing the number of targets improves the chances that, in any single attentional fixation, a target motion will be processed.

The 'pure target' conditions included set-sizes of one, two and four. Trials for these conditions were randomly intermixed with trials in which no targets were present (50% of all trials) and trials in which a mixture of targets and distractors were present. For the 'pure target' trials of rotation, no 'redundancy gains' were observed. For contraction, expansion and translation however, 'redundancy gains' were observed. Their corresponding F-A rates showed no evidence of a speed-accuracy trade-off and there were no systematic decreases in the RT variances as a function of set-size. From these results it was concluded that the contractions, expansions and translations of the multiple-target conditions must have been processed in parallel, despite the fact that their singleton set-size slopes showed significant positive inclinations.

7.4.2 'Pure target' conditions: visual search or signal detection?

In Thornton & Gilden's (2001) study, the 'pure target' trials were intermixed with other trials, but within a single experimental session there was no uncertainty about the type of target that was to be searched for. Given the above, are these 'pure target' trials best considered in terms of visual search? Typically, visual search tasks have required observers to detect odd-items from a number of distractor items. In the 'pure target' trials of Thornton & Gilden however, the observers could not have based their performances upon odd-item detection, as distractor motions were never present. These conditions may therefore be more appropriately considered in terms of a signal detection task in which RT was being used as a performance measure.

7.4.3 The source of the redundancy gains

The authors remained uncommitted as to the nature of the parallelism they reported. The positive set-size slopes of the singleton searches could have arisen either from decision noise, the existence of a limited attentional resource, or inhibition from surrounding distractors.

If the decision noise hypothesis were correct, the 'redundancy gains' in the 'pure target' conditions could be ascribed to probabilistic advantages arising at the decision stage. In the visual search literature these explanations are often referred to as Race Models, and would occur because increases in the number of target patches would also lead to increases in the number of channels being monitored. Because the channels are thought to be noisy, the time that is needed for each of them to reach detection criterion would vary on a trial-to-trial basis. If it is assumed that on average processing times are comparable, the addition of target motions would act to decrease the average amount of time needed to complete the processing of at least one of these targets.

An alternative explanation for the redundancy gains comes from accounts in which the signals from the multiple target patches are summed at some stage prior to decision. In the literature covering divided attention such models have been termed Coactivation models (Miller, 1982). Because their 'pure target' conditions were blocked and there was no uncertainty with respect to the target type, observers need only have monitored the channels carrying out summation across motion patches. Such a strategy would be akin to the summation/averaging strategies that have been reported in the visual search literature that has measured JNDs (Morgan *et al*, 1998; Burr & Baldassi, 2000). Thornton & Gilden did cite weak evidence in favour of such a Coactivation account but they left the issue largely unresolved.

In summary, Thornton & Gilden's (2001) study showed that, as the number of displayed target patches is increased (for expansion, contraction or translation), RTs for detection decrease. Furthermore, based upon the directions of the trends in error rates and RT variances this effect does not seem to be attributable to explanations based upon a serial analysis. In contrast no redundancy gains were observed for rotation. A serial processing of rotation could not therefore be ruled out. Its greater set-size effect for singleton search also suggests that it is processed more inefficiently than the other motions. From the results of an entirely different methodology, the researchers claim further support for the notion that rotations stimuli are more attentionally demanding than radial and translational stimuli.

7.4.4 Disadvantages of using a multiple-target RT search paradigm

(1) As previously mentioned, in the pure-target trials, observers could not have based their performance upon the detection of odd items. However, in other trials containing distractor motions, target detection could have relied upon the presence of odd items. It is possible that search could have been carried out in very different ways depending upon the condition; the intermixing of the pure-target and non-pure-target trials might therefore have been inappropriate and served only to confuse interpretation of the data. Insight into this issue might have been gained from the C-R slopes but these were not reported.

(2) To ensure redundancy gains are not due to an increased chance that a single attentional fixation is able to process a target stimulus, it must be confirmed that RT variances do not decrease with target number. For the motion stimuli tested by Thornton & Gilden (2001) no systematic decreases were observed and parallelism for radial and translation could be diagnosed. However, this method of diagnosis is even more cumbersome than standard RT visual search: to confirm parallelism there must be (a) redundancy gains present, (b) RT variances that do not decrease with pure-target number and (c) error rates that do not increase with pure-target number.

7.5 Future direction

Up until now, the primary concern of this thesis has been with the nature of perceptual access to 3-D representations that are defined by motion. Freeman *et al* (1996b) hypothesised that when motions induce three-dimensional representations they can be accessed more rapidly than two-dimensional representations induced from motion. This hypothesis was tested using singleton search tasks but despite trying a number of different motion stimuli and display configurations, no evidence has been found to support it.

In experiment 1 of this thesis, the search for radial and deformation (*def1*) stimuli was found to be highly inefficient. Search for rotations was efficient but this was most probably due to cues deriving from inter-patch interactions. In this chapter another category of search has been discussed; one in which all stimuli are processed in parallel but the search is inefficient. Using a modified search paradigm Thornton & Gilden (2001) set out to determine whether radial and rotation motions are processed in serial or inefficient parallel. Parallelism was confirmed for both signs of radial motion but not for rotational motions. Because the key conditions that they used to diagnose parallel search (the ‘pure target’ conditions) did not require the observer to carry out an odd-item-out detection, the parallelism could have derived either from a probabilistic summation at the decision stage or a within channel summation at the perceptual level.

In order to examine which source of summation is the most likely for complex motions, in experiment 5a-e, a subthreshold summation technique is employed. This permits quantitative estimates of summation while circumventing issues such as speed/accuracy trade-offs. Prior to this, chapter 8 will first describe experiment 4, which confirms processing to be parallel for pairs of orthogonal motion components.

Chapter Eight: Experiment 4 – Orthogonal motion components

8.1 Chapter Introduction

Using a subthreshold summation technique, Meese & Harris (2001a; 2001b) have found that complex motion detection tends to improve slightly when a second (orthogonal) complex motion is superimposed on the first. These small advantages are consistent with a probabilistic summation at the decision stage, and the idea that orthogonal motion components are processed by independent mechanisms. In another study, that made use of suprathreshold motion stimuli under restricted viewing conditions (i.e. 85ms stimulus duration followed by a 400ms backward mask), De Bruyn & Orban (1993) concluded that multiple, superimposed, global motions are probably not processed by independent mechanisms, but by common general purpose mechanisms (see section.1.3.2). These, they propose, can process different global motions in a rapid serial manner. Their hypothesis could also be used to explain the findings of Meese & Harris (2001b). In which case, the superior performance of their compound conditions over their component conditions could have arisen because one orthogonal motion component was always processed to a greater extent than the other, during the 300ms stimulus presentation. After stimulus offset the observer may, owing to visual persistence in the early visual system, have been able to retrospectively analyse the neglected motion component.

However, the experiments of De Bruyn & Orban have not conclusively ruled out the possibility that superimposed orthogonal motions are processed in parallel by independent mechanisms. There is another possible explanation for their data. It could be that the process of identifying direction sign, which their task demanded, requires serial analysis at the decision-making stage but not the perceptual stages. In this scenario both of the orthogonal motion flows would be processed in parallel up to the level of decision, but because of the restricted viewing times, the two decisions concerning motion direction could not be made (experimental evidence for this explanation, with alphanumerical stimuli, is described by Duncan, 1980). Such serial analysis at the decision stage could be a manifestation of attentional limitations for example.

In one experiment, the authors do explore the above possibility using a negative cueing paradigm. In each experimental trial, two superimposed motion components were presented under the same restrictive viewing conditions, followed by a prompt to report the direction of one of these motions. In any single condition the observer would be certain as to the type of just one of the motions, but would be uncertain as to the type of the other (it could be one of two types; for example radial or horizontal translation). Only the second motion that had the a priori uncertainty was tested. It was found that despite the observer having foreknowledge as to which of the two motions was to be ignored, performance in the direction discrimination task remained poor. The authors argue that this result is consistent with the way in which a general-purpose mechanism might be expected to operate: top-down processes would be able to impose templates upon the general-purpose mechanism ready for it to respond to a certain motion, but templates would not be able to configure it to be an "everything except..." mechanism. However, it is also possible that the negative cueing result can be explained by properties of the decision stage. For example, it may be the case that, at the decision stage, outputs from 'non-target' mechanisms are not successfully ignored under circumstances of divided attention (i.e. uncertainty about motion type).

If the poor performance of De Bruyn & Orban's superimposed motion conditions can be attributed to limits at the decision stage then, in a detection experiment, such as that of Meese & Harris (2001b), performance advantages might still be expected, even under conditions of restrictive viewing. This prediction derives from the fact that in a simple detection experiment, the decision stage would not be required to identify the mechanisms that gave the high output.

Here a detection experiment, similar to the one of Meese & Harris (2001b), was carried out to see whether probability summation could still be observed for orthogonal motions under restricted viewing. The restrictive viewing was achieved through the use of an interruption mask. This should have disrupted any processing that persists in the visual system after stimulus offset. Previously, Burr & Santoro (2001) have reported long periods of (ideal) temporal summation for complex motions (2-3 sec). The time taken for the stimulus plus mask is much briefer than these summation periods. A strategy of serially processing the two motions of the compound condition would not, therefore, be expected to yield performance advantages over the component motion conditions.

8.2 Models of summation

As mentioned previously, signal summation might arise from within-channel summation at the perceptual level (also referred to as physiological summation) or probabilistic summation at the decision stage.

8.2.1 Within-channel summation

If performance were determined by within-channel summation, then ideal linear summation is predicted. Fig. 8.1 illustrates a within-channel summation model. Each of the input lines (1-m) represents the processing carried out by a set of component channels. For the purpose of the experiments within the next two chapters, the component signals may be thought of as representing the processing of individual local motions. At a second perceptual stage, a process sums these component responses. The entity that carries out this summation is referred to here as a central integrator and is illustrated in fig.8.1 as the larger box.

In this two-stage summation model the effect of increasing the proportion of component channels that carry signal will inevitably lead to increases in sensitivity. If the sensitivities for each of the component channels are assumed equal, then behavioural sensitivity can be expressed using the following signal-to-noise ratio:

$$Sens_{cmpd} = \frac{nS}{\sqrt{kN_{\mathrm{ext}} + mN_{\mathrm{int}} + N_{\mathrm{cent}}}} \qquad (8.1)$$

Where m is the number of component channels that the central integrator pools from, k is the number of these channels that can be considered as being stimulated by external noise and n is the number of these channels that also signal. N_{ext} denotes the variance of the external stimulus noise, N_{int} the variance of the intrinsic noise associated with each component channel and, N_{cent} the variance of intrinsic noise associated with the central integrator.

The summation studies of this and the next chapter will assume that both the noise associated with the component channels (mN_{int}) and the noise associated with the central integrator channel (N_{cent}) remain constant, irrespective of experimental condition.

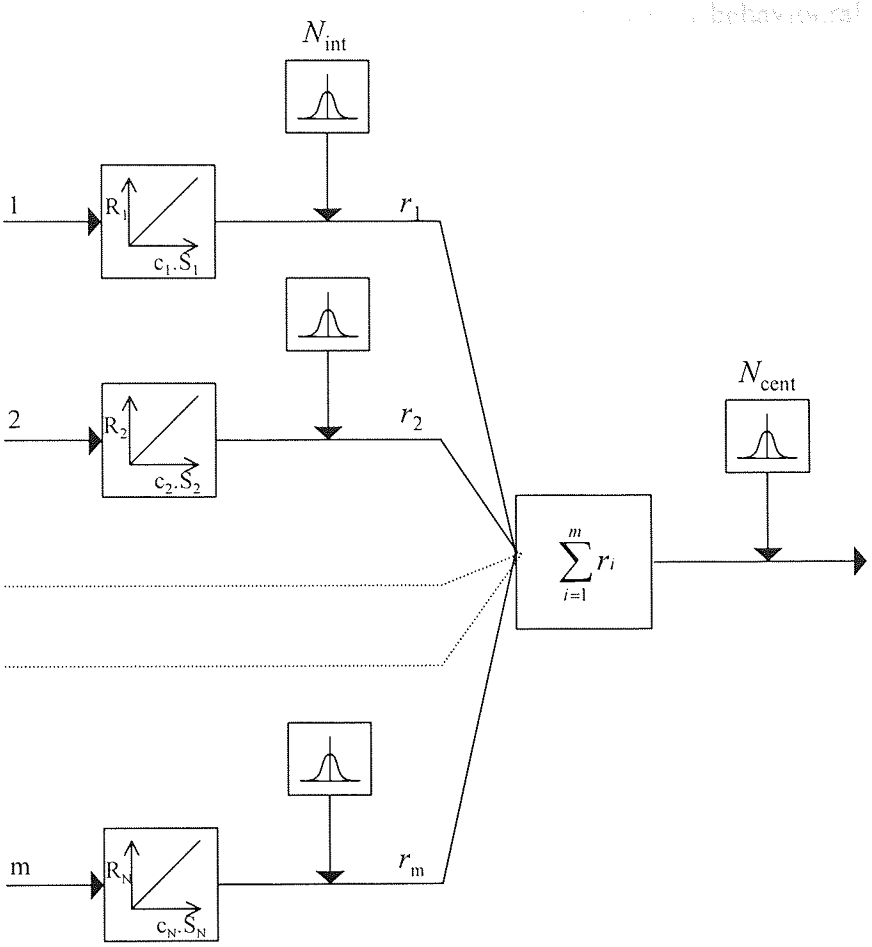


Fig 8.1. Each input line represents the processing of individual components (1-m). Some transducer function operates on each of the stimulus-driven signals $c_i \cdot S_i$, eliciting a response r_i (in this exposition the function is linear). These signals feed into a central integrator. Inset in the box denoting this central integrator is the simple summing operator that describes its behaviour. Two sources of noise are also illustrated. N_{int} denotes the intrinsic noise associated with the component channels and, N_{cent} denotes the intrinsic noise associated with the central integrator.

Extrinsic stimulus noise dominates

If the external noise of the stimulus dominates over the intrinsic noise that is associated with the channels, then eqn.8.1 will simplify to the following:

$$Sens_{cmpd} = \frac{nS}{\sqrt{kN_{\text{ext}}}} \tag{8.2}$$

In which case, if the amount of extrinsic noise is kept constant while the number of channels stimulated by signal is increased, linear increases in behavioural sensitivity are expected (linear summation), i.e.

$$Sens_{cmpd} = n \cdot Sens_i \qquad (8.3)$$

Where $Sens_i$ is the behavioural sensitivity for signal presented to a single channel.

However, if the number of channels stimulated by extrinsic noise is not kept constant but increases at the same rate as those stimulated by signal, then square root increases in behavioural sensitivity are expected (quadratic summation), i.e.

$$Sens_{cmpd} = \sqrt{n} \cdot Sens_i \qquad (8.4)$$

Intrinsic channel noise dominates

If the intrinsic channel dominates over the extrinsic stimulus noise, then eqn.8.1 will simplify to the following:

$$Sens_{cmpd} = \frac{nS}{\sqrt{mN_{\mathrm{int}} + N_{\mathrm{cent}}}} \qquad (8.5)$$

In this case, irrespective of whether the external stimulus noise is kept constant or not linear summation is expected (eqn.8.3).

Minkowski summation formula

The summation equations 8.3 and 8.4 are often expressed in terms of the more general Minkowski formula:

$$S_{cmpd} = \left(\sum_{i=1}^{n} Sens_i^{\beta} \right)^{\frac{1}{\beta}} \qquad (8.6)$$

Where β is the Minkowski coefficient. For linear summation, β is set to 1.0, and for quadratic summation it is set to 2.0.

8.2.2 Probability summation

Probability summation occurs at the decision stage by means of a maximum-output decision rule. This means that if the output from any one of the monitored mechanisms exceeds a certain criterion level, then a 'detected' response will be given. It can be isolated from within-channel summation and observed if the following conditions are met

(1) Any mechanisms that are monitored at the decision stage are sensitive to only one of the several stimulus components that may be presented to the observer (i.e. stimulus components are far apart).

(2) These mechanisms are subject to some degree of internal noise.

(3) This internal noise has some degree of independence for each of the mechanisms that are involved in the detection of a particular stimulus component.

(4) More than one group of these independent mechanisms is monitored at the decision stage during a single stimulus presentation.

(5) The sensitivity to one component is not so high relative to the others that it dominates the sensitivity to the compound stimuli.

(6) The experimental paradigm is sensitive enough to measure the performance facilitation of the compound condition over the component conditions.

The implication is that when an observer detects a compound stimulus near threshold, the detection is typically due to a single relatively high response from just one of the groups of independent mechanisms (sensitive to one of the stimulus components).

High-threshold theory: The standard Quick Pooling model

A probability summation model that has frequently been employed to explain data in spatial vision has been the Quick Pooling model (Quick, 1974). It has had considerable success in predicting levels of summation for far-apart stimuli in various domains (e.g. spatial-frequency, spatial-position domains). The model does have weak points however in that several of the assumptions it makes are known to be untrue, at least under certain conditions. These assumptions are listed below. They elaborate and are additional to those that were mentioned previously. Suspected violations of the assumptions will be discussed when appropriate. Despite these, it is hard to dismiss the model's successes as mere coincidence. It may be that, for the typical summation experiment, the assumptions simply do not become notably violated.

(1) **Guessing parameter**: If the output from any of the monitored analysers does not exceed criterion then the observer will make a guess. The probability of guessing a stimulus to be present remains constant.

(2) **High-Threshold theory:** The implication is that thresholds are so high relative to the internal noise of the mechanisms that internal noise in itself is never sufficient to exceed the criterion level for detection. This means that the visual system cannot be driven to a 'detect' state when presented with a blank/null stimulus. This is often expressed by the statement that 'no true false-alarms' occur; any F-As made by the observer must be attributed to guessing. This only occurs when the visual system is in a 'non-detect' state. A HTT assumption has been present in the many accounts of probability summation including the first by Sachs *et al* (1971). HTT has however been unable to account for certain aspects of psychophysical data.

(3) **Additive Gaussian noise & its approximation by the Weibull function:** The internal noise associated with mechanisms is assumed to be Gaussian and with a variance that is independent of signal strength. Quick (1974) used the more manageable Weibull function to approximate the cumulative Gaussians used to describe psychometric functions.

$$\psi = 1 - e^{-R^{\beta}} \tag{8.7}$$

Where, ψ = probability of detection; R = stimulus driven activity; β = slope steepness

Tyler & Chen (2000) have shown that the Weibull function becomes an increasingly inaccurate approximation to a cumulative Gaussian as β deviates from a value of 4.

(4) **Perfect unlimited-capacity monitoring:** All mechanisms sensitive to the components of the compound condition are monitored with the same efficacy as when they are monitored in the component conditions. Olzak & Thomas (1981) found what appeared to be probability summation for far-apart stimuli but when the components were set to be very far apart, no summation was observed. This seems to suggest that the assumption of perfect monitoring may be violated under certain circumstances.

(5) **Fully independent analysers:** The internal noise is completely independent for each of the mechanisms or groups of mechanisms that are involved in the detection of different stimulus components. If their internal noise were not independent then the estimation of β for the psychometric functions would be greater than that when estimated from the summation of component sensitivities. However, studies that have looked at both estimates (Meese & Williams, 2001, Robson & Graham, 1981, Watson,

1979) have found no significant differences despite β estimates varying between subjects.

In addition, if there was correlation between far apart analysers this might be expected to decrease the further apart they are. If observed summation is due exclusively to probability summation then this would predict greater summation the further apart the analysers are. There is no evidence for such a trend. Indeed there has been suggestion for the opposite: an absence of summation for very far apart analysers (Olzak & Thomas, 1981; Bonneh & Sagi, 1998).

