The Role of Binocular Disparity and Motion Parallax Information in the Perception of Depth and Shape of Physical and Simulated Stimuli.

Andrew D. Parton

Ph.D. Psychology

2000
Abstract

A series of experiments is reported that examined the perception of the depth structure of a visual scene on the basis of binocular disparity and motion parallax information. Initial experiments (2.1-2.4) revealed that there are considerable differences in the perception of depth in computer simulated surfaces specified by each cue individually. These differences were interpreted as indicating a variation in the relative sensitivity of the visual system to different components of the geometric transformations generated between retinal images within the two domains. Subsequent experiments assessed observers' perception of depth, on the basis of disparity and/or parallax information, in configurations of point light sources in a dark limited cue environment viewed (i) directly (experiments 3.1, 3.2, 3.3 and 5) and (ii) through a head mounted CRT display (experiments 4.1-4.3). They showed that in these viewing conditions the visual system cannot derive a metric (Euclidean) representation of scene structure but it has access to a range of strategies (and consequently representations) that may enable it to complete tasks accurately. The strategy used depended upon the range of available information sources, an assessment of their reliability and the nature of the experimental task. Two final experiments examined the effect of increasing the available depth cues by (i) performing a task in an illuminated structured viewing environment and (ii) introducing surface texture cues. They showed that biases in depth perception persist in such environments and that they cannot be entirely explained by conflicting depth information signalled by accommodation (Frisby et al, 1996). It is argued that strong fusion models of depth cue interaction best describe the range of interactions found across of the all
experiments. However, there are limitations on the types of strong interaction used by the visual system, i.e. no evidence was found for Richards' (1985) proposal that the simultaneous presence of disparity and motion information allow the recovery depth structure.
Table of Contents

Title Page 1
Abstract 2
Table of Contents 4
Acknowledgements 10

Chapter 1
The Derivation of Depth from Binocular Disparity and Motion Parallax Information.

1.1 Introduction 11
1.2 The Interaction between Different Sources of Depth Information 12
1.3 The Geometry of Binocular Vision and Structure from Motion 16
  1.3.1 Binocular Vision and the Derivation of 3-d Structure 16
  1.3.2 Motion Parallax and the Derivation of 3-d Structure 21
  1.3.3 Binocular Disparity and Motion Parallax Compared 24
1.4 The Structure of Visual Space on the Basis of Binocular Disparity and Motion Information 28
  1.4.1 Mathematical Models of Visual Space 33
1.5 Combining Binocular Disparity and Motion Information to Retrieve 3-d Scene Structure 34
  1.5.1 A Theoretical Model of Disparity - Motion Combination 34
  1.5.2 Empirical Evidence of Disparity - Motion Combination 36
1.6 Outline of the Thesis 37
Chapter 2
The Effect of Surface Orientation on the Perception on Corrugated Surfaces Specified by Binocular Disparity or Motion Parallax Information.

2.1 Introduction 41

2.1.1 Previous Studies of the Depth Anisotropy 42
2.1.2 Objectives of the Current Experiments 49

2.2 Experiment 2.1 - Binocular Disparity Detection Threshold Task 51

2.2.1 Method 52

2.2.1.1 Observers 52
2.2.1.2 Stimuli 52
2.2.1.3 Apparatus 53
2.2.1.4 Design and Procedure 53

2.2.2 Results and Discussion 54

2.3 Experiment 2.2 - Binocular Disparity Depth Matching (Suprathreshold) Task 59

2.3.1 Method 59

2.3.1.1 Observers 59
2.3.1.2 Stimuli and Apparatus 60
2.3.1.3 Design and Procedure 60

2.3.2 Results and Discussion 61

2.4 Experiment 2.3 - Motion Parallax Detection Threshold Task 66

2.4.1 Method 67

2.4.1.1 Observers 67
2.4.1.2 Stimuli 67
Chapter 3

The Task Dependant Use of Binocular Disparity and Motion Parallax Information with Physical Stimuli.

3.1 Introduction 85

3.2 Experiment 3.1 – Nulling Task 89

3.2.1 Method 90

3.2.1.1 Observers 90

3.2.1.2 Apparatus 90

3.2.1.3 Design and Procedure 91

3.2.2 Results and Discussion 92

3.3 Experiment 3.2 - Shape Task 95

3.3.1 Method 96

3.3.1.1 Design and Procedure 96

3.3.2 Results and Discussion 96
3.4 Experiment 3.3 - Depth Matching Task

3.4.1 Method

3.4.1.1 Apparatus

3.4.1.2 Design and Procedure

3.4.2 Results and Discussion

3.5 General Discussion

Chapter 4

The Design of Telepresence Systems: An Application of the Task-Dependant use of Binocular Disparity and Motion Parallax Information.

4.1 Introduction

4.2 The Telepresence Equipment

4.2.1 The Head Mounted Display.

4.2.2 The Cameras

4.2.3 Visual Acuity Using the Telepresence Equipment

4.3 Experiment 4.1 - Telepresence Nulling Task

4.3.1 Method

4.3.1.1 Observers

4.3.1.2 Apparatus

4.3.1.3 Design and Procedure

4.3.2 Results and Discussion

4.4 Experiment 4.2 - Telepresence Equilateral Triangle Task

4.4.1 Method

4.4.1.1 Design and Procedure
4.4.2 Results and Discussion 132
4.5 Experiment 4.3 - Telepresence Depth Matching Task 135
4.5.1 Method 136
  4.5.1.1 Observers 136
  4.5.1.2 Design and Procedure 136
4.5.2 Results and Discussion 137
4.6 General Discussion 141

Chapter 5
Perceived Size and Depth on the Basis of Binocular Disparity and Motion Parallax Information.