(6) **The slope (β) values for component stimuli are the same.** Slopes values for psychometric functions of different stimulus components are assumed to have the same values (Graham, 1989. pp.174). This assumption has even been made for stimulus components stimulating different sensory modalities (Wuerger *et al*, 2003).

Making the above assumptions enables the following psychometric function to be derived for multiple, far-apart components (see Appendix 2 for derivation).

$$\psi = 1 - e^{-\sum_{i=1}^{n} R_i^{\beta}} \tag{8.8}$$

R_i = stimulus driven activity of analyser i

To predict the observers' sensitivity $Sens_{\mathrm{cmpd}}$ to a compound stimulus from the component sensitivities $Sens_i$ the Minkowski summation formula can be derived and used:

$$Sens_{cmpd} = \left(\sum_{i=1}^{n} Sens_i^{\beta} \right)^{\frac{1}{\beta}} \tag{8.9}$$

Empirical estimates from psychometric functions have found β values to be around 4±1 (e.g. Robson & Graham, 1981). Hence, estimates between 3 and 5 for β are often taken to be indicative of probability summation.

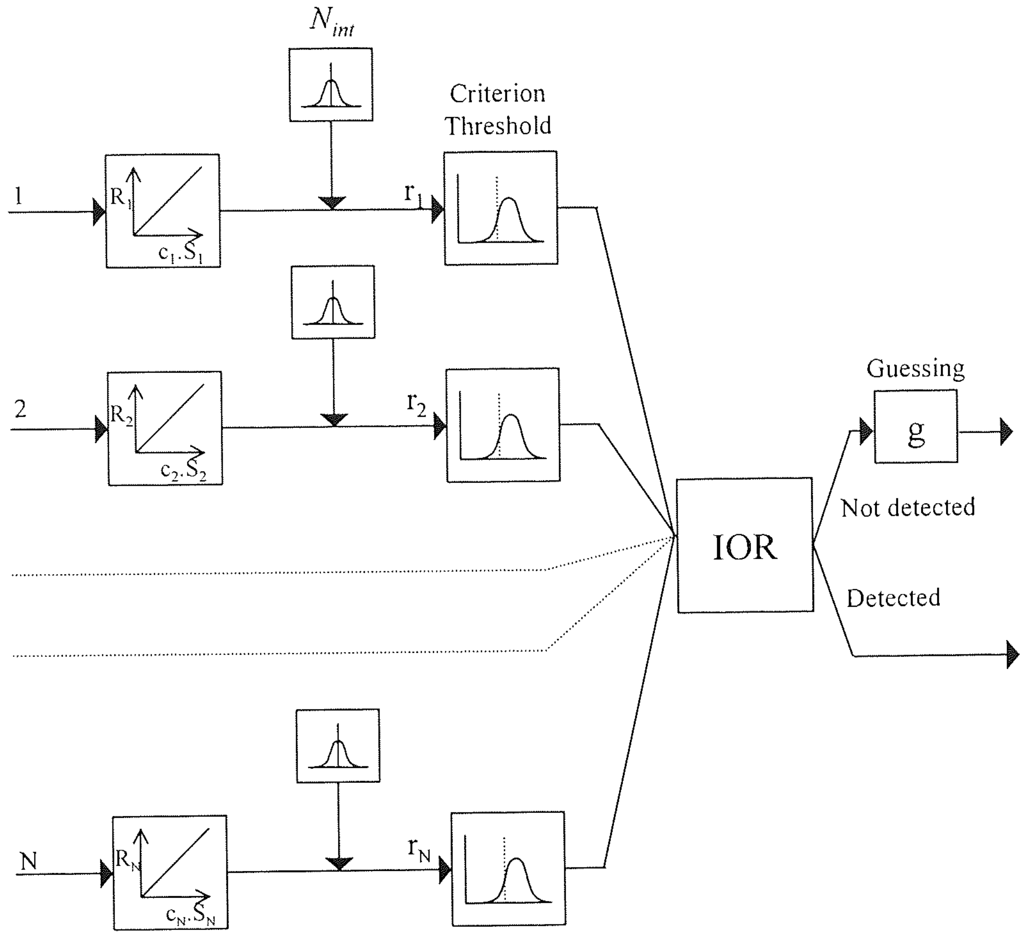


Fig 8.2. Schematic diagram illustrating how a multiple-independent-analyser model can lead to probability summation. The input lines 1-N represent the independent analysers that are being monitored by the decision stage. The responses, R_i, are then subject to uncorrelated mechanism specific noise. The resulting internal responses, r_i, are then subject to a hard threshold. If r_i exceeds this threshold then the analyser i may be said to be in a 'detect' state. The IOR operator – presumed to be at the decision stage – evokes a 'detect' state in the observer if at least one of these analysers are themselves in a detect state. In this case, the observer would respond that a stimulus had been detected. If all the analysers are in a 'non-detect' state then the observer will also be in a 'non-detect' state. However, F-As do occur; and these are accounted for by observer guessing. Observers guess with a probability, g, that a stimulus had been present in the display when the observer was in a 'non-detect' state.

Signal detection theory

Tyler & Chen (2000) made predictions about the amount of summation that would be expected if probability summation were cast in a signal detection theory framework. Assuming a linear transducer function, they predicted that for situations in which there was no extrinsic uncertainty and the numbers of component stimuli are low, summation could be approximated by forth-root summation. At higher stimulus numbers, the level of summation becomes progressively less, leading to increasingly higher beta estimates. Under conditions of extrinsic uncertainty, however, forth-root summation remains a good approximation for all stimuli containing signal, although summation does increase slightly for higher stimulus numbers.

Extrinsic uncertainty

In the following experiments the assumption is made that if a component channel is stimulated by external, stimulus noise (it may or may not be associated with a signal) then that channel is monitored. Any channels not stimulated by external noise are ignored with a 100% efficacy[1]. An assumption akin to this was clearly implicit in the previous visual search tasks - only additional distractors were considered to add to the spatial uncertainty of the target, a spatial jittering of the patches, for example, did not.

8.3 Stimulus and experimental design

A temporal 2IFC detection paradigm is used to investigate levels of summation between pairs of orthogonal complex motions. This is achieved by measuring the detection thresholds for when the components are presented alone and for when the components are presented concurrently within the same region of space. The experimental procedure is similar to that used in experiment 2 of Meese & Harris (2001b).

8.4 Stimuli and procedure

For a discussion of the general methods of stimulus generation and the estimation of detection threshold, refer to section.2.8.

8.4.1 Stimuli

In each stimulus interval, an RDK was presented and curtailed within a single large annulus. The annulus had an inner radius of 0.4° and an outer radius of 5° and a fixation spot appeared at its centre. Subjects viewed the stimuli binocularly from a VD of 69cm. Dot luminance was ramped in and out for 1.2° at the inner and outer annulus boundaries. An average dot density of 4.8dots/deg^2 was maintained throughout the stimulus presentation. Individual dots would have lifetimes of 150ms unless their trajectories crossed one of the annulus boundaries. Each image frame would be presented for 9 frames. During an image frame, 50% of the dots present would disappear and reappear in new random locations. Signal dots adhered either to an, expansion (exp), contraction (cntr), or

---

[1] In has been suggested however that in reality the observer may not be able to successfully ignore all irrelevant (unstimulated) channels (Pelli, 1985).

clockwise-rotation (clk-rot) transform depending upon the condition. Linear speed gradients were always present, with dots at the outer edge travelling at 2.7°sec^{-1}; this resulted in a speed gradient of 0.55sec^{-1}.

For each trial, the two intervals of a 2IFC trial consisted of a 300ms test period, followed by a 33ms blank period, before the presentation of a 600ms backward mask. In the test period, either noise dots alone were presented (N interval) or some combination of noise and signal dots were presented (S+N interval). The backward mask was constructed in the same way as the N interval: randomly allocated noise dots of the same luminance and density were displayed within the same annulus region. A 500ms inter-stimulus-interval (ISI) would separate the S+N and N intervals.

Blanking periods were included in an attempt to ensure that coherence thresholds remained sufficiently low to permit their measurement. In the absence of a blanking period it was suspected that the immediate effects of integrative masking might raise thresholds to un-measurable levels. Because the effects of interruption masking are thought to be optimal after short delays (30-80ms) (unlike integration masking) a stimulus-onset-asynchrony (SOA) of 333ms seemed to be a suitable compromise.

8.4.2 Procedure

Two orthogonal motion pairs were tested: clockwise rotation and expansion (*clk-rot+exp*) and clockwise rotation and contraction (*clk-rot+cntr*). A single session consisted of two stages and tested one of the two conditions. A total of 6 staircases would be completed within a session. The recording of responses occurred during the last 12 staircase reversals. Estimates of detection (coherence) threshold for the two component motions (e.g. *clk-rot* and *exp*) were obtained in the first stage. The staircase interleaving ensured that the observer was always uncertain as to which of the two component motions was going to be presented on a particular trial. In the second stage, a detection threshold for the compound condition was estimated. The displays for this condition contained both motion components, with a fixed proportion of the signal dots adhering to the first component and the remaining adhering to the second component. Signal contributions from the two motions were normalised using the threshold estimates of the first stage. The amount of signal from component 1 (Sig_{cmpd1}) and component 2 (Sig_{cmpd2}) present in the compound stimulus (Sig_{cmpd}) can be determined using the following expressions:

$$Sig_{cmpd1} = \left(\frac{C_{1alone}}{C_{1alone} + C_{2alone}} \right) Sig_{cmpd} \tag{8.10}$$

$$Sig_{cmpd2} = \left(\frac{C_{2alone}}{C_{1alone} + C_{2alone}} \right) Sig_{cmpd} \tag{8.11}$$

C_{1alone} and C_{2alone} represent the coherence thresholds of the two orthogonal motions when presented alone.

In the compound conditions, both the signal dots from component 1 and from component 2 were randomly distributed over the entire aperture. A previous comparison made by Meese & Harris (2001a) had confirmed that detection thresholds were similar irrespective of whether the signal dots from the two component motions are allocated to spatially distinct regions or to a single common region. Their finding gives confidence to the assumption that was made here, that there are no local spatial interactions between the signal dots of the two global motions.

Data were gathered from three observers (PAA, HRC, and AEF). Each carried out 12 experimental sessions (6 per condition) in a pseudorandom order. Prior to these sessions, observers had carried out 4 practice sessions (2 per condition) to familiarise themselves with the task and stimuli.

8.5 Results

Average summation indices (SI) are plotted for each observer and condition (fig.8.3-8.5); they were calculated using the following formula:

$$SI(dB) = 20 \cdot \log\left(\frac{C_{1alone}}{C_{1cmpd}} \right) \tag{8.12}$$

In all cases, plotted values represent the mean of six summation indices and the error bars represent the corresponding SEs. Inset in the figures 8.3-8.5 are the relative contributions of the two components to the compounds. These appeared to vary between observers. Also shown are the predicted amounts of summation for the four Minkowski metric exponents 1, 3, 4, and 5. In this case, where there are only two components, the corresponding levels of summation, expressed in dB, are 6, 2, 1.5, and 1.2 respectively. Exponents between 3 and 5 have previously been shown to yield levels of summation that are consistent with probability summation under the Quick Pooling model. Probability summation under SDT would also predict SIs much less than 6dB (Tyler & Chen, 2000).

Clearly, no condition showed levels of summation that approached linear. For observer PAA, summation levels for both conditions were most consistent with the predictions of probability summation. For observers HRC and AEF, the levels of summation for the *exp+clk-rot* condition lay between the lower bounds of probability summation and zero summation; for the *cntr+clk-rot* condition, the levels of summation appeared slightly negative, but clearly, in neither case would this negative summation be significant.

Consistent across all three observers was that the *cntr+clk-rot* condition showed less summation than the *exp+clk-rot* condition. Between observers however, there appeared to be differences in the general amounts of summation that took place. For both conditions, observer PAA showed the greatest levels of summation followed by HRC and then AEF.

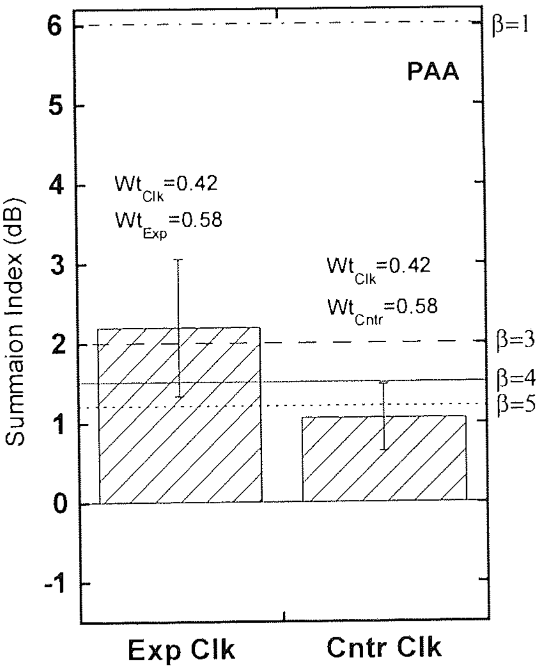


Fig.8.3. Showing data for observer PAA. Summation indices (from an average of 6 estimates) are plotted for both experimental conditions: expansion-clockwise rotation (*Exp Clk*) and contraction-clockwise rotation (*Cntr Clk*). The inset values denote the component weightings for each compound condition. Summation levels are shown for the Minkowski exponents, 1 (dot-dashed line), 3 (dashed line), 4 (solid line), and 5 (dotted line).

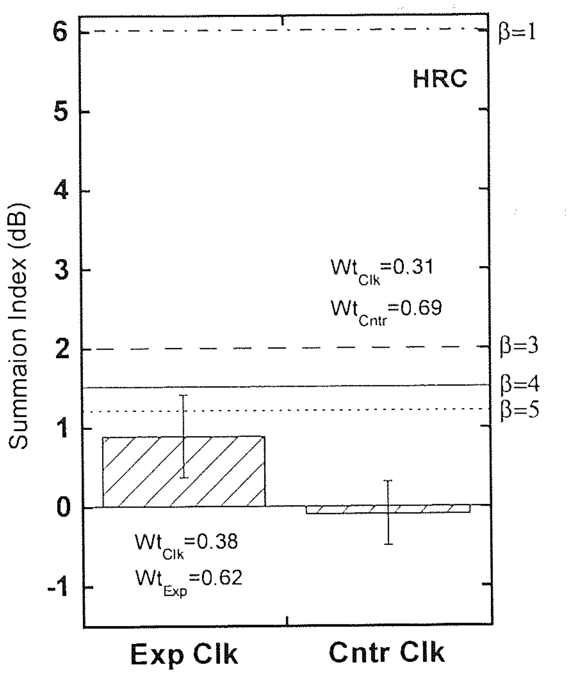


Fig.8.4. Showing data for observer HRC. Summation indices (from an average of 6 estimates) are plotted for both experimental conditions: expansion-clockwise rotation (*Exp Clk*) and contraction-clockwise rotation (*Cntr Clk*). The inset values denote the component weightings for each compound condition. Summation levels are shown for the Minkowski exponents, 1 (dot-dashed line), 3 (dashed line), 4 (solid line), and 5 (dotted line).

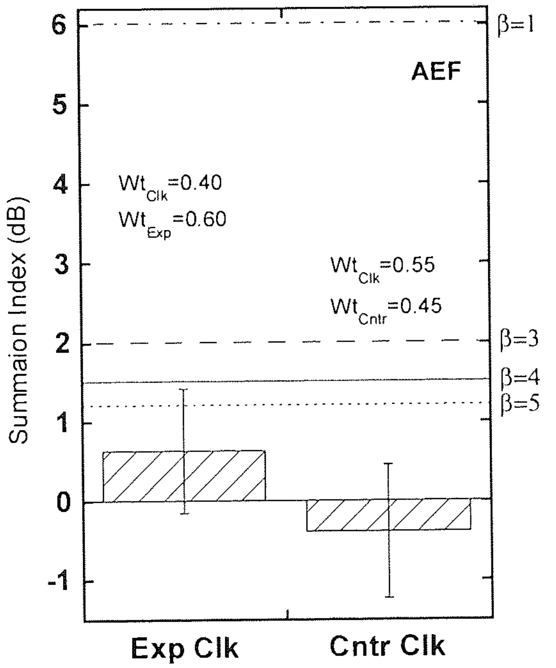


Fig.8.5. Showing data for observer AEF. Summation indices (from an average of 6 estimates) are plotted for both experimental conditions: expansion-clockwise rotation (*Exp Clk*) and contraction-clockwise rotation (*Cntr Clk*). The inset values denote the component weightings for each compound condition. Summation levels are shown for the Minkowski exponents, 1 (dot-dashed line), 3 (dashed line), 4 (solid line), and 5 (dotted line).

8.6 Discussion

The data from observer PAA appears to be best explained by probability summation. For the data of observers HRC and AEF the interpretation becomes more dependent upon condition. For the *exp+clk-rot* condition, summation levels seem to be equally indicative of probability summation and zero summation, while in the *cntr+clk-rot* condition they could be reflecting either zero summation or negative summation. Negative summation implies inhibitory interactions between the processing of global motions. Although Meese & Harris (2001b) have reported such inhibition for opposing signs of *def1*, they did not report it for any of the orthogonal motion pairs that were tested here (Meese & Harris, 2001a; 2001b; although they did not report a contraction-clockwise rotation condition). A more likely explanation for the results of this experiment relates to the effects of fatigue. Each experimental session was long (35-40min) and both naïve observers reported the task to be particularly arduous. Because the sessions were split into two stages (an initial component stage and a latter compound stage), it is possible that performance in the latter stage - in which the compound sensitivities were determined – was worse than in the first. If this were the case, then levels of summation would be underestimated and such fatigue effects could therefore be masking probability summation and sometimes be driving SI(dB)s to negative values.

Although evidence for probability summation was found for only a single observer, the existence of these performance advantages, even under conditions of restricted viewing, adds plausibility to the hypothesis that superimposed global motion components are processed in parallel with independent mechanisms. The absence of any clear summation with the two naïve observers is not inconsistent with the above hypothesis, even if fatigue had not been a factor. It may have been that at the decision stage independent mechanisms were not being monitored with a perfect efficiency.

Chapter Nine: Experiment 5 - Spatial summation of complex motions

9.1 Experiment 5a: Main experiment

9.1.1 Overview

In chapter 7 a study by Thornton & Gilden (2001) was discussed, in which evidence was provided, which favoured the idea that spatially discrete radial motions (of the same motion sign) are processed in parallel, albeit with the processing being subject to limitations. Whether these radial motions were being processed by independent mechanisms or, instead by common set of mechanisms was not fully explored in the paper. If the former were true, then the performance gains that were observed with increasing patch numbers would have been due to probabilistic advantages at the decision stage, whilst in the latter case they would have been derived from within-channel summation at an earlier perceptual stage. Although the researchers were unable to draw firm conclusions on the issue, there was some weak evidence against the former. However, the search paradigm used was not ideal for determining the magnitudes of performance gains, as the performance measure (RT) was free to confound with criterion level. A more appropriate measure, which does not confound with criterion, would have been coherence threshold.

The present chapter describes a number of subthreshold summation experiments, which use coherence measures to establish the existence of spatial parallelism, and its likely origin, for a number of stimulus types including a sign of radial motion (expansion). Experiment 4 of this thesis was a restricted viewing version of Meese & Harris' (2001b) subthreshold summation experiment. It aimed to test whether the summation that they had observed for orthogonal motion components had been due to parallel processing or a retrospective serial processing that was utilising visual persistence. The results favoured the former explanation and so afforded confidence that the summation of complex motions obtained with this paradigm was arising from parallel processing (at least for reasonably short display presentations).

9.1.2 Spatial summation experiment of Morrone et al (1995)

Previously, spatial summation experiments have been used to investigate the processing of single complex motions. In a series of experiments by Morrone and colleagues dot coherence thresholds were measured for three types of motion (radial; rotation; horizontal translation) under a number of configuration conditions (Morrone *et al*, 1995; Burr *et al*, 1998). In one key experiment (Morrone *et al*, 1995) the displays were constructed by dividing up a single circular patch, equally, into 16 sectors. A global motion transform (e.g. a clockwise rotation, a contraction, or a homogeneous left-to-right translation) would span this patch with its centre aligned to the midpoint of the patch. On any particular trail a number of the sectors would contain some mixture of signal and noise dots (designated here as the S+N sectors). Signal dots would move along trajectories consistent with the global motion transform while the noise dots would move in an incoherent manner.

In one set of conditions all 16 of the sector locations contained dots of an approximately equal density, irrespective of whether some of these sectors did not carry signal dots. Those not allocated signal dots were filled just with noise dots (designated here as the N sectors). During a single experimental session a number of these displays were tested; they differed in the number of S+N sectors that were present. In this experiment the displays could contain 1, 2, 4 or 8 S+N sectors and during a single session the proportion of signal dots presented in these sectors were staircased based upon the subjects' performance in a direction discrimination task. Threshold estimates for each display condition were obtained at the end of the session. In another set of conditions only the sectors that were designated to be S+N sectors contained dots; the other sectors were left blank (i.e. in these conditions there were no N sectors present).

Sensitivity estimates were made from the threshold estimates, with sensitivity being taken as the reciprocal of the proportion of signal dots present in the S+N sectors at coherence threshold (simplified to the expression $1+N/S$; where N and S denote the average number of noise and signal dots that are present at threshold).

Morrone and colleagues plotted their data as log sensitivity against log number of S+N sectors. They found that slopes of 1.0 could adequately describe the data from those display conditions in which the additional N sectors were present, and slopes of 0.5 could adequately describe the data from the display conditions in which no N sectors were present. When viewed in terms of the Minkowski summation formula of eqn. 8.6 these slope values correspond to the β values of 1.0 and 2.0 respectively. These are the same β values that are predicted by the within-channel summation model in which extrinsic stimulus noise dominates (section 8.2.1).

The authors made the following assumptions so that their data could be viewed in terms of the within-channel summation model. (1) The extrinsic stimulus noise has a variance that is proportional to the number of sectors that are present. (2) Any increases in the number of S+N sectors leads to proportionate increases in the total amount of signal in the display. If this were the case then eqn 8.3 (rewritten below) can be used to describe the data from those displays containing the additional N sectors, as the amount of extrinsic stimulus noise would have remained constant irrespective of the number of S+N sectors that were present. Consequently these display conditions can be referred to as the *constant-noise* conditions.

$$Sens_{cmpd} = n \cdot Sens_i \tag{9.1}$$

Where $Sens_{cmpd}$ is the sensitivity to the display, $Sens_i$ is the sensitivity to a single S+N patch and n is the number of S+N patches that are present in the display.

Eqn 8.4 (rewritten below) can be used to describe the data from the display conditions that did not contain the additional N sectors. In these the total variance of the extrinsic stimulus noise is assumed to increase proportionally with the number of S+N sectors. Providing the extrinsic stimulus noise dominates throughout, increasing the number of S+N sectors by a factor of n would lead to an n factor increase in the amount of signal but will lead only to a $\sqrt{n}$ factor increase in the amount of noise. Consequently this set of conditions can be referred to as the *varying-noise* conditions.

$$Sens_{cmpd} = \sqrt{n} \cdot Sens_i \tag{9.2}$$

If the slopes had been much shallower they would have been more consistent with typical predictions of probability summation (section 8.2.2). The good correspondence between their data and the predictions of within-channel summation supports the notion that specialised motion mechanisms exist, capable of pooling local motion signals not only across a sector but also across all 16 of the sectors.

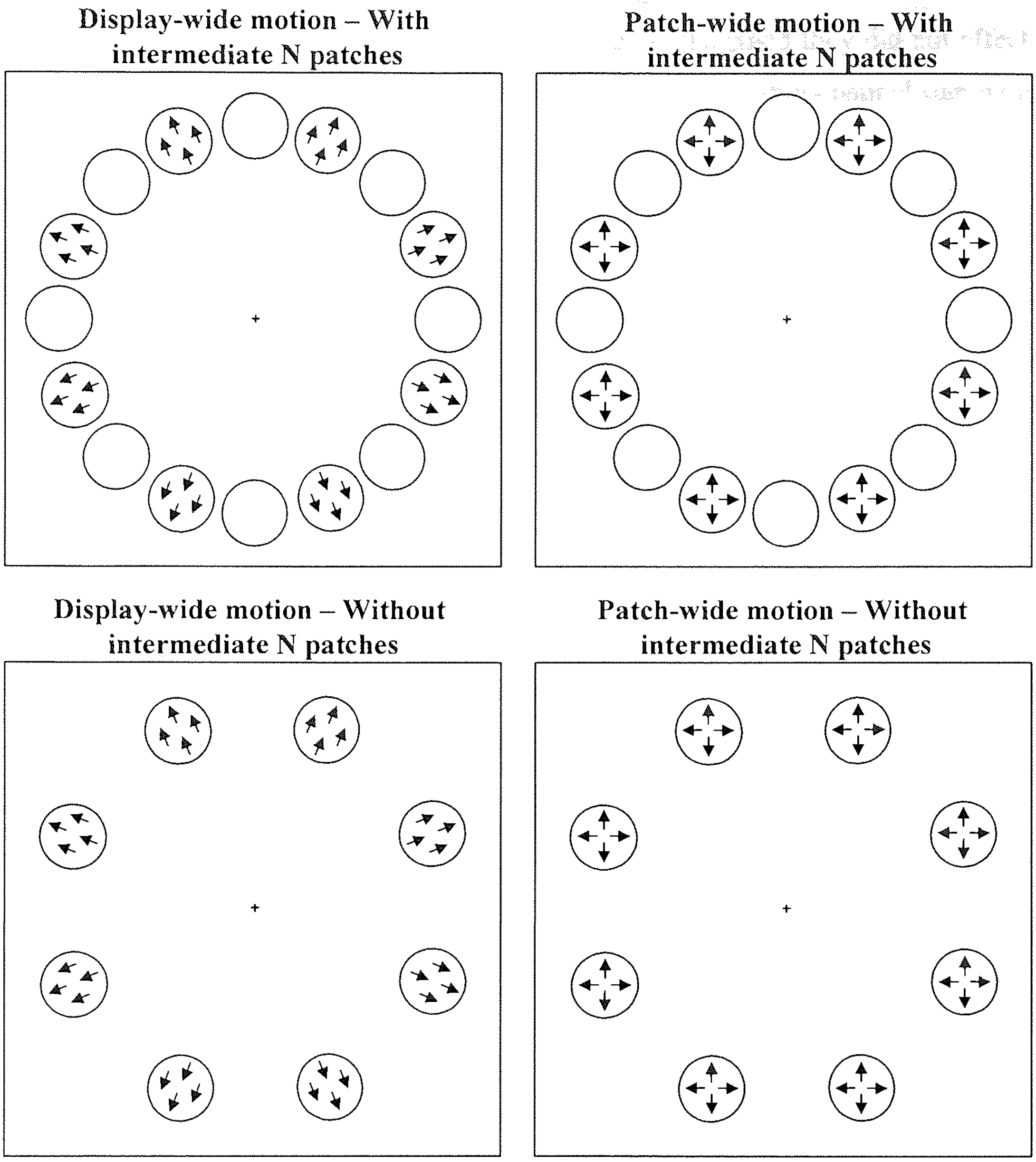
9.1.3 Stimulus and experimental design

Within this chapter, a similar experiment to the one described above was carried out in order to establish the magnitudes of spatial summation that are associated with multiple patches of motion, as opposed to the summation associated with a single motion spanning multiple sectors. To do this displays were constructed from discrete patches positioned on the circumference of a notional circle (see fig 9.1 for illustration). A total of 16 possible patch positions were available. As in the sector experiments of Morrone *et al* (1995) there were both constant-noise conditions and varying-noise conditions.

In the constant-noise conditions patches would appear in all 16 of the locations. On a certain trial some number of these patches would carry signal and noise dots. These will therefore be referred to as S+N patches. The signal dots of these patches followed trajectories that were consistent with motion transforms centred on, and restricted to each of the patches (*patch-wide* motion). Any patches that did not contain signal dots displayed only noise dots and will therefore be referred to as N patches. Irrespective of the type of patch, dot density was kept the same. As in the Morrone *et al* (1995) experiment the proportion of signal dots in the S+N patches was staircased, based upon observer performance, and estimates of coherence thresholds made for each display condition (i.e. displays containing different number of S+N patches) at the end of an experimental session. In the varying-noise conditions only the S+N patches were filled with dots, any remaining patch locations were left blank.

An additional group of display-wide motion conditions were used as control conditions. In these the signal dots from all the S+N patches would follow trajectories consistent with a single motion transform that extended over the entire array of patches and was centred to the origin of the notional circle. The two example displays depicted in the left column of fig.9.1 illustrate the resulting display-wide motion for expansion, in which 8 S+N patches are present. As with the patch-wide displays thresholds were determined for both the constant-noise and the varying-noise conditions. Both of these conditions were therefore similar to those used by Morrone *et al* (1995) and Burr et al (1998), except here the dots were restricted to patches and not sectors.

As in experiment 4 a temporal 2IFC detection paradigm was used (see section 2.8 for details of the generation of the random-dot displays and the method used to estimate coherence thresholds). In the signal intervals the displays containing S+N patches were presented. In the noise intervals only N patches were ever present. The number of these N patches was set so as to be the same as the number of patches that were present in the corresponding signal interval. In the noise intervals of the constant-noise conditions 16 N patches would always be present, while in the varying-noise conditions the number was equal to the number of S+N patches in the corresponding signal interval.



Display-wide motion – With intermediate N patches

Patch-wide motion – With intermediate N patches

Display-wide motion – Without intermediate N patches

Patch-wide motion – Without intermediate N patches

Fig.9.1. The four primary display conditions for experiment 5a. Each circular cluster of patches illustrates a particular display configuration for expansive motion. Arrows represent the approximate motion directions taken on by signal dots (in the S+N patches). Outlined patches containing no signal denote noise only (N) patches. The two configurations on the left show the patch motions that result from applying a display-wide motion transform. The two configurations on the right show the resulting motions from applying a patch-wide motion transform.

9.1.4 Predictions

The findings of Morrone and colleagues (see section 9.1.2) have previously shown that display-wide conditions could yield summation levels that are consistent with within-channel summation, in which the extrinsic stimulus noise dominates (see eqns 9.1 & 9.2). In the present experiment the same levels of summation might be expected (linear summation for the constant-noise conditions and quadratic summation for the varying-noise conditions) as the only major difference between the displays is the shape of the apertures that were used: sectors for the previous experiments and patches for the present experiment. An implication of the data from the varying-noise conditions of Morrone et al (1995) was that even when the aperture borders could be discerned they did not effect the operation of the putative mechanisms that were underlying the within-channel summation.

With regards to the patch-wide conditions there are three main possibilities. Firstly, no summation may be observed between the S+N patches, indicating that only a single pattern of motion can be processed at time, either due to attentional limits or the absence of specialised neural mechanisms ($\beta = \infty$). A second possibility is that there are significant levels of summation, similar to those predicted by within-channel summation ($\beta = 1.0$ for constant-noise; $\beta = 2.0$ for varying-noise). This would suggest that the motions of the different patches are processed by a common set of mechanisms. The third possibility is that there are intermediate levels of summation (typically $5.0 > \beta > 3.0$). Such a finding would be consistent with models of probability summation in which each patch is processed by independent sets of mechanisms. A strict HTT model of probability summation, such as the standard Quick Pooling model (described in section 8.2.2), would predict similar β values for both the constant-noise and varying-noise conditions. In a signal detection theory framework, however, the constant-noise conditions would predict β estimates to be slightly higher in the varying-noise conditions.

9.1.5 Stimulus details

Viewing was monocular and from a distance of 114cm. The notional circle that the stimulus patches were arranged upon had a radius of 3.74° and a fixation spot was positioned at the origin. In each interval, the orientation of the circle was set randomly. The displayed patches were allocated so as to ensure a maximal spacing between them. Each patch had a diameter of 1.38° and a central region of diameter of 0.23° (no dots were present within this central region). Dot luminances' were not ramped at the boundaries, as had been the case in the experiment 4, instead boundaries were abrupt. This led to well-defined impressions of aperture boundary when the static frames of the movies were viewed. A dot density of 20dots/degree^2 was used. This high dot density was necessary to ensure that a sufficient number of dots were available for staircasing. All dots travelled at speeds of 0.5°sec^{-1}. If, during its trajectory, a dot breached one of the patch boundaries, it would be randomly repositioned within the patch. Conversely, if a dot entered the patch during the course of its trajectory, a randomly selected, but corresponding dot type (noise or signal) would be removed from the patch. These two measures ensured that dot density remained constant throughout the RDK movies.

Limited-lifetime dots were constructed from 6 instances of their trajectories. On each movie image, 16.7% of the dots disappeared and reappeared in new locations. S+N and N intervals each lasted for 450ms and permitted the presentation of movie sequences that were 6 images long (a single image was presented for 9 frames). A 300ms ISI separated the two intervals of single trial.

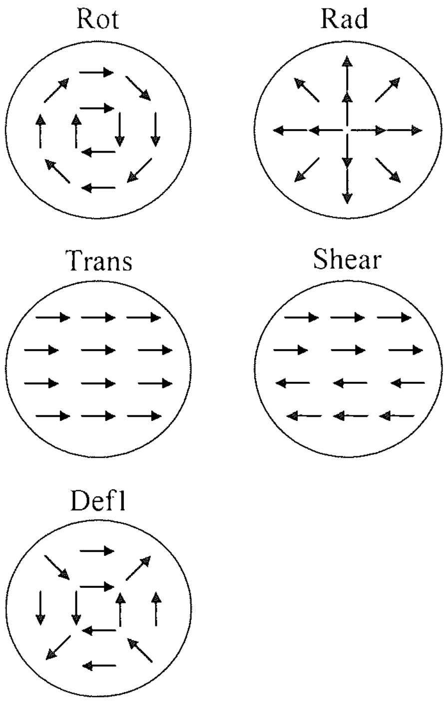


Fig.9.2. Schematics illustrating all five of the motion types used in the spatial summation experiments 5a-e. Note that, because of the removal of the speed gradients, the shear motion gives rise to an abrupt change in direction about the central horizontal axis.

Absence of speed gradients

To permit meaningful comparison between the patch-wide and display-wide conditions, speed gradients were removed. This ensured that the average distribution of dot velocities over the entire display remained the same for both sets of display conditions. Because the patch diameters were reasonably small, the absence of a speed gradient was not particularly conspicuous in any of the displays.

Display configuration

By positioning the stimulus patches on the circumference of a notional circle around fixation any variation in sensitivity arising from variations in acuity, due to viewing eccentricity, should have been lessened. In the visual search experiments (experiments 1-3b) an iso-eccentric positioning of the patches could not be achieved because patch sizes that were necessary to afford compelling impressions of 3-dimenionality were too large to allow it.

The distances between the S+N patches within a signal interval was made equal for the purposes of (a) encouraging the observer to adopt a strategy of attending to the whole display, (b) discouraging observers from initiating eye-movements, and (c) to be consistent with the configurations of Morrone *et al* (1995). If performances were derived from within-channel summation then the motion signals from the various patches would presumably be summed by mechanisms with RFs covering the entire display, in which case, the relative positioning of the patches would be expected to be irrelevant. Burr *et al* (1998) did investigate summation in varying-noise (N) conditions for two other methods of area increase, in addition to the increasing sectors method ('windmill' configurations). In these experiments, the area containing S+N was increased, either by increasing the display radius ('circles' configurations), or increasing the sector angle of a single contiguous sector ('sector' configurations). They concluded that levels of summation were approximately the same, irrespective of how area was increased. However, data for all three of the display conditions were obtained from only a single observer, and for this observer there was a tendency for greater summation in the 'sector' and 'circles' conditions than in the 'windmill' conditions. For example, best-fit Minkowski exponents (β) for the 'windmill' conditions were, 2.50 (radial), 2.04 (rotation), and 2.63 (translation), while for the 'circles' conditions they decreased to, 1.75, 1.72, and 1.96 respectively (note that the expected β values were 2.0).

Inter-patch cues in the patch-wide conditions

For the *rot* and *rot+trans* conditions of experiment 1 of this thesis, it was speculated that the efficient target detection had been based not upon the sign of the rotation components in themselves but upon spatial interactions between these motions from neighbouring patches (section 4.4.2). Might similar cues contribute to detection of coherent motion? It is possible that the putative interactions seen in experiment 1 only occurs for rotation stimuli that contain motions with levels of coherence that are significantly higher than threshold. However, to lessen the chance of any such effects within experiment 5a, a minimum distance that was greater than that of a single patch diameter always separated S+N patches. Moreover, in the constant-noise conditions S+N patches were separated by at least one N patch. It seems likely that the inclusion of these intermediate patches of noise would be capable of disrupting any interaction. Indeed, the lack of efficient search for the rotational motions of Thornton & Gilden's (2001) study may have resulted from their inclusion of a textured surround preventing inter-patch cues.

9.1.6 Procedure

Separate sessions were carried out for the different configuration and motion conditions. Within a single session, threshold estimates were made for each of the S+N patch numbers that were being tested. For observer PAA these were 1, 2, 4, and 8. For observer HRC they were 2, 4, and 8. All staircases were interleaved ensuring the observer was uncertain as to the exact number of S+N patches that were to appear on any particular trial. The recording of responses took place on the last 8 staircase reversals.

Data were gathered from two observers (PAA and HRC). Both carried out sessions for the following three motion types: clockwise rotation (*Rot*), positive radial (*Exp*), and rightward translation (*Trans*). PAA carried out additional sessions for deformation (*Def1*), and positive horizontal shear (*Shear*). HRC gathered data for the constant-noise conditions only, while PAA gathered data for both constant-noise and varying-noise conditions (for schematics of these five stimulus types see fig.9.2).

Thresholds were averaged from 5-6 estimates, which were obtained from experimental sessions carried out in a pseudorandom order. Prior to these experimental sessions, observers carried out 3 practice sessions on each motion condition to reduce thresholds to consistent levels.

9.1.7 Results

Figs 9.3-9.6 show plots of average sensitivity as a function of S+N patch number on double-log axes. Because coherence thresholds are measured as a proportion of signal dots, sensitivity can be expressed as $1+N/S$, in which N and S denote the average number of signal and noise dots that are present at threshold. Fig. 9.3 describes data for varying-noise (without intermediate N patch) conditions. Figures 9.4 and 9.5 describe data for constant-noise (with intermediate N patch) conditions.

Where appropriate, both the patch-wide and display-wide conditions have been included in the plots. Clearly, displays for patch-wide conditions of translation would be identical to those of the display-wide conditions. Many of the deformation (*Def1*) conditions proved to be too hard to determine coherence thresholds for. Even after extensive practice *Defl* thresholds could only be reliably estimated for the 2, 4, and 8 S+N patch numbers of the display-wide varying-noise condition.

Best-fit Minkowski exponents (β) were determined for each display and motion condition and are included in the plots (figs.9.3-9.5); these correspond to the reciprocal of the slope magnitude. Sensitivities tended to increase monotonically with S+N patch number under the varying-noise condition. Furthermore, to a first approximation, these increases are well described by the Minkowski summation formula (eqn.8.6). The best-fit β exponents are largely dependent upon motion condition, and for *Shear* transforms, they were also dependent upon display condition: little to no summation was present for the patch-wide conditions while, for the display-wide conditions, summation was closer to quadratic. Conversely, both *Rot* and *Exp* motions yielded similar levels of summation and general sensitivity for their patch-wide and display-wide configurations. The correspondence between the patch-wide and display-wide fits is particularly compelling for *Exp*.