5.1 Introduction 146
5.2 Experiment 6 – Size and Depth Task 150
  5.2.1 Method 150
     5.2.1.1 Observers 150
     5.2.1.2 Apparatus and Stimuli 150
     5.2.1.3 Design and Procedure 151
  5.2.2 Results 152
  5.2.3 Discussion 167

Chapter 6
The Recovery of the 3-d Structure of Physical Stimuli from Binocular Disparity Information within an Illuminated Structured Environment.

6.1 Introduction 172
6.2 Experiment 6.1 – Illuminated Structured Environment Task

6.2.1 Method

6.2.1.1 Observers

6.2.1.2 Apparatus

6.2.1.3 Design and Procedure

6.2.2 Results

6.2.3 Discussion

6.3 The Derivation of Surface Slant from Texture Information

6.4 Experiment 6.2 – Textured Surface Task

6.4.1 Method

6.4.1.1 Observers

6.4.1.2 Apparatus and Stimuli

6.4.1.3 Procedure

6.4.2 Results

6.4.3 Discussion

6.5 General Discussion

Chapter 7

General Summary and Conclusions

References
I would like to express my thanks to a number of people who have provided help and support during the production of this thesis:

Dr. Mark Bradshaw, my primary supervisor, to whom I owe a considerable debt of gratitude. This thesis would certainly not exist without his enthusiasm, insight and guidance.

Dr. Paul Hibbard whose critical comments have contributed considerably to the early chapters of the thesis.

Professor Ian Davies and the Department of Psychology at the University of Surrey for providing a departmental studentship which supported me during the writing of this thesis.

Finally, the following who all contributed time, ideas or equipment to the experimental work in the thesis: Suzy Adamson, Chris Blair, Charley Clarke, Moira Dean, Dave Field, Jo Graham, Julia Hamon, Jorg Huber, Pete Kearton, Rachael Maguire, Julia Noakes, James Parton, John Pretlove, Penny Rolling, Terry Thacker, Lada Tijomevievic, Heather Tomlinson, Si Watt, Denise Welles, Alison Wheeler, Marian White, Rob van der Willigen and Tasha Young.
CHAPTER 1

The Derivation of Depth from Binocular Disparity and Motion Parallax Information.

1.1 Introduction

When an observer views a 3-d scene their visual percept of the depth, and spatial layout, of objects within the scene can potentially be derived from a number of different sources of information (e.g. Cutting & Vishton, 1995). These depth cues include binocular disparities (the differences in the positions of objects in the two retinal images), the angle of convergence of the two eyes, relative motion created by observer or object movement (structure from motion), changes in the accommodative state of the eyes, and pictorial cues such as perspective, occlusion and changes in surface texture density. The information which observers can derive from each of these depth cues varies in both its (i) qualitative nature and (ii) precision (see below). Consequently, there has been a considerable amount of research examining the theoretical and practical limits of individual depth cues and the extent, if at all, to which they interact. In particular, a large number of studies have examined the extent to which observers can accurately recover the depth structure of a 3-d scene on the basis of individual or groups of cues.

The principal aim of the experiments in the current thesis is to examine the limits and interaction of binocular disparity and motion parallax information in the perception of depth. Both cues have been shown in isolation to evoke a powerful sensation of the
3-d structure of a visual scene (Julesz, 1971; Rogers & Graham, 1979) but despite considerable amount of research there is still not a consensus on which components of the retinal images are utilised to compute depth from these cues. There is, however, considerable interest in comparing the limits of these cues and examining their potential interaction because of the similarities in the geometric nature of the information which they specify (see section 1.2) and the way in which they are processed by the visual system (Graham & Rogers, 1982; Rogers & Graham, 1982; 1983; Bradshaw & Rogers, 1996).

1.2 The Interaction between Different Sources of Depth Information

There are a number of possible reasons why the visual system makes use of a range of different information sources to derive the 3-d structure of a scene. Firstly, Todd (1985) pointed out that perceiving the layout of a scene could be of critical importance and that fault tolerance within the available information guards against failure of a specific information source, or a failure of the assumptions on which its derivation is based. Secondly, as noted earlier, different depth cues vary in the qualitative nature of the information that they specify about a 3-d scene. For example, some cues (e.g. convergence, accommodation and vertical binocular disparities) can, in principal, specify the egocentric (absolute) distance from a fixated object to the observer (see for example Foley, 1980). In contrast, other depth cues can detail information about the exocentric (relative) depth of different objects within a scene, either specifying (i) the depth between objects up to a scaling factor (e.g. horizontal binocular disparities), (ii) the relative depth between objects but not which is closer
(e.g. motion parallax information and accommodative blur) or (iii) merely distinguishing the depth order of objects (e.g. occlusion). Thirdly, even though a particular depth cue, or combination of depth cues, may theoretically specify certain information about the depth relationships within a scene, practical considerations may affect an observer's ability to utilise it. For example, difficulties registering precisely the position of the eyes will impose limits on the precision of distance information derived