Data for constant-noise conditions were available from two observers. Consistent with the predictions of within-channel summation, but not with the predictions of probability summation with a high threshold, summation levels were greater in the constant-noise conditions than in the varying-noise conditions. The addition of noise to the displays decreased sensitivity in all cases. As with the varying-noise conditions, the data are well described by the Minkowski summation formula. For the *Rot* and *Rad* conditions of both observers, the patch-wide and display-wide configurations yielded similar levels of summation. Between observers, the levels of summation for *trans* were almost identical. However, observer differences were apparent in the summation levels of the *Rot* and *Rad* conditions: observer HRC exhibited greater general levels of summation than PAA.

For both observers in the constant-noise conditions, summation levels were greater than those predicted by probability summation (typically β coefficients of between 3 and 5) indicating within-channel summation. However, the levels of summation fell slightly short of those expected from an ideal integrator (linear summation in constant noise conditions and quadratic or greater in the varying noise conditions) and those previously reported from the sector summation experiments of Morrone *et al* (1995). To investigate this issue further, another observer (KAM) was tested with *Rot*, *Rad*, and *Trans* but only for the global conditions. The following S+N patch numbers were tested: 2, 4, 8 and 16. Data from KAM are plotted in fig.9.6. This observer showed greater levels of summation for

Rot and *Exp* conditions with constant-noise displays; the *Rot* varying-noise condition also showed greater summation. However, compared to observers PAA and HRC, less summation was shown for the *trans* conditions.

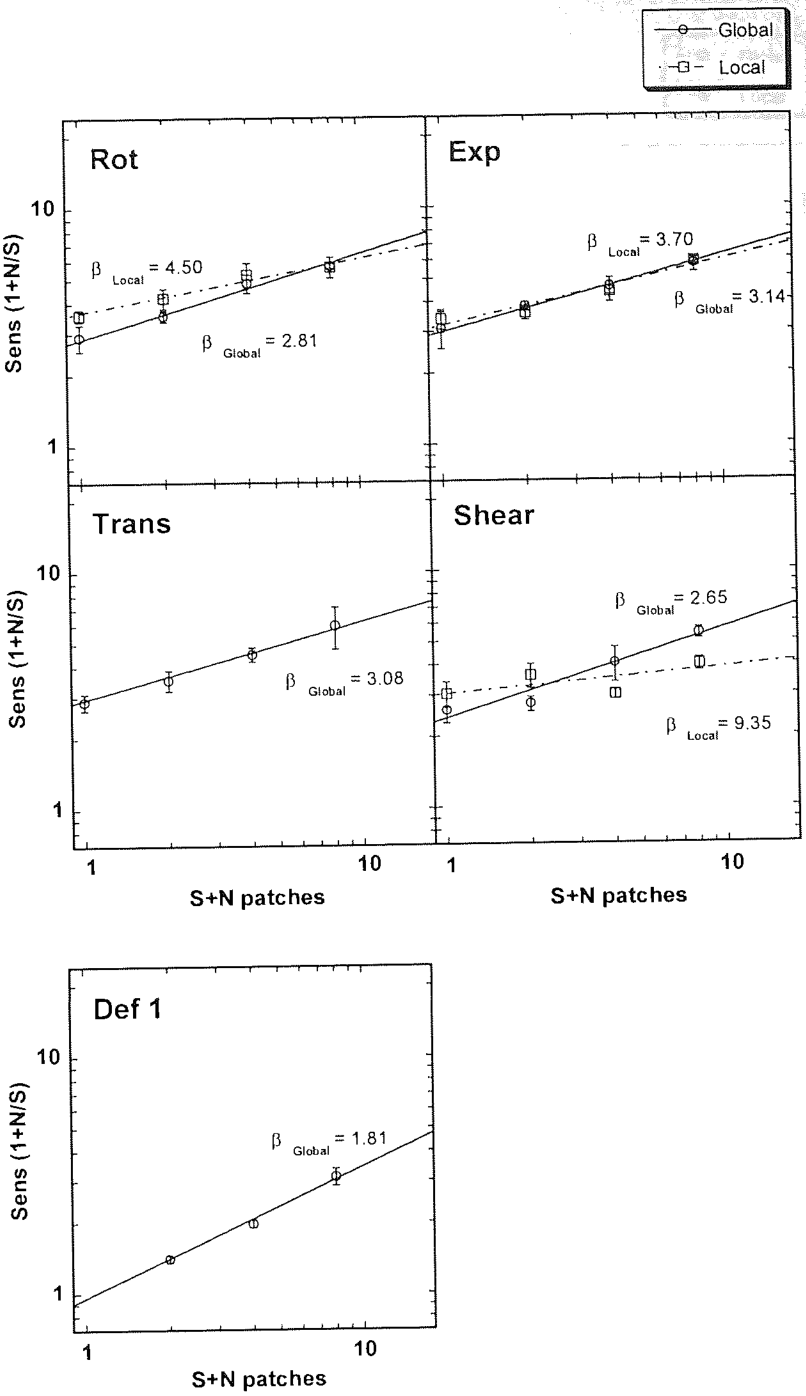


Fig. 9.3. Observer PAA. Data are shown for the varying-noise (with intermediate N patch) conditions. Where appropriate, plots show data for both display-wide and patch-wide configurations. The curves represent the best fits to the Minkowski summation formula, where β and $Sens_i$ were free to vary (see eqn.8.6). Best-fit β exponents are inset (β_{Global} and β_{Local} denote display-wide and patch-wide conditions). Each data point represents the mean of 5-6 threshold estimates; error bars represent the corresponding SEs.

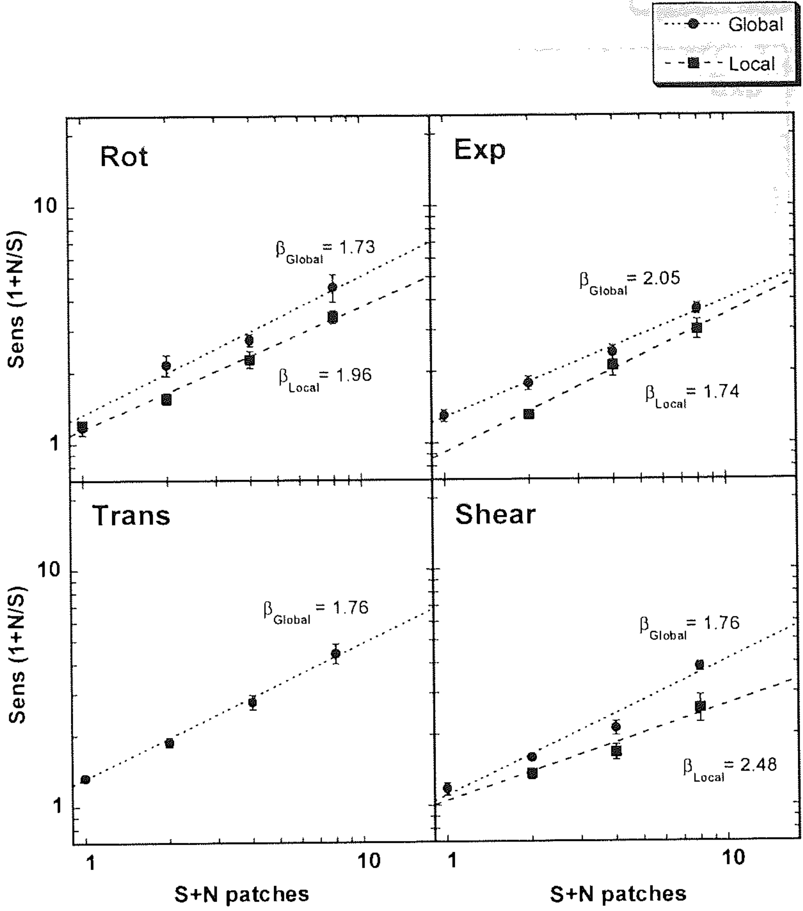


Fig. 9.4. Results for observer PAA. As for fig.9.3 except that data are shown for constant-noise (with intermediate N patches) conditions.

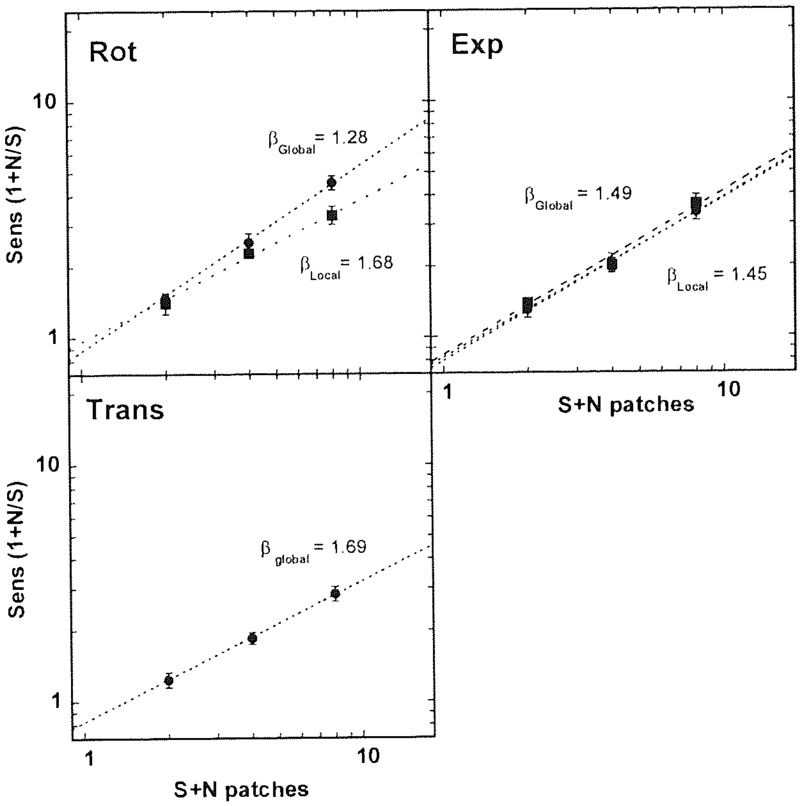


Fig. 9.5. Observer HRC. Same as above: data are shown for constant-noise (with intermediate N patches) conditions.

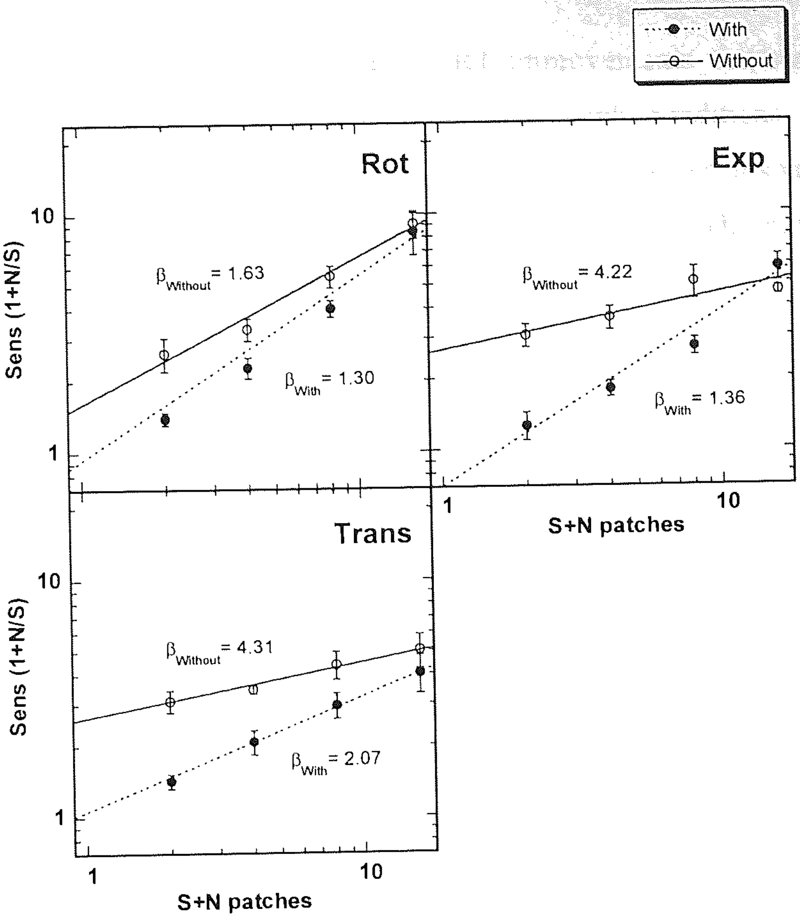


Fig.9.6. Results for observer KAM. Data are shown for display-wide configurations. Plots show data for both constant-noise and varying-noise conditions. β_{With} and $\beta_{Without}$ values denote the β exponents for the constant- and varying-noise conditions respectively.

9.1.8 Discussion

This experiment showed that for both the display-wide (single motion transform) and the patch-wide (multiple motion transforms) displays, there were reliable improvements in motion sensitivity as the number of S+N patches increased. These results suggest that, for the patch-wide conditions, in which signal dots adhered to one of many spatially discrete motion transforms, the observers were capable of encoding multiple representations of motion. For rotation and expansion, the improvements in sensitivity (i.e. levels of summation) were remarkably similar for the display-wide and patch-wide conditions. Furthermore, because the general sensitivities were also similar, it seems unlikely that the improvements in the patch-wide conditions can be attributed to the summing of the local motions from certain patch regions[1]. Inspection of summation levels for the patch-wide constant-noise conditions suggested that performance improvements were greater than those expected from probability summation, implying some form of within-channel

---

[1] As patches were arranged on the circumference of a notional circle, in the *Rad* and *Rot* patch-wide conditions, a pooling of dot motions from just the more outer regions of the patch could lead to approximations of the display-wide versions. Such a selective pooling strategy was not, however, in concordance with the subjective experiences of the two observers.

summation. It is therefore speculated that the RT improvements that were observed by Thornton & Gilden (2001) in their "pure target" search conditions (for expansion, contraction, and translation) also arose from a summation of signals at some stage prior to decision-making. In contrast to the present experiment, however, Thornton & Gilden did not report performance improvements for displays containing additional patches of rotation.

It is acknowledged, however, that in addition to differences between the task of Thornton & Gilden's research and the studies described here, their stimuli also differed in a number of ways; of particular significance might be their use of stimuli with suprathreshold coherence levels.

For the one observer (PAA) that did obtain estimates of sensitivity for the patch-wide varying-noise conditions, summation levels were less than those expected for perfect within-channel summation. This finding is consistent with the hypothesis that the processing of multiple motion representations is dependent upon some limited attentional resource. As each patch was reasonably well demarcated by dot presence, observers might have been capable of dividing attention between the displayed patches. Lower patch numbers, would consequently have meant that each patch would be devoted larger amounts of attention than in sessions with higher numbers of patches. In these cases, sensitivity to the individual patches would, consequently, have been higher. This division of a limited attentional resource would therefore have the general effect of decreasing the observed levels of spatial summation. Conversely, in the constant-noise conditions, attention would have always been divided between the 16 patches, and therefore sensitivity to individual patches would not have varied.

The tendency for summation levels to fall short of those expected from perfect within-channel summation is not, however, exclusive to varying-noise patch-wide conditions, it was also apparent for all other conditions. Clearly, for the display-wide conditions, these results conflict with the findings of Morrone *et al* (1995). As the notional circle, to which the patches had been allocated, had been randomly orientated on every interval, the possibility that this less than linear summation was due to differences in the weightings of individual detector outputs can be ruled out (Wilson & Wilkinson, 1998).

Meese & Harris (2001b) have previously reported less than perfect area summation for expansion, rotation and deformation flows. Discrepancies for the levels of area summation with static paired dot dipoles (Glass patterns) are also apparent (Wilson *et al*, 1997; Wilson & Wilkinson, 1998; Kurki *et al*, 2003; Dakin & Bex, 2002; 2003). Moreover, for the conditions of Burr et al (1998) that most closely match the display-wide conditions of the present experiment, summation levels were also found to vary both between observers and between motion types[2]. However, the β exponents that they observed in the constant- and varying-noise conditions only ever reached values as high as ~1.3 and ~3.0 respectively[3].

To investigate whether the tendency for high β exponents is the result of differences in experiment or stimulus design, summation levels were established for the display-wide conditions in a number of control experiments. These addressed some of the differences between the present experiment (experiment 3a) and the ones of Morrone *et al* (1995) and Burr *et al* (1998). The following three differences were investigated: viewing condition (binocular/monocular), task type, and noise type. A sector (rather than patch) version of the experiment was also carried out to replicate the main experiment of Morrone et al (1995). It is acknowledged that there are numerous other differences, but these were considered less worthy of investigation. For example, Morrone and co-workers used both black and white dots[4] of a relatively low density, whereas here white dots at a relatively high density were used. There seems little theoretical reason, however, to suspect that the use of black and white dots would affect threshold levels. Signal and noise dots of either contrast polarity have been shown to be similarly effective (Edwards & Badcock, 1994), suggesting that local motion signals are rectified prior to summation. Similarly, the difference in dot density alone is unlikely to have had any appreciable effect (Barlow & Tripathy, 1997).

9.2 Experiment 5b: binocular viewing

Viewing in both the Morrone *et al* (1995) and Burr *et al* (1998) studies was binocular. In the experiment described here, however, viewing was monocular. This provided consistency with the viewing conditions in experiment 1, where performance could, conceivably, for certain conditions, have been based upon 3-D interpretations. Here, binocular viewing was used to investigate the possibility that ideal area summation is contingent upon a pooling of signals that are derived from a dichoptic processing of local motions.

---

[2] Burr *et al*'s (1998) closest matching configurations were constructed from near square patches rather than circular patches. In the constant-noise conditions the patches would form a perfectly contiguous annulus.

[3] Only their smallest annulus diameter of 4.5° is considered.

[4] More specifically, their dots were generated from small Gaussian patches.

9.2.1 Stimuli and procedure

The stimuli and task were as in experiment 3a with the exception that (a) viewing was binocular, (b) the computer and display equipment had changed, and (c) as a consequence of the latter, the VD was changed from 114cm to 154cm. A single observer (PAA) replicated data collecting for the display-wide conditions of the main experiment. The *Defl* motion type was omitted because, in the earlier experiment, threshold estimates could only be made for three of its eight conditions.

9.2.2 Results

Fig.9.7 shows the data for the varying-noise conditions, and fig.9.8 the data for the constant-noise conditions. Appropriate data from experiment 5a have been included to enable easy comparison between the monocular and binocular conditions. (Data for all of these control experiments will be displayed on plots with axes matching those of fig.9.3-9.6, again to permit easy comparison between all conditions.)

Both constant-noise and varying-noise conditions for all four of the motion types show very similar levels of summation for monocular and binocular viewing conditions. There is a possible tendency for β exponents to be slightly higher in the *Trans* and *Shear* conditions but in terms of the fitted slope magnitudes, these differences are slight. Sensitivities are also remarkably similar, considering that there was an intervening period of five months between this and experiment 5a.

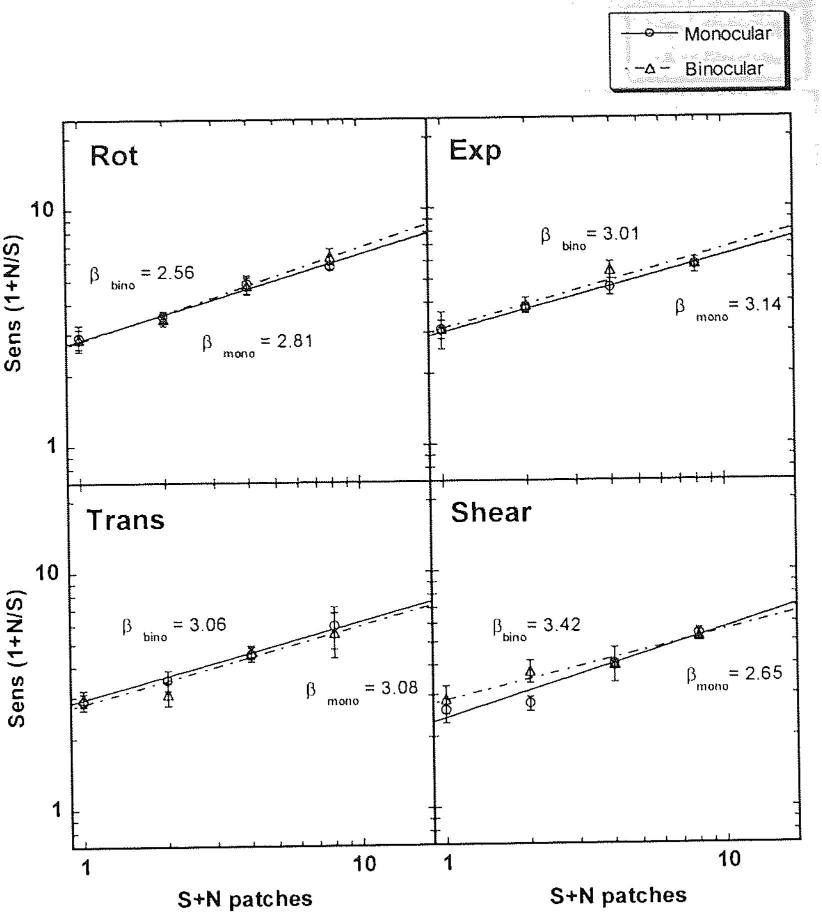


Fig. 9.7. Results for observer PAA. Data are shown for the display-wide varying-noise (without intermediate N patch) conditions. Within a single graph, data have been plotted from both the present experiment (binocular viewing) and the previous main experiment (monocular viewing). As in Figs.9.3-9.6 the curves represent best fits to the Minkowski summation formula, where β and $Sens_i$ were free to vary. Best-fit β exponents are inset. β_{bino} and β_{mono} denote exponent estimates for the present and previous experiments respectively.

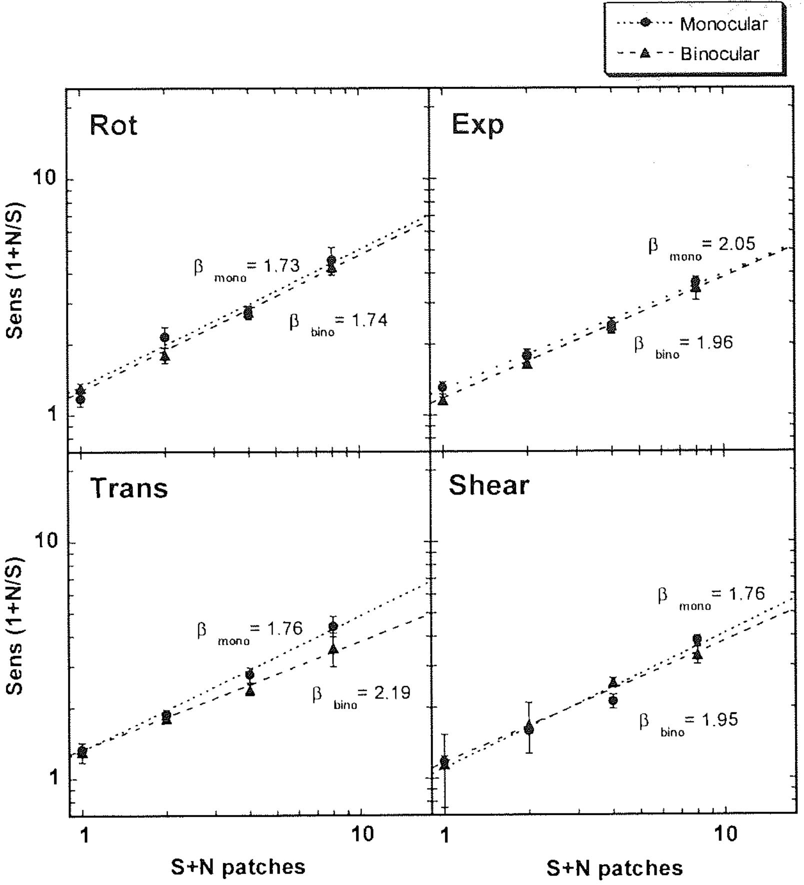


Fig. 9.8. Results for observer PAA. Same as above except that data are shown for constant-noise (with intermediate N patches) conditions.

			Global Transform								
			Varying Noise					Constant Noise			
			Rot	Exp	Trans	Shear	Defl	Rot	Exp	Trans	Shear
Exp 5A	*Main*	HRC						1.28 (1.00-1.78)	1.49 (1.15-2.09)	1.69 (1.34-2.27)	
		KAM	1.63 (1.28-2.24)	4.22 (2.65-10.33)	4.31 (2.67-11.20)			1.30 (0.96-2.00)	1.36 (1.12-1.73)	2.07 (1.52-3.24)	
		PAA	2.81 (2.16-4.05)	3.14 (2.20-5.51)	3.08 (2.10-5.83)	2.65 (1.92-4.28)	1.81 (1.53-2.21)	1.73 (1.36-2.38)	2.05 (1.75-2.46)	1.76 (1.52-2.10)	1.76 (1.49-2.15)
Exp 5B	*Binocular viewing*	PAA	2.56 (2.01-3.52)	3.01 (2.28-4.42)	3.06 (2.34-4.41)	3.42 (2.41-5.88)		1.74 (1.50-2.07)	1.96 (1.60-2.52)	2.19 (1.70-3.07)	1.95 (1.57-2.58)
Exp 5C	*Direction Discrimination*	PAA	1.83 (1.47-2.42)	3.36 (2.37-5.79)	3.48 (2.37-6.58)	2.25 (1.71-3.32)		1.39 (1.19-1.68)	1.55 (1.29-1.93)	1.76 (1.48-2.17)	1.72 (1.40-2.22)
Exp 5D	*Dynamic-6*	PAA			6.01 (3.58-18.60)					1.35 (1.11-1.72)	
	Dynamic-2	PAA			3.56 (2.40-6.90)					1.30 (1.11-1.57)	
	Twinkle-2	PAA			3.28 (2.28-5.82)					1.57 (1.31-1.97)	
Exp 5E	*Sectors*	PAA	2.06 (1.64-2.77)	3.64 (2.19-10.72)	3.85 (2.56-7.77)			1.16 (1.00-1.38)	1.43 (1.20-1.75)	1.36 (1.16-1.63)	
		JBS	1.98 (1.50-2.92)	2.53 (1.69-5.05)	5.82 (2.74- ∞)			1.05 (0.89-1.26)	1.19 (0.99-1.50)	1.32 (1.18-1.50)	

Table. 9.1 A summary table showing the β values and 95% confidence intervals for the display-wide conditions of the various experiments described in this chapter.

			Local Transform					
			Varying Noise			Constant Noise		
			Rot	**Exp**	**Shear**	**Rot**	**Exp**	**Shear**
Exp 5A	*Main*	**HRC**				1.68 (1.30-2.35)	1.45 (1.14-1.98)	
		PAA	4.50 (2.95-9.52)	3.70 (2.68-5.99)	9.35 (4.25-∞)	1.96 (1.64-2.44)	1.74 (1.36-2.39)	2.48 (1.68-4.72)

Table. 9.2 A summary table showing the β values and 95% confidence intervals for the patch-wide conditions of experiment 5a.

9.3 Experiment 5c: 1IFC (direction discrimination) task

Morrone and colleagues used a single interval (1IFC) direction discrimination task rather than a 2IFC detection task in their area summation studies. Here, observers had to indicate, which, out of two possible motion directions the signal dots were travelling in. For example, a choice of contraction and expansion would have been available for the radial conditions. The choice of the 2IFC detection task stemmed from the observation that for *Defl* motions, comprehension of its motion directions is less intuitive than for *Rot*, *Rad*, *Trans*, and *Shear* motions.

In general, it may be of little surprise that different tasks yield different estimates of sensitivity. However, typically there is little reason to expect differences in summation levels between different tasks. The use of coherent motion stimuli in the 1IFC direction discrimination task and the present 2IFC detection task might be an exception. Geisler (1999) has found evidence to support the hypothesis that, under certain circumstances, motion detection can be aided by static orientation cues that are derived from "motion streaks". Geisler has shown that line noise orientated parallel to the direction of motion is more effective at masking the motion than perpendicular line noise. If such orientation cues can aid in the detection of coherent motion, they may do so by stimulating specialised global form mechanisms rather than global motion mechanisms. For example, the detection of rotational motion may, under certain circumstances, stem from stimulation of the putative circular mechanisms in V4 (as posited by Wilson & Wilkinson, 1998) rather than specialised mechanisms for rotation. However, it is conceivable that these form based cues would be least effective in the direction discrimination task, simply because, in this task, orientated streaking alone would provide insufficient cuing of direction.

9.3.1 Stimuli and procedure

The stimuli and task were as in the main experiment (5a) of this chapter with the exceptions that (a) the task changed from a temporal 2IFC detection task to a 1IFC direction discrimination task, (b) the computer and display equipment had changed, and (c) as a consequence of the latter, the VD was changed from 114cm to 154cm. A single observer (PAA) replicated data collecting from the display-wide conditions of experiment 5a and as in experiment 5b the *Def1* motion type was left out.

In the previous 2IFC detection experiments only one direction sign was tested, for each motion type: *Rot* (clockwise), *Rad* (positive), *Trans* (rightward), and *Shear* (positive). In this 1IFC task, the sign of motion in each trial was chosen randomly, with an equal probability for each direction. The observers' task was to indicate motion sign using the left/right buttons of a PC mouse. Control of the staircases and estimation of the threshold remained the same as before.

9.3.2 Results

Fig.9.9 shows the data for varying-noise conditions, and fig.9.10 shows data for constant-noise conditions. Corresponding data from experiment 5a have also been plotted to assist comparison between the two types of task. For all varying-noise conditions, measured sensitivities were higher in the direction discrimination task than in the 2IFC detection task. Three out of the four motion conditions showed no apparent differences in the summation magnitudes. The direction discrimination data for the *Rot* condition did yield a slope x1.5 steeper than that of the 2IFC detection data. This resulted in an β exponent of 1.83, which is consistent with the predictions of perfect within-channel summation. The constant-noise conditions for all four motions showed similar levels of sensitivity and summation between task types.

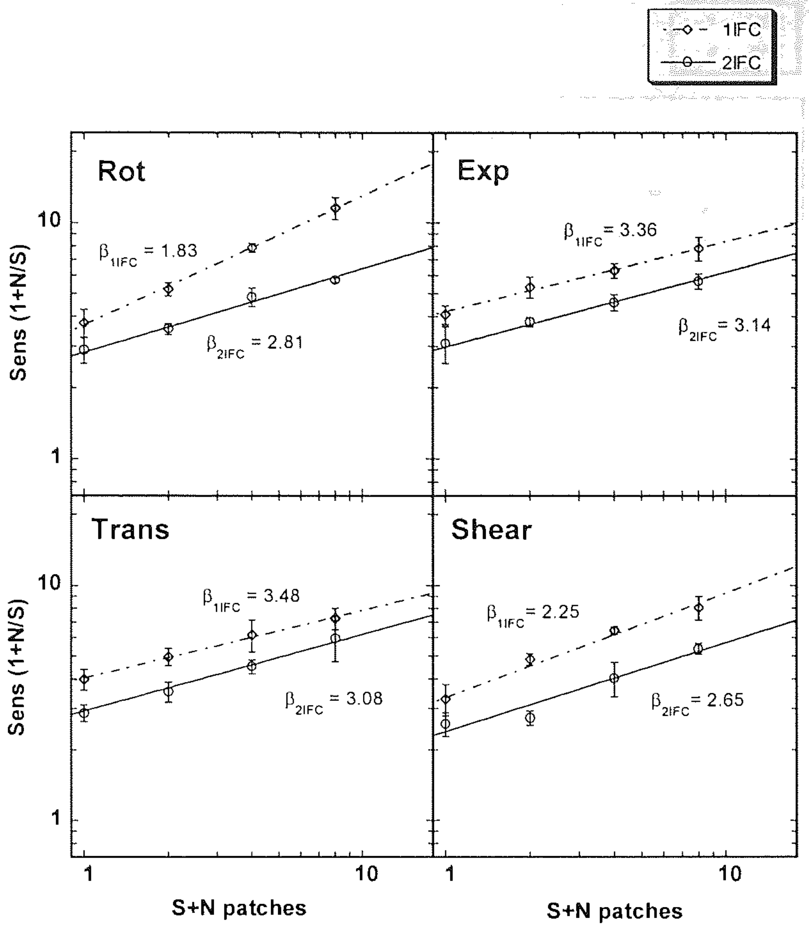


Fig. 9.9. Data from observer PAA for the 1IFC and 2IFC display-wide varying-noise conditions. Within a single panel, data are plotted from both the present and main area summation experiments. The curves represent Minkowski model fits, in which β and $Sens_i$ were free to vary. The best-fit β exponents are inset. β_{1IFC} and β_{2IFC} denote exponent estimates for the two task types.

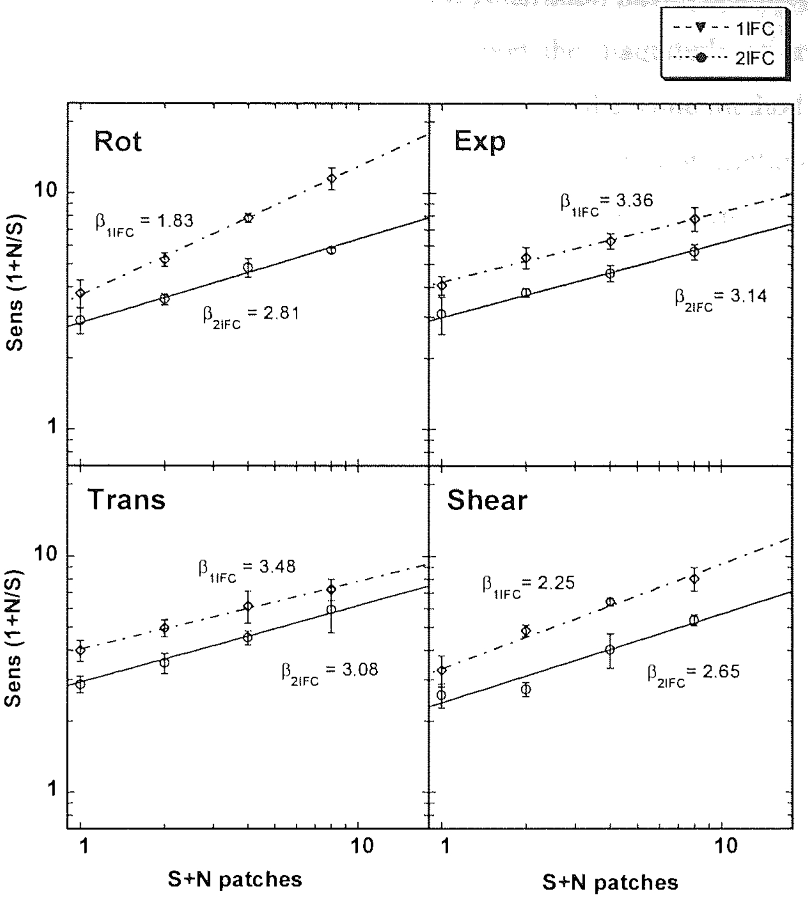


Fig. 9.10.. Data from observer PAA for the display-wide, constant-noise conditions, of the 1IFC and 2IFC experiments. Other details are the same as above.

9.4 Experiment 5d: Comparison of noise types

The method of noise and signal dot generation in the studies of Morrone *et al* (1995) and Burr *et al* (1998) differed slightly to the method employed in the previous experiments. Firstly, in experiments 5a-c dots were used with a maximum lifetime of 6 movie images, where each image corresponded one instance of a straight-line trajectory. Morrone *et al* generated their signal dots in a similar fashion with the exception that dot lifetimes were only ever two movie images long, unless the trajectory had determined the second image to be outside of the aperture, in which case it would instead be randomly reallocated to somewhere within the visible aperture (this ensured constant dot densities). Secondly, noise dots in the study of Morrone *et al* did not take on dot trajectories with the same dot displacement as their signal dots; instead, they were randomly allocated on each movie frame. Scase et al (1996) has termed this noise "random position", and the noise used in experiments here, "random direction".

In this experiment, three different methods of dot generation have been used to investigate the possible effects of noise generation method upon the magnitude of area summation. The first, which will be referred to as *Dynamic-6* noise, is the same method as that used in experiments 5a-c ("random direction" noise with maximum dot lifetimes of 6 movie frames). The second is the same method that was used by Morrone *et al* ("random position" noise with a maximum of 2 movie frames) and is referred to here as *Twinkle-2* noise. The third is an intermediate to the two, in which the noise dots are "random direction" but all dot lifetimes have been limited to 2 movie frames. In the following discussion, differences in the relationships of signal dot proportion to actual signal level are considered for the three methods of dot generation. These differences are most easily described through the separate comparison of *Dynamic-6* and *Twinkle-2* noise conditions to the *Dynamic-2* noise condition.

Dynamic-6 versus *Dynamic-2*

In comparing these two "random direction" conditions, the first thing to note is that for a given dot density and signal dot percentage, the *Dynamic-6* condition will contain a greater amount of 'actual' signal than the *Dynamic-2* condition. This arises from the fact that the longer dot lifetimes produce greater numbers of signal dot correspondences by virtue of their lower dot "birth" and "death" rates[5]. However, 'actual' signal level will always rise proportionately with signal dot percentage, irrespective of dot lifetime.

As noise dot directions are picked randomly from a 360° range, for a given movie length, an increase in dot lifetime will lead to corresponding decreases in the number of independent noise direction samples that are made available. As Scase *et al* (1996) have pointed out, "this lack of independence will increase the variability contributed by the noise dots to any global measure...". Of course, such an assertion makes an assumption about the visual system. It assumes that the encoding of dot correspondences is not necessarily independent from the encoding of other dot correspondences (presumably this would arise because multiple dot pairings from a single trajectory may be activating common local detectors). Indeed, Scase *et al* have reported subtle yet significant elevations in coherence threshold (with a left/right translation discrimination task) for "random direction" noise conditions relative to "random walk" noise conditions[6]. Inclusion of the

---

[5] In the *Dynamic-2* condition on average only 50% of the possible signal pairings will result in correspondences representative of the signal. In the *Dynamic-6* condition, this increases to 83%.

[6] "Random walk" noise can, for the purposes of this comparison be considered analogous to our *Dynamic-2* noise.

Dynamic-2 noise condition was therefore motivated by the possibility that lower levels of noise variability might permit higher levels of summation to be observed.

Dynamic-2* versus *Twinkle-2

Noise dots, that by chance are reallocated to positions that are consistent with those expected from signal dot pairings (i.e. correspondence noise) might occur, on average, more frequently in *Dynamic-2* displays than the *Twinkle-2* displays. This would result from the fact that the reallocation of noise dots in the former case is restricted to displacement distances that are equal to those of the signal dots; hence, they will always be matched to them in terms of speed. While many formulations of ideal observer performance have attributed the observer with a perfect knowledge of the signal (e.g. Barlow & Tripathy, 1997) it is acknowledged that there is empirical evidence to suggest that performance is more likely to be reflecting a pooling of local signals from a wide range of speeds and directions (Watamaniuk *et al*, 1989). Irrespective of exactly which dot pairings may constitute correspondence noise, it seems plausible that performance would be limited by it to a greater extent in the *Dynamic-2* noise displays than in the *Twinkle-2* noise displays. In the former case, it can be envisaged that as the proportion of signal dots is increased the amount of correspondence noise would decrease (at least for sufficiently low dot densities). This may be less so in the *Twinkle-2* noise displays, where even at low signal levels, the displacements for many more noise dot pairings might exceed that needed for motion detection (d_{max}). Inclusion of the *Twinkle-2* noise condition was therefore motivated by the possibility that *Dynamic-6* noise may have contained higher levels of correspondence noise. If so, it would be expected to decrease with increasing signal levels, which may explain the less than linear levels of spatial summation.

9.4.1 Stimuli and procedure

Coherence thresholds were estimated for the *Trans* motion type with three types of noise: *Dynamic-6*, *Dynamic-2*, and *Twinkle-2* (see above for definitions). Both noise conditions were also tested: varying-noise and constant-noise. In a single experimental session, data were gathered from displays containing 2, 4, 8, and 16 S+N patches, for a particular noise type and condition. Single S+N patches were not tested because of the possibility that for the two shorter dot lifetime conditions sensitivities might be immeasurable. To speed up data gathering, the 1IFC direction discrimination task (as described in experiment 3c) was used. Other details remained the same as the main experiment (5a), with the exceptions that the computer and display equipment had changed, and as a consequence the VD was changed from 114cm to 154cm. A single observer (PAA) collected the data.

9.4.2 Results

The upper panels of fig.9.11 compare the data for the *Dynamic-6* noise conditions of the present and previous experiment (3c). For the constant-noise (with intermediate N patches) displays, the three S+N patch numbers common to both experiments (2,4, and 8) gave almost identical sensitivities. Sensitivity for the single S+N patch of the previous experiment did however, deviate somewhat from the Minkowski fit and so acted to increase the β exponent. In the varying-noise (without intermediate N patches) displays, agreement between the two experiments did not seem as close, due possibly to practice effects. The lower two panels of fig.9.11 plot data for all three types of the noise condition used in the present experiment. For the constant-noise conditions sensitivities both for *Dynamic-2* and *Twinkle-2* tended to show greater deviations from their fitted slopes, although the resulting slope magnitudes were similar to that of *Dynamic-6*. In the varying-noise conditions sensitivities to coherent translation were very similar for *Dynamic-2* and *Twinkle-2*, but higher for the *Dynamic-6*. This finding is in accordance with the idea that more signal is present in this latter condition.

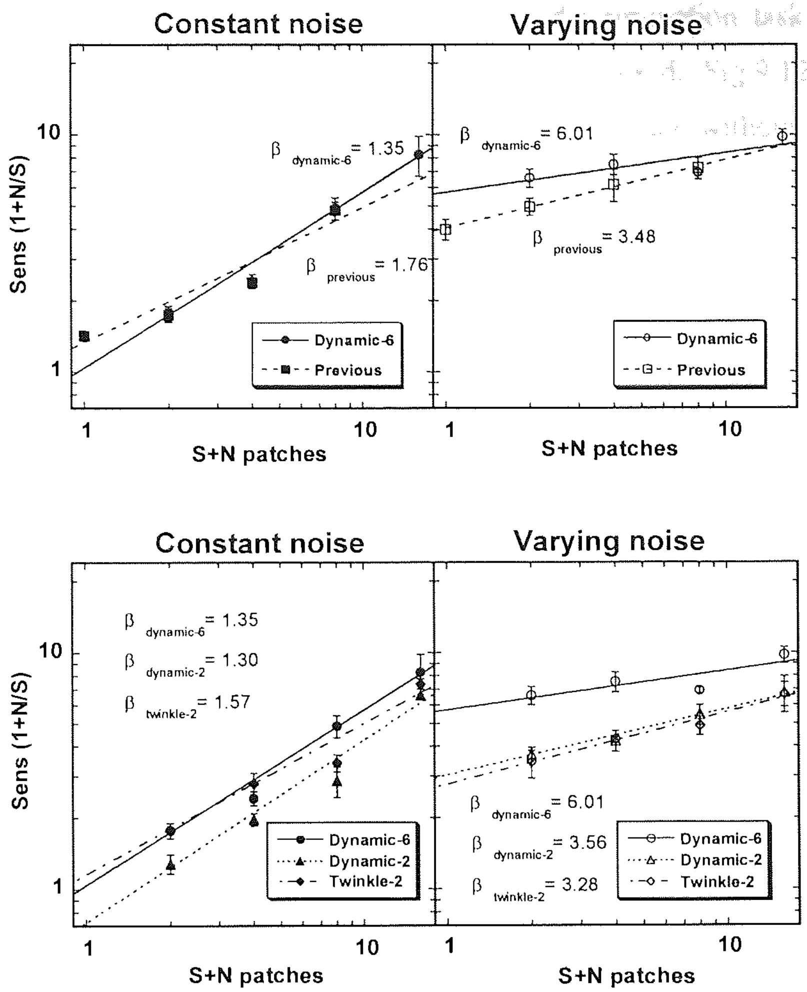


Fig. 9.11. Data from observer PAA for the various noise types. All data are for the *Trans* motion type. The two upper panels plot results from the *Dynamic-6* condition of the present experiment (*Dynamic-6*) and the same condition as carried out in Experiment 5c. The lower two panels plot results for all three noise conditions of the present experiment. As before the curves represent Minkowski model fits.

9.5 Experiment 5e: Sectors

In this control experiment, sectors were used rather than patches. These displays therefore bore a closer similarity to the ones used by Morrone *et al* (1995). One reason for suspecting that performance differences could occur between sector and patch displays – especially for the constant-noise conditions – was that, for circular arrangements of sectors a grouping by form (aperture shape) seems much more likely than for a circular arrangement of patches. Whether this could also lead to a difference in the levels of area summation is unclear, but the potential importance that aperture shape can have upon the perception of motion has been touched on previously in section 1.3.4 (Watanabe, 1997).

9.5.1 Stimuli and procedure

Data were gathered again using a 1IFC direction discrimination task. Instead of using circular patches of random dot stimuli, sectors were used. Fig.9.12 depicts example displays for when four S+N patches are present (both with and without intermediate noise sectors). Note that the displays were notionally divided into 16 sectors of equal area. The outer display diameter was set to 8.85° to match those of the previous patch based experiments. A small central blank region with a diameter of 0.23° was also included. Other stimulus and task details remained the same as those in experiment 5c. However, because of the greater stimulus area in this experiment, the absence of a speed gradient with the *Rot* and *Exp* conditions did become noticeable. Two observers participated in the study. Each gathered data for the three motion types *Rot*, *Exp*, and *Trans*, in both constant-noise and varying-noise conditions.

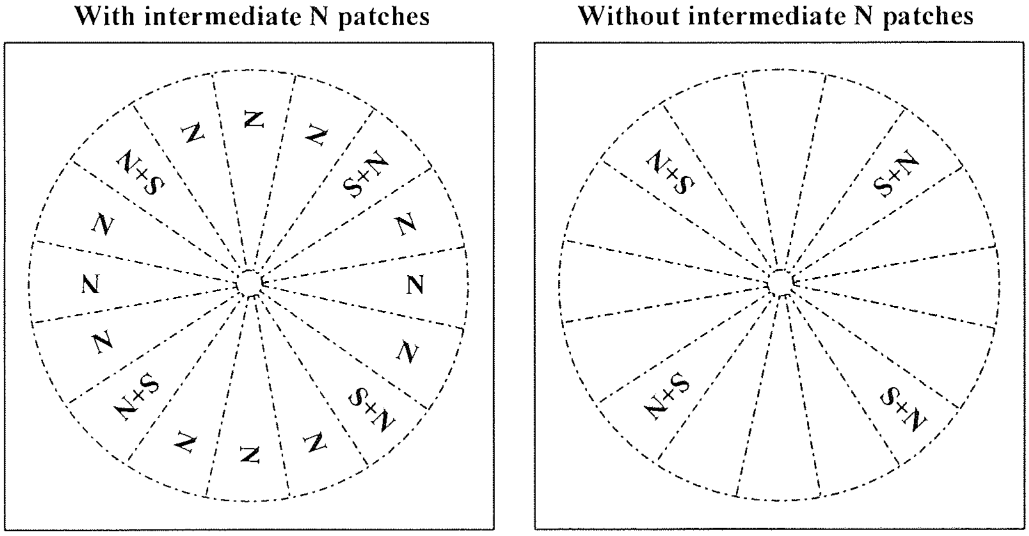


Fig.9.12. Example displays from the sector experiment 5e. Each display is notionally divided into 16 equal sectors. In these two examples, only four contained signal. As with experiments 5a-d S+N areas were always maximally spaced from each other. The left schematic depicts a constant-noise display and the right, a varying-noise display.

9.5.2 Results

In fig.9.13 and fig.9.14, data are plotted for observer PAA and JBS respectively. All three motion types in the constant-noise conditions, yielded best-fit β exponents that were close to 1.0 for both observers. In the varying-noise conditions, however, there was still a trend for β exponents in to be greater than 2.0.

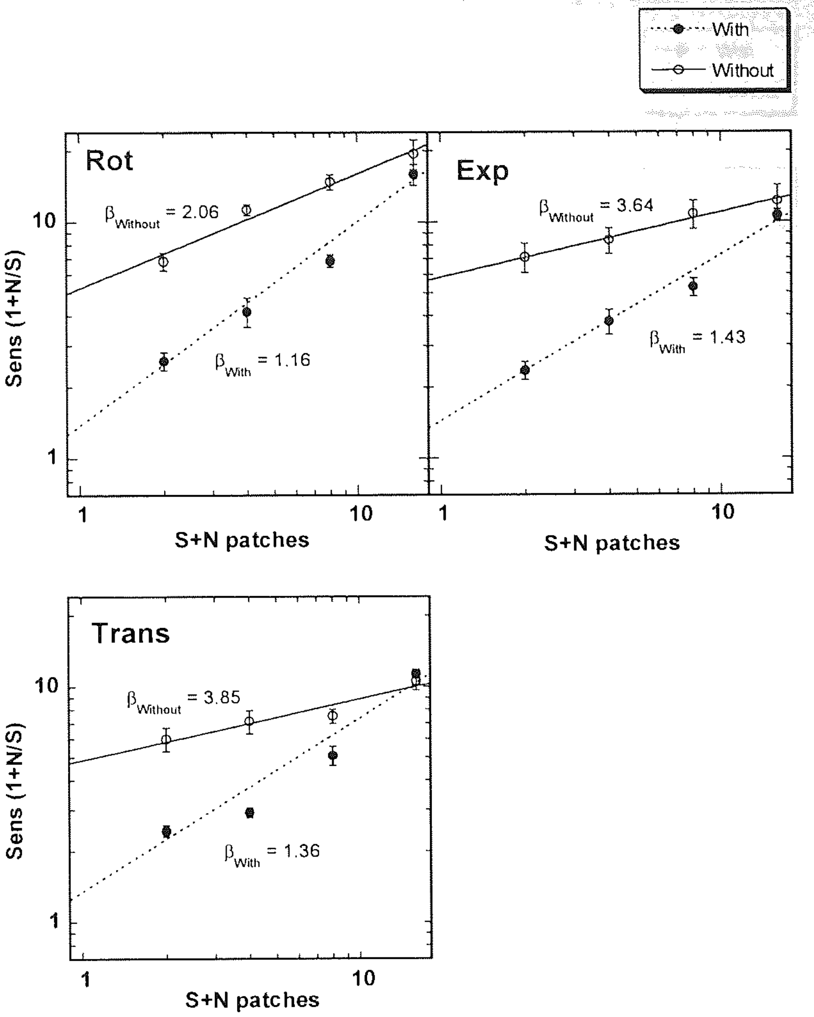


Fig.9.13. Data from observer PAA for the three types of motion. Within each panel, mean sensitivities are plotted as a function of S+N number for both the constant-noise and varying-noise conditions. The β_{With} and β_{Without} values denote the β exponents for the constant- and varying-noise conditions respectively.

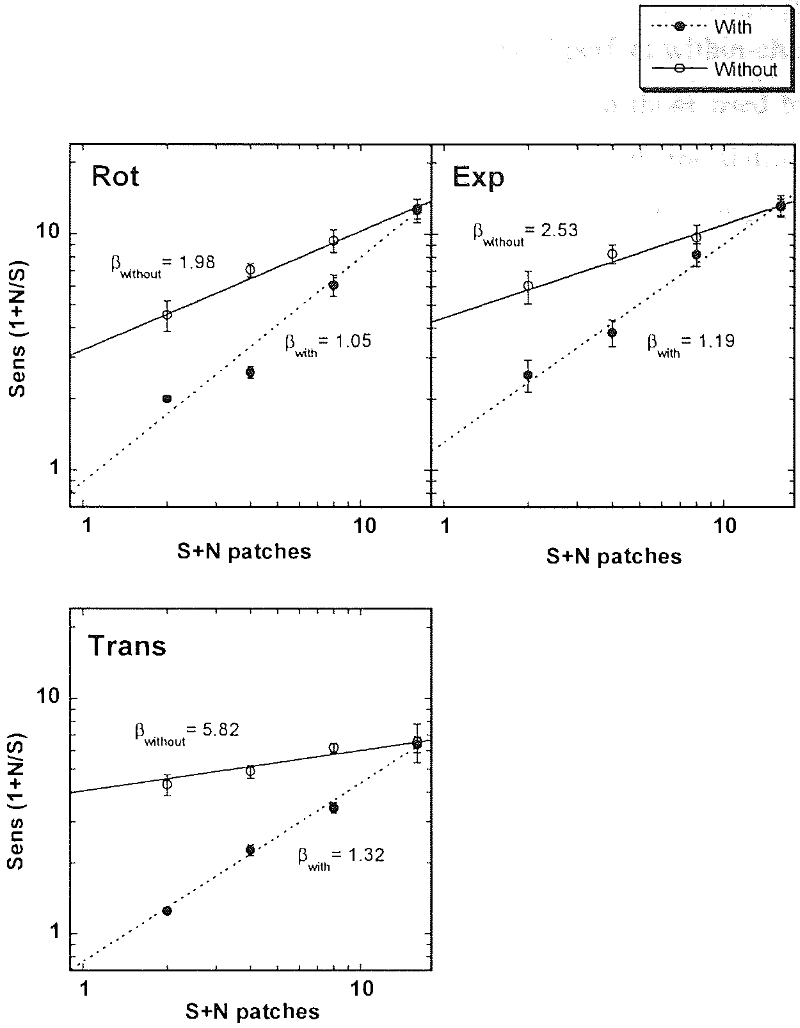


Fig.9.14. Data from observer JBS for the three types of motion. As in fig 9.13, mean sensitivities are plotted as a function of S+N number for both the constant-noise and varying-noise. The β_{With} and $\beta_{Without}$ values denote the β exponents for the constant- and varying-noise conditions respectively.

9.6 Conclusions from the control experiments 5b-e

Changing from a monocular to binocular viewing clearly had little to no effect upon summation levels. Changing from a 2IFC detection task to a 1IFC discrimination task also seemed to have little effect upon summation levels, with one exception. For the rotation motion conditions, the Minkowski (β) coefficients derived from the 1IFC data were in close accordance with the predictions of perfect within-channel summation. Why this should occur for rotation alone is not clear. The interpretation of the data for the different noise conditions was confused by the fact that the variations between the summation levels were not necessarily as great as those for the identical conditions carried out in different experiments. Clearly, however, the three different noise conditions did not give vastly differing degrees of summation.

The manipulation that yielded the clearest indications of perfect within-channel summation came from the use of contiguous sector displays, similar to those used by Morrone *et al* (1995). However, even using this display configuration, both translation and expansion still showed a tendency to give slightly lower levels of summation than those reported by Morrone *et al* (1995) and Burr *et al* (1998).

Chapter Ten: Summary and conclusions

10.1 Overview

The aim of this work was to explore the representation of higher-level motion in the brain. Of specific interest were the following questions. (1) Are motion representations more readily accessible to the observer if they elicit impressions of 3-dimensionality? (2) If so is this reflected in a clear parallel/serial search dichotomy? (3) Can multiple complex motions of the same type, but positioned at different spatial locations, be processed in parallel? (4) If so, when near coherence threshold, are they processed independently up until the level of decision or they pooled at some stage prior to decision?

In all the experiments of this thesis, motion representations were achieved through the use of random-dot-kinematograms. Experiments 1-3b employed a standard singleton search paradigm to test the hypothesis that motion representations are more readily accessible to observers if they elicit impressions of 3-dimensionality. Experiments 4-5e describe subthreshold summation experiments. In experiment 4, pairs of superimposed orthogonal complex motions were tested in an attempt to establish whether the summation effects are due to parallel processing. In the remaining experiments (5a-e), a spatial summation technique was used to confirm summation between multiple patches of (spatially discrete) motion and the likely origin of this summation.

10.2 Are motion representations more readily accessible when they elicit impressions of 3-dimensionality?

In experiment 1, the '3-D access' hypothesis was tested using motions that were constructed from 2-D complex motion components. This choice of motion stimuli seemed like a sensible starting point given the suggestions that biological motion analysis could involve the initial extraction of 2-D components (e.g. Koenderink, 1986; Meese *et al*, 1995). Moreover, a preliminary study by Freeman *et al* (1996b) had found tentative evidence to suggest that the efficient detection of odd 2-D motion components might depend upon whether their resulting motions are searched based upon 3-D representations or not. The key comparison that was to be made in experiment 1 was between deformation flows and deformation-plus-translation flows. Freeman *et al* (1996b) had posited that both could yield compelling impressions of surface slant in trained observers, but only the later could yield the unambiguous impressions that would be required were slant sign to be used as the basis for search. Contrary to these predictions search for both the deformation and the deformation-plus-translation conditions were found to be highly inefficient and most importantly the deformation-plus-translation condition was the less efficient of the two. In addition, the '3-D access' hypothesis predicted efficient search for radial motion as they produce compelling impressions of motion-in-depth. Again, no evidence for efficient search was found.

In summary, the results from a single observer in this search experiment failed to provide any evidence to support the '3-D access' hypothesis when motion cues are used. However, a potential problem was identified. One of the premises of the experiment was that the deformation alone condition would be incapable of supporting search based upon 3-D representations. The subjective experience of the observer suggested otherwise; to this observer other 3-D cues were available. This observation brought into question the usefulness of search slope differences for testing the '3-D access' hypothesis.

In experiment 2, translation motions containing 1-D (vertical) speed gradients were used. These simulated, moving planar surfaces rather than the curved/non-rigid surface impressions that tended to be elicited from deformation-plus-translation motions. These motions therefore, more closely resembled the planar surfaces simulated in a study by Holliday & Braddick (1991) for singleton search experiments. This similarity was of significance because, using disparity defined planes, Holliday & Braddick had reported parallel search for odd directions of slant. It was speculated that if generic cue-independent representations had been utilised in these searches, then parallel search would also be expected when similar slant representations are derived from motion parallax. However, despite providing closer matches to the simulated surfaces of Holliday & Braddick (1991), and them, subjectively, appearing more salient than the horizontal slant from the deformation-plus-translation motions, motions containing 1-D speed gradients did not show evidence of efficient parallel search. This did not disprove the notion that cue-independent slant representations can be accessed simultaneously. It may have been that motion parallax stimuli, generated with RDKs, are incapable of producing sufficiently compelling impressions of slant.

In experiments 3a-b the following question was addressed: if the 3-D representations from motion stimuli could be grouped together, would their individual representations become more accessible to the observer? One reason for suspecting that this might occur comes from studies that suggested that when display items group into common (3 dimensional) surfaces they are attended to more easily, and are therefore searched with better efficiency

(He & Nakayama, 1995; Morita & Kumada, 2003). An absence of powerful grouping factors in experiment 2 may have rendered access to 3-D representations inefficient. Three types of display configuration were conceived of to the test this grouping hypothesis. All three were designed to promote grouping amongst distractor patches containing motion parallax with the grouping based upon a principle of good-continuity in 3-dimensions.

In the first configuration (experiment 3a), the distractors were arranged so as to permit distractor grouping by common surface on a row-wise basis. The results were consistent with the above grouping hypothesis. However, in the second, when search displays were arranged so that distractors were free to group on the basis of a single, approximately corrugated surface (experiment 3b), no evidence was found in favour of the grouping hypothesis. One reason why search performance may not have been superior in the corrugated surface condition was that the continuity between the distractor slants of the different rows was not sufficiently salient. Using distractors that simulate slants conforming to a single coplanar surface might ensure better impressions of surface continuity. However, the ability to simulate coplanar surface impressions from motion parallax is dependent upon how the visual system obtains the estimates of translation that are used to derive representations of slant. If estimates were based purely upon an analysis of translation, which is restricted to the individual patches, the distractor patches would not be expected to adhere to impressions of a display-wide coplanar surface. Alternatively, if slant representations were based upon a global pan-patch analysis of translation then the slants from each distractor patch would be expected to provoke impressions of an implicit coplanar surface. This empirical issue was addressed in a preliminary slant matching experiment (3c). The results provided no evidence in favour of a pan-patch analysis of translation and therefore no reason to believe that, using motion parallax stimuli, individual distractor patches would group by common coplanar surface; this search experiment was therefore not carried out.

The results from the search experiments therefore failed to provide any evidence in support of the '3-D access' hypothesis, despite using several different types of motion stimuli and making attempts to group distractor patches by common surface. There are four possible explanations. (1) Multiple surface representations simply cannot be processed simultaneously. This would be the account given by classical FIT. (2) Another possibility is that with sufficient grouping, the attentional demands of the display can be lessened, so permitting an efficient monitoring of the multiple surface representations. (3) Irrespective as to the extent of stimulus grouping, surface representations derived from RDK motions alone can not be made sufficiently salient for them to act as effective cues to segmentation. This, however, does not rule out the possibility that motion cues can contribute to more generic representations of surface that can be accessed efficiently. Further insight into this idea might be garnered by measuring odd-item-out search performances for stereoscopic defined slant, for conditions in which the stereoscopic cues are either in concordance or conflict with motion cues to slant. The study by Holliday & Braddick (1991) had found that stereoscopic defined slant can be searched in parallel. If performance had been based upon slant representations specific to stereoscopic cues then the addition of motion parallax cues should little effect the efficient search performance. If, however, performance had been based upon slant representations that can also be influenced by motion cues, then when the motion cues are in conflict, search performance would be expected to worsen.

10.3 Are multiple representations of complex motion processed independently or pooled at some perceptual level?

In experiment 1 the search efficiencies of the three complex motion types were tested: radial, rotation, and deformation. Consistent with a previous report by Braddick & Holliday (1991), searches for an odd sign of radial or deformation motion were highly inefficient. Contrary to the interpretations given by theories such as FIT, these poor search efficiencies were not taken as evidence against the existence of specialised mechanisms that are selective for these motions. Instead, the search inefficiencies were attributed to the observers' inability to monitor the relevant mechanisms effectively. This interpretation seems the most likely (at least for the case of radial motion) given the psychophysical evidence favouring the existence of specialised complex motion detectors (cf. Regan, 1993; Braddick, 1993). In chapter 7, it was also pointed out that low threshold theory accounts of visual search might also be able to account for positive search slopes, even when stimulus representations are efficiently processed and accessed. Using a modified search paradigm Thornton & Gilden (2001) were able to confirm that multiple representations of both expansion and contraction are processed in parallel. However, because the task did not require the observers to carry out odd-item-out detection, it was speculated that the parallelism could have derived either from independent detectors or a pooling of the motion signals at some earlier perceptual level.

To address whether spatially distinct patches of complex motion are processed independently, are pooled, or neither, a subthreshold summation technique was used. This permitted coherence threshold to be used as a performance measure, therefore circumventing issues of speed-accuracy trade-off that are inherent to RT measures. Coherence thresholds were obtained for displays containing various motion patch (S+N) numbers, both for when the total noise of the display was kept constant and for when it varied with S+N patch number. Thresholds for rotations and expansions, showed clear decreases with increasing S+N patch number. This was true for both the constant and the varying noise conditions. In both, these threshold decreases were attributed to parallel processing. It could have been argued that the threshold decreases might also have been expected from a serial analysis (De Bruyn & Orban, 1993), in which observers adopt a strategy of shifting attention between the display patches during the stimulus interval. Burr & Santoro (2001) have, however, reported long summation periods for complex motion coherence thresholds; much longer than the interval durations that were used in this experiment. It therefore seems unlikely that a strategy of switching between patches would have led to improvements in performance. Moreover, in experiment 4, further support was given to the notion that multiple complex motions are indeed processed and accessed in parallel.

In addition to the conditions in which the S+N patches contained their own discrete representations of complex motion (patch-wide) another set of conditions were tested, in which a single representation of complex motion spanned all S+N patches in the display (display-wide). Comparison between the patch-wide and display-wide conditions revealed decreases in coherence thresholds (summation levels) to be very similar, both for the rotation and expansion motion types. Moreover, for the constant-noise conditions these summation levels were greater than those typically expected from probability summation. While high levels of spatial summation have been reported previously for single display-wide complex motion flows, levels of summation have not been reported for multiple spatially discrete complex motion flows (Morrone *et al*, 1995; Burr *et al*, 1998; Meese & Harris, 2001b). From these results it seems likely that coherence thresholds for both the display-wide and patch-wide motions were based upon a pooling of signals from across the display patches, and therefore was probably occurring at some perceptual stage.

In a supplementary study reported in chapter 8 (experiment 4), a variation of experiment 2 (independent detectors for complex motion) from Meese & Harris (2001b) was carried out. Its aim was to establish whether the decreases in coherence threshold, which they had observed when pairs of (orthogonal) complex motions were superimposed, was due to parallel processing or serial processing. To test this, a similar experiment was rerun under restricted viewing conditions. Despite these restricted viewing conditions, for at least one of the subjects, decreases in coherence threshold were still observed in the compound conditions. Furthermore, these decreases were consistent with those that have previously been attributed to probability summation. This finding supports the notion that independent detectors are responsible for the processing of orthogonal complex motion flows when near threshold.

10.5 Further study

The results from the main spatial summation experiment (5a) are of particular interest, as it raises the possibility that complex motion mechanisms exist, which respond in similar ways irrespective of the number of individual motion flows falling within their RF (provided the total amount of signal remains the same). Additional experiments might address the following questions. (1) Are the high levels of summation for the patch-wide motions best explained by general-purpose mechanisms, capable of configuring themselves to many types of coherent motion, or mechanisms selective for specific complex motions? One approach to answering this question would be to test summation levels when different types of patch-wide complex motion are used. If lower summation levels are observed then this would argue against matched filter accounts. (2) If the high summation levels are best explained by processes selective for particular components of complex motion, can spatially discrete components be summed irrespective of their polarity? This might be answered by testing summation levels in which half the display patches always contain motion of one polarity and the other half contain motion of the other polarity. Lower summation levels in this case would indicate polarity to be important. Answers to the above questions may lead to fresh insights into the properties of the putative mechanisms responsible for the detection of complex motion.

References

Adelson, E.H.; & Bergen, J.R. (1985) Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, **2**, 284-299.

Adelson, E.H. & Movshon, J.A. (1982) Phenomenal coherence of moving visual patterns. *Nature*, **300**, 523-525.

Albright, T.D. & Desimone, R. (1987) Local precision of the visuotopic organisation in the middle temporal area (MT) of the macaque. *Experimental Brain Research*, **65**, 582-592.

Andersen, G.J., & Atchley, P. (1997) Smoothness of the velocity field and three-dimensional surface detection from optic flow. *Perception and Psychophysics*. **59**. 358-369.

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

Bakin, J.S., Nakayama, K., & Gilbert, C.D. (2000) Visual Responses in Monkey Areas V1 and V2 to Three-Dimensional Surface Configurations. *The Journal of Neuroscience*. **20**, 8188-8198.

Baldassi, S., & Burr, D.C. (2000) Feature-based integration of orientation signals in visual search. *Vision Research*, **40**, 1293-1300.

Baldassi, S., & Verghese, P. (2002) Comparing integration rules in visual search. *Journal of Vision*, **2**, 559-570.

Ball, K., & Sekuler, R. (1980) Models of stimulus uncertainty in motion perception. *Psychological Review*, **87**, 435-467.

Ball, K.; Sekuler, R. (1981) Cues reduce direction uncertainty and enhance motion detection. *Perception and Psychophysics*, **30**, 119-128.

Barlow,H.B.; Tripathy,S.P. (1997) Correspondence noise and signal pooling in the detection of coherent visual motion. *The Journal of Neuroscience*, **17**, 7954-7966.

Bennett, P, & Jaye, P.D. (1995) Local Localization, Not Discrimination, Is Constrained by Attention. *Canadian Journal of Experimental Psychology*, **49**, 460-503.

Beverley, K.I., & Regan, D. (1979) Visual perception of changing size: the effect of object size. *Vision Research*, **19**. 1093-1104.

Bex, P.J., Metha, A.B., & Makous, W. (1998) Psychophysical evidence for a functional hierarchy of motion processing mechanisms. *Journal of the Optical Society of America A*, **15**, 769-776.

Blakemore, C., & Campbell, F.W. (1969) On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, **203**, 237-260.

Bonneh, Y., & Sagi, D. (1998) Effects of spatial configuration on contrast detection. *Vision Research*, **38**, 3541-3553.

Braddick, O.J. (1993) Adaptation and 'pop-out' in the analysis of divergence-sensitive mechanisms. *Perception*, **22**, 500.

Braddick, O.J., & Qian, N. (2001) The organization of global motion detection and transparency. In Zanker, J.M., & Zeil, J (Eds), *Motion vision - Computational, neural, and ecological constraints*, 85-112, Springer Verlag. New York.

Braddick, O.J., & Holliday, I.E. (1991) Serial search for targets defined by divergence or deformation of optic flow. *Perception*. **20**, 345-354.

Bradley, D.C., Chang, G.C., & Andersen, R.A. (1998) Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature*, **392**, 714-717.

Bradshaw,M.F., & Rogers, B.J. (1996) The interaction of binocular disparity and motion parallax in the computation of motion. *Vision Research*, **36**, 3457-3468.

Braunstein, M.L. (1968) Motion and texture as a sources of slant information. *Journal of Experimental Psychology*, **78**, 247-253.

Bravo, M.J. (1998) A Global Process in Motion Segregation. *Vision Research*, **38**, 853-863.

Bravo, M.J, & Nakayama, K (1992) The role of attention in different visual search tasks. *Perception and Psychophysics*, **51**, 465-472.

Briand, K., & Klein, R.M. (1987) Is Posner's "beam" the same as Treisman's "glue"? On the relationship between visual orienting and feature integration theory. *Journal of Experimental Psychology: Human Perception and Performance*, **13**, 228-241.

Britten, K.H., & Newsome, W.T. (1998) Tuning bandwidths for near-threshold stimuli in area MT. *Journal of Neurophysiology*. **80**, 762-770.

Bruce, V., Green,P.R., & Georgeson, M.A. (2003) Visual perception: Physiology, psychology and ecology (4th edn). Psychology Press. Hove.

Buracas, G.T., & Albright, T.D. (1996) Contribution of area MT to perception of three-dimensional shape: a computational study. *Vision Research*, **36**, 869-887.

Burr, D., Badcock, D.R., & Ross, J. (2001) Cardinal axes for radial and circular motion, revealed by summation and by masking. *Vision Research*, **41**, 473-481.

Burr, D.C., Morrone, M.C., & Vaina, L.M. (1998) Large receptive fields for optic flow detection in humans. *Vision Research*, **38**, 1731-1743.

Burr, D.C., & Santoro, L. (2001) Temporal integration of optic flow, measured by contrast and coherence thresholds. *Vision Research*, **41**, 1891-1899.

Campbell, F.W., & Robson, J.G. (1968) Application of fourier analysis to the visibility of gratings. *Journal of Physiology*, **197**, 551-566.

Cavanagh, P., Arguin, M., & Triesman, A. (1990) Effect of surface medium on visual search for orientation and size and size features. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 479-492.

Cohn, T.E., & Lasley, D.J. (1974) Detectability of a luminance increment: Effect of spatial uncertainty. *Journal of the Optical Society of America*, **64**, 1715-1719.

Corbetta, M., Miezin, R.M., Dobmeyer, S., Shulman, G.L., & Petersen, S.E. (1991) Selective and divided attention during visual discriminations of shape, colour, and speed: Functional anatomy by positron emission tomography. *Journal of Neuroscience*, **11**, 2382-2402.

Cornsweet, T.N. (1962) The staircase method in psychophysics. *American Journal of Psychology* ,**75**, 485-491.

Crick, F. & Koch, C. (1995) Are we aware of neural activity in primary visual cortex. *Nature*, **375**, 121-123.

Cumming, B.G. & Parker, A.J. (2000) Local disparity not percieved depth is signaled by binocular neurons in cortical area V1 of the macaque. *The Journal of Neuroscience*, **20**, 4758-4767.

Dakin, S.C. (2001) Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A,* **18**, 1016-1026.

Dakin, S.C., & Bex, P.J. (2002) Summation of concentric orientation structure: seeing the Glass or the window? *Vision Research*, **40**, 2013-2020.

Dakin, S.C., & Bex, P.J. (2003) Response to Wilson & Wilkinson: Evidence for global processing but no evidence for specialised detectors in the visual processing of Glass patterns. *Vision Research*, **43**, 565-560.

Davis, E.T., Kramer, P., & Graham, N. (1983) Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Perception and Psychophysics*, **33**, 20-28.

Davis, E.T., Shikano, T., Peterson, S.A., & Michel, R.K. (2003) Divided attention and visual search for simple versus complex features. *Vision Research*, **43**, 2213-2232.

Davis, G., & Driver, J.S. (1994) Parallel detection of Kanizsa subjective figures in the human visual system. *Nature*, **371**, 791-793.

Davis, G., & Driver, J.S. (1998) Kanizsa Subjective figures can act as occluding surfaces at parallel stages of visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 169-184.

De Bruyn, B., & Orban, G.A. (1990) The role of direction information in the perception of geometric optic flow components. *Perception and Psychophysics*, **47**. 433-438.

De Bruyn, B., & Orban, G.A. (1993) Segregation of spatially superimposed optic flow components. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 1014-1027.

Domini, F, & Caudek, C (1999) Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 426-444.

Donnelly, N., Humphreys, G.W., & Riddoch, M.J. (1991) Parallel Computation of Primative Shape Descriptions. *Journal of Experimental Psychology: Human Perception and Performance*, **17**, 561-570.

Driver, J., & McLeod, P. (1992) Reversing visual search asymmetries with conjunction search of movement and orientation. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 22-33.

Driver, J., McLeod, P., & Dienes, Z. (1992) Are direction and speed coded independently by the visual system? Evidence from visual search. *Spatial Vision*, **6**, 133-147.

Duffy, C.J., & Wurtz, R.H. (1991a) Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, **16**, 1329-1344.

Duffy, C.J., & Wurtz, R.H. (1991b) Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small stimuli. *Journal of Neurophysiology*, **16**, 1346-1359.

Duncan, J. (1980) The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, **87**, 272-280.

Duncan, J., & Humphreys, G.W. (1989) Visual search and visual similarity. *Psychological Research*, **96**, 433-458.

Edwards, M., & Badcock, D.R. (1993) Asymmetries in the sensitivity to motion in depth: A centripetal bias. *Perception*, **22**, 1013-1023.

Edwards, M., & Badcock, D.R. (1994) Global motion perception: Interaction of the ON and OFF pathways. *Vision Research*, **34**, 2849-2858.

Edwards, M., & Nishida, S. (1999) Global motion detection with transparent motion signals. *Vision Research*, **39**, 2239-2249.

Egeth, H.E., Virzi, R.A., & Garbart, H. (1984) Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 32-39.

Enns, J.T., & Rensink, R.A. (1991) Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*, 98, 335-351.

Estes, W.K., & Taylor, H.A. (1964) A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Science*, **52**, 446-454.

Finney, D.J. (1971) *Probit analysis*. Cambrifge University Press. London.

Foster, D.H., & Ward, P.A. (1991) Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society (London B)*, **243**, 83-86.

Freeman, R.D., & Ohzawa, I. (1990) On the neurophysiological organisation of binocular vision. *Vision Research*, **30**, 1661-1676.

Freeman, T.C.A., & Fowler, T.A. (2000) Unequal retinal and extra-retinal motion signals produce different perceived slants of moving surfaces. *Vision Research*, **40**, 1857-1868.

Freeman, T.C.A., & Harris, M.G. (1992) Human sensitivity to expanding and rotating motion: effects of complementary masking and directional structure. *Vision Research*, **32**, 81-87.

Freeman, T.C.A., Harris, M.G., & Meese, T.S. (1996a) On the Relationship Between Deformation and Perceived Surface Slant. *Vision Research*, **36**, 317-322.

Freeman, T.C.A., Harris, M.G., & Meese, T.S. (1996b) Discrimination of optic flow components: access to the third dimension? *Perception*, **25**, (Supplement).

Friedman-Hill, S.R., & Wolfe, J.M. (1995) Second-order parallel processing: Visual search for the odd item in a subset. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 531-551.

Gautama, T., & Van Hulle, M.M (2001) Function of centre-surround antagonism for motion in visual area MT/V5: a modeling study. *Vision Research*, **41**, 3917-3930.

Geisler, W.S. Motion streaks provide a spatial code for motion direction. *Nature*, **400**, 65-69.

Georgeson, M, Freeman,T.C.A., & Scott-Samuel, N.E. (1996) Sub-pixel accuracy: Psychophysical validation of an algorithm for fine positioning and movement of dots on visual displays. *Vision Research*, **36**, 605-612.

Graham, N. (1989) *Visual pattern analysers*. Oxford University Press. New York.

Graziano, M.S.A., Andersen, R.A., & Snowden, R.J. (1994) Tuning of MST neurons to spiral motions. *Journal of Neuroscience*, **14**, 54-67.

Grosof, D.H., Shapley, R.M., & Hawken, M.J. (1993) Macaque V1 neurons can signal 'illusory' contours. *Nature*, **365**, 550-552.

Gurnsey,R., Humphrey,G.K., & Kapitan,P. (1992) Parallel discrimination of subjective contours defined by offset gratings. *Perception and Psychophysics*, **52**, 263-276.

Haenny, P.E.; Maunsell, J.H.R.; Schiller, P.H. (1988) State dependent activity in monkey visual cortex: II. Retinal and extraretinal factors in V4. *Experimental Brain Research*, **69**, 245-259.

Harris, M.H, (1994) Optic and retinal flow. In Smith, A.T., & Snowden, R.J, *Visual detection of motion*. **11**, 307-332, Academic Press. London.

Harris, M.G., Freeman, T.C.A., & Hughes, J. (1992) Retinal Speed Gradients and the Perception of Surface Slant. *Vision Research*, **32**. 587-590.

He, S., Cavanagh, P., & Intriligator, J. (1996) Attention resolution and the locus of visual awareness. *Nature*, **383**, 334-337.

He, S., & MacLeod, D.I.A. (2001) Orientation-selective adaptation and tilt after-effect from invisible patterns. *Nature*, **411**, 473-476.

He, Z.J., & Nakayama, K. (1992) Surfaces verses features in visual search. *Nature*, **359**, 231-233.

He, Z.J., & Nakayama, K. (1994) Perceiving Textures: Beyond Filtering. *Vision Research*, **34**, 151-162.

He, Z.J., & Nakayama, K. (1995) Visual Attention to surfaces in three-dimensional space. *Proceedings of the national academy of sciences of the USA*, **92**, 11155-11159.

Heathcote, A., & Mewhort, D.J.K. (1993) Representation and selection of relative position. *Journal of Experimental Psychology: Human Perception and Performance*, **19**, 488-516.

Hick, W.E. (1952) On the rate of gain of information. *Quarterly Journal of Experimental Psychology* ,**4**, 11-26.

Holliday, I.E., & Braddick, O.J. (1991) Pre-attentive detection of a target defined by stereoscopic slant. *Perception*, **20**, 355-362.

Howard, I.P., & Rogers, B.J. (2002) *Seeing in depth (Volume 2: Depth perception)*, I.Porteous. Ontario.

Hubel, D.H., & Wiesel, T.N. (1968) Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, **195**, 215-243.

Huk, A.C.; Heeger, D.J. (2000) Task-related modulation of visual cortex. *Journal of Neurophysiology*, **83**, 3525-3536.

Humphrey, G.K., Goodale, M.A., Bowen, C., Gati, J.S., Vilis, T., Rutt, B.K., & Menon, R.S. (1997) Differences in perceived shape from shading correlate with activity in early visual areas. *Current Biology*, **7**, 144-147.

Janssen, P., Vogels,R.,& Orban, G.A. (1999) Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proceedings of the National Academy of Sciences of the USA*, **96**, 8217-8222.

Joseph, J.S., Chun, M.M., & Nakayama, K (1997) Attentional requirements in a "preattentive" feature task. *Nature*, **387**, 805-807.

Julesz, B. (1986) Texton gradients: The texton theory revisited. *Biological Cybernetics*, **54**, 245-251.

Julesz,B., & Hesse, R.I. (1970) Inability to percieve the direction of rotation movement of line segments. *Nature*, **255**, 243-244.

Kleffner, D.A., & Ramachandran, V.S. (1992) On the perception of shape from shading. *Perception and Psychophysics*, **52**, 18-36.

Klein, J.A., & Lane, D.M. (1986) Locus and persistence of capacity limitations in visual information processing. **12**, 200-210.

Klein, R. (1988) Inhibitory tagging system facilitates visual search. *Nature*. **334**, 430-431.

Koenderink, J.J. (1986) Optic Flow. *Vision Research*, **26**, 161-179.

Koenderink, J.J., & van Doorn, A.J. (1975) Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, **22**, 773-791.

Koenderink, J.J., & van Doorn, A.J. (1991) Affine structure from motion. *Journal of the Optical Society of America A*, **8**, 377-385.

Kurki, I. Laurinen, P., Peromaa, T., & Saarinen, J. (2003) Spatial integration in Glass patterns. *Perception*, **32**, 1211-1220.

Lamme, V.A.F., Zipser, K., & Spekreijse, H. (1999) Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proceedings of the National Academy of Sciences of the USA*, **95**, 3263-3268.

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

Lee, D.K., Itti, L., Koch, C., & Braun, J. (1999) Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, **2**, 370-374.

Lennie, P. (2003) The cost of cortical computation. *Current Biology*, **13**, 493-497.

Logan, G.D. (1994) Spatial Attention and the Apprehension of Spatial Relations. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 1015-1036.

Luschow, A., & Nothdurft, H.C. (1993) Pop-out of orientation but not pop-out of motion at isoluminance. *Vision Research*, **33**, 91-104.

MacLeod, D.I.A., & He, S. (1992) Visible flicker from invisible patterns. *Nature*, **362**. 256-258.

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

Marr, D., & Poggio, T. (1976) Cooperative computation of stereo disparity. *Science*, **194**, 283-287.

Marrara, M.T., & Moore, C.M. (2000) Role of perceptual organization while attending to depth. *Perception and Psychophysics*, **62**, 786-799.

Mattingley, J.B., Davis, G., & Driver, J. (1997) Preattentive Filling-in of Visual Surfaces in Parietal Extinction. *Science*, **275**, 671-674.

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

McCarley, J.S., & He, Z.J. (2000) Asymmetry in 3-D perceptual organization: Ground-like surface superior to ceiling-like surface. *Perception and Psychophysics*, **62**, 540-549.

McCarley, J.S.; He, Z.J. (2001) Sequential priming of 3-D perceptual organization. *Perception and Psychophysics*, **63**, 195-208.

McLeod, P., Driver, J., & Crisp, J. (1988) Visual search for a conjunction of movement and form is parallel. *Nature*, **332**, 154-156.

Meese, T.S. (1995) Using the standard staircase to measure the point of subjective equality - A guide base on computer simulations. *Perception and Psychophysics*, **53**, 267-281.

Meese, T.S., & Anderson, S.J. (2003) Spiral mechanisms are required to account for summation of complex motion components. *Vision Research*, **42**, 1073-1080.

Meese, T.S., & Harris, M.G. (1997) Computation of surface slant from optic flow: Orthogonal components of speed gradient can be combined. *Vision Research*, **37**, 2369-2379.

Meese, T.S., & Harris, M.G. (2001a) Broad direction bandwidths for complex motion mechanisms. *Vision Research*, **41**, 1901-1914.

Meese, T.S., & Harris, M.G. (2001b) Independent detectors for expansion and rotation, and for orthogonal components of deformation. *Perception*, **30**, 1189-1202.

Meese, T.S., Harris, M.G., & Freeman, T.C.A. (1995) Speed Gradients and the Perception of Surface Slant: Analysis is Two Dimensional not One-dimensional. *Vision Research*, **35**, 2879-2888.

Meese,T.S., & Holmes, D.J. (2004) Performance data indicate summation for pictorial depth-cues in slanted surfaces. *Spatial Vision*, **17**, 127-151.

Meese, T.S., & Williams, C.B. (2000) Probability summation for multiple patches of luminance modulation. *Vision Research*, **40**, 2101-2113.

Miller, J. (1982) Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, **14**, 247-279.

Moore, C.M., & Egeth, H.E. (1997) Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, **23**, 339-352.

Moore, C.M., Elsinger, C.L., & Lleras, A. (2001) Visual attention and the apprehension of spatial relations: The case of depth. *Perception and Psychophysics*, **63**, 595-606.

Moran, J., & Desimone, R. (1985) Selective attention gates visual processing in the extrastriate cortex. *Science*, **229**, 782-784.

Morgan, M., Ward, R.M., & Castet, E. (1998) Visual search for a tilted target: tests od spatial uncertainty models. *Quarterly Journal of Experimental Psychology*, **51**, 347-370.

Morita, H., & Kumada, T. (2003) Effects of pictorially-defined surfaces on visual search. *Vision Research*, **43**, 1869-1877.

Morrone, M.C., Burr, D.C., & Pietro, S.D. (1999) Cardinal directions for visual optic flow. *Current Biology*, 9, 763-766.

Morrone, M.C., Burr, D.C., & Vaina, L.M. (1995) Two stages of visual processing for radial and circular motion. *Nature*, **376**, 507-509.

Motter, B.C. (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, **70**, 909-919.

Nakayama, K. (1990) The iconic bottleneck and the tenuous link between early visual processing and perception. In Blakemore, C. (Eds), *Vision: coding and efficiency*, Cambridge University Press.

Nakayama, K. & Joseph, J.S. (1998) Attention, Pattern Recognition, and Pop-Out in Visual Search. In Parasuraman, R., *The Attentive Brian*, 279-297, MIT Press, London.

Nakayama, K; Shimojo, S.; Silverman, G.H. (1989) Stereoscopic depth: Its relation to image segmentation, grouping and the recognition of occluded objects. *Perception*, **18**, 55-68.

Nakayama, K, & Silverman, G.H. (1986) Serial and parallel processing of visual feature conjunctions. *Nature*, **320**, 264-265.

Newsome, W.T., & Pare, E.B.A. (1988) A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*. **8**. 2201-2211.

Olzak, L.A., & Thomas, J.P. (1981) Gratings: Why frequency discrimination is sometimes better than detection. *Journal of the Optical Society of America*, **71**, 64-70.

Orban, G.A. (1997) Visual processing in Macaque area MT/V5 and its satellites (MSTd and MSTv). In Rockland, K.S., Kaas, J.H., & Peters, A. (Eds), *Extrastriate cortex in primates*, 359-434. Plenum Press. London.

Pack, C.C., Berezovskii, V.K., & Born, R.T. (2001) Dynamic properties of neurons in cortical area MT in alert and anaethetised macaque monkeys. *Nature*, **414**, 905-908.

Palmer, J., Ames, C.T., & Lindsey, D.T. (1993) Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, **19**, 108-130.

Palmer, J. (1994) Set-Size Effects in Visual Search: the Effect of Attention is Independent of the Stimulus for Simple Tasks. *Vision Research*. **34**, 1703-1721.

Parkes, L., Lund, J., Angelucci, A., Solomon, J.A., & Morgan, M. (2001) Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, **4**, 739-744.

Pashler, H.E. (1987) Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception and Psychophysics*, **41**, 191-201.

Pashler, H.E. (1998) The Psychology of Attention. MIT Press. London.

Pelli, D.G. (1985) Uncertainty explains any aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, **2**, 1508-1532.

Poggio,G.F., Gonzalez,F., & Krause, F. (1988) Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *Journal of Neuroscience*, **8**, 4531-4550.

Posner, M.I.; Snyder, C.R.R.; Davidson, B.J. Attention and the detection of signals. *Journal of Experimental Psychology: General*, **109**, 160-174.

Quick, R.F. (1974) A Vector-Magnitude Model of Contrast Detection. *Kybernetik*, **16**, 65-67.

Ramachandran, V.S.; Cavanagh, P. (1985) Subjective contours capture stereopsis. *Nature*, **317**, 527-531.

Raymond, J.E. (1993) Movement direction analyzers: independence and bandwidth. *Vision Research*, **33**, 767-775.

Regan, D. (1993) The divergence of velocity and visual processing. *Perception*, **22**, 497-499.

Regan, D. & Beverley, K.I. (1978) Looming detectors in the human visual pathway. *Vision Research*, **18**, 415-421.

Regan, D. & Beverley, K.I. (1979) Visually guided locomotion: psychophysical evidence for a neural mechanism sensitive to flow patterns. *Science*, **205**, 311-313.

Remington, R.W., Johnston, J.C., & Yantis, S. (1992) Involuntary attentional capture by abrupt onsets. *Perception and Psychophysics*. **51**, 279-290.

Rensink, R.A., & Enns, J.T. (1995) Preemption effects in visual search: Evidence for low-level grouping. *Psychological Review*, **102**, 101-130.

Rensink, R.A., & Enns, J.T. (1998) Early completion of occluded objects. *Vision Research*, **28**, 2489-2505.

Robson, J.G., & Graham, N. (1981) Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, **21**, 409-418.

Rodman, H.R., & Albright, T.D. (1987) Coding of visual stimulus velocity in area MT of the macaque. *Vision Research*, **27**, 2035-2043.

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

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

Sachs, M.B., Nachmias, J., & Robson, J.G. (1971) Spatial-Frequency Channels in Human Vision. *Journal of the Optical Society of America*, **61**, 1176-1186.

Salzman, C.D., Britten, K.H., & Newsome, W.T. (1990) Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, **346**, 174-177.

Scase, M.O., Braddick, O.J., & Raymond, J.E. (1996) What is Noise for the Motion System? *Vision Research*, **36**, 2579-2586.

Shiffrin, R.M., & Gardner, G.T. (1972) Visual processing capacity and attentional control. *Journal of Experimental Psychology*, **93**, 72-83.

Shiffrin, R.M., & Schneider, W. (1977) Controlled and automatic information processing: II. Perception, learning, automatic attending and a general theory. *Psychological Review*, **84**, 127-190.

Snowden, R.J. (1994) Motion processing in the primate cerebral cortex. In Smith, A.T., & Snowden, R.J. (Eds), *Visual detection of motion*, 307-332. Academic Press, London.

Snowden, R.J., & Milne, A.B. (1996) The effects of adapting to complex motions: position invariance and tuning to spiral motions. *Journal of Cognitive Neuroscience*, **8**, 435-452.

Sperling, G. (1960) The information available in brief visual presentations. *Psychological Monographs: General and Applied*. **74**. 1-29.

Styles, E.A. (1998) *The psychology of attention*. Psychology Press. Hove.

Sugita, Y. (1999) Grouping of image fragments in the primary visual cortex. *Nature*, **401**, 269-272.

Susak, C, & Cavanagh, P. (1995) Facial organization blocks access to low-level features: An object inferiority effect. *Perception and Psychophysics*, **21**, 901-913.

Steinman, S.B. (1987) Serial and parallel search in pattern vision? *Perception*, **16**, 389-398.

Takeuchi, T. (1997) Visual search of expansion and contraction. *Vision Research*, **37**, 2083-2090.

Tanaka, K., Hikosaka, H., Saito, H., Yukie, Y., Fukada,Y., & Iwai, E. (1986) Analysis of local wide-field movements in the superior temporal visual area of the macaque monkey. *Journal of Neuroscience*, **6**, 134-144.

Theeuwes, J., & Kooi, J.L. (1994) Parallel search for a conjunction of shape and contrast polarity. *Vision Research*, **34**, 3013-3016.

Thornton, T., & Gilden, D.L. (2001) Attentional Limitations in the Sensing of Motion Direction. *Cognitive Psychology*, **43**, 23-52.

Treue, S. (2003) Climbing the cortical ladder from sensation to perception. *Trends in Cognitive Science*. **7**, 469-471.

Treue,S., & Andersen, R.A. (1996) Neural responses to velocity gradients in macaque cortical area. *Visual Neuroscience*, **13**, 797-804.

Treue,S., & Trujillo, J.C.M. (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, **399**. 575-579.

Triesman, A., & Gelade, G. (1980) A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.

Triesman, A., & Gormican, S. (1988) Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, **95**, 15-48.

Treisman, A., & Sato, S. (1990) Conjunction search revisited. *Journal of Experimental Psychology: Human Pereception and Performance*, **16**, 459-478.

Tyler, C.W., & Chen, C-C. (2000) Signal Detection Theory in the 2AFC paradigm: attention, channel uncertainty and probability summation. *Vision Research*, **40**, 3121-3144.

Wandell, B.A. (1995) *Foundations of vision*. Sinauer Associates. Sunderland.

Wang, Q., Cavanagh, P., & Green, M. (1994) Familiarity and pop-out in visual search. *Perception and Psychophysics*. **56**, 495-500.

Ward, R., & McClelland, J.L. (1989) Conjunction search for one and two identical targets. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 664-672.

Warner, C.B., Juola, J.F., & Koshino, H. (1990) Voluntary allocation versus automatic capture of visual attention. *Perception and Psychophysics*, **48**, 243-251.

Watamaniuk, S.N.J., Sekuler, R., & Williams, D.H. (1989) Direction perception in complex dynamic displays - the integration of direction information. *Vision Research*, **29**, 47-59.

Watanabe, T. (1995) Motion aftereffects show that attention may selectively activate local motion units located in V1. *Investigative Ophthalmology and Visual Science*, **36**, 634.

Watanabe, T. (1997) Velocity decomposition and surface decomposition - reciprocal interactions between motion and form processing. *Vision Research*, **37**, 2879-2889.

Watanabe, T., & Miyauchi, S. (1998) Roles of attention and form in visual motion processing: Psychophysical and brain imaging studies. In Watanabe, T. (Eds), *High-level motion processing: Computational, neurobiological, and psychophysical perspectives*. 95-113, MIT Press. London.

Watanabe, T., Sasaki, Y., Miyauchi, S., Putz, B., Fujimaki, N., Nielsen, M., Takino, R., & Miyakawa, S. (1998). Attention-regulated activity in human primary visual cortex. *Journal of Neurophysiology*, **79**, 2218-2221.

Watson, A.B. (1979) Probability summation over time. *Vision Research*, **19**, 515-522.

Welch, L. (1989) The perception of moving plaids reveals two motion-processing stages. *Nature*, **337**, 734-736.

Wetherill, G.B., & Levitt, H. (1965) Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology*, **18**, 1-10.

Wilson, H.R., & Wilkinson, F. (1998) Detection of global structure in Glass patterns: implications for form vision. *Vision Research*. **38**, 2933-2947.

Wilson, H.R, Wilkinson, F., & Asaad, W.(1997) Concentric Orientation Summation in Human Form Vision. *Vision Research*, **37**, 2325-2330.

Wolfe, J.M. (1998) *Visual Search*. In Pashler, H.E. (Eds), *Attention*, 13-73, Psychology Press. Guildford.

Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989) Guided Search: An alternative to the Feature Integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 419-433.

Wolfe, J.M., & Franzel, S.L. (1988) Binocularity and visual search. *Pereception and Psychophysics*, **44**, 81-93.

Wolfe, J.M., Yu,K. P., Stewart, M.I., Shorter, A.D., & Friedman-Hill, S.R., & Cave, K.R. (1990) Limitations on the parallel guidance of visual search: Colour x colour and orientation x orientation conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 879-892.

Wuerger, S.M., Hofbauer, M., & Meyer, G.F. (2003) The integration of auditory and visual motion signals at threshold. *Perception and Psychophysics*, **65**, 1188-1196.

Xiao, D.K., Marcar, V.L., Raiguel, S.E., & Orban, G.A. (1997) Selectivity of macaque MT/V5 neurons for surface orientation in depth specified by motion. *European Journal of Neuroscience*, **9**. 956-964.

Xiao, D.K., Raiguel, S., Marcar, V., Koenderink, J., & Orban, G.A. (1995) Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area. *Proceedings of the National Academy of Sciences of the USA*, **92**, 11303-11306.

Yager, D., Kramer, P., Shaw, M., & Graham, N. (1984) Detection and identification of spatial frequency: Models and data. *Vision Research*, **24**, 1021-1035.

Zeki, S.M. (1974) Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of* Physiology, **236**, 827-832.

Zeki, S.M., Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology*, **236**, 827-832.

Zhang, K., Sereno, M.I., & Sereno, M.E. (1993) Emergence of Position-independent Detectors of Sense of Rotation and Dilation with Hebbian Learning: An Analysis. *Neural Computation*, **5**, 597-612.

Zipser, K., Lamme, V.A.F., & Schiller, P.H. (1996) Contextual modulation in primary visual cortex. *Journal of Neuroscience*, **16**, 7376-7389.

Appendix 1: The complex motion transforms

Transform for spiral group motion

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

Exp ($\theta = 0°$); Contr ($\theta = 180°$); Clk-Rot ($\theta = 90°$) CClk-Rot ($\theta = -90°$)

Transform for deformation group

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

Def1 ($\theta = 90°$); -Def1 ($\theta = -90°$); Def2 ($\theta = 0°$); -Def2 ($\theta = 180°$)

Appendix 2: Derivation of the psychometric function from assumptions of probabilistic summation

First the psychometric function for a single analyser is considered. To relate the stimulus driven activity, R to the signal strength, c and analyser sensitivity S, some assumption needs to be made about the function: $R = f\left(c \cdot S_1\right)$. For the present, R is assumed to be a linear function of $c \cdot S$ (i.e. the analyser has a linear transducer function). The psychometric function is therefore

$$\psi = 1 - e^{-(c \cdot S)^\beta} \tag{A2.1}$$

The probability of an observer not detecting can be expressed as the product of the probabilities associated with each analyser being monitored not detecting.

$$1 - \psi = \prod_{i=1}^{n} (1 - \psi_i) \tag{A2.2}$$

Substituting (A2.1) into (A2.2):

$$1 - \psi = \prod_{i=1}^{n} \left[1 - 1 - e^{-\left(c_1 \cdot S_1\right)^\beta} \right] \tag{A2.3}$$

Simplifying to:

$$\psi = 1 - e^{-\sum_{i=1}^{n} \left(c_i \cdot S_i\right)^\beta} \tag{A2.4}$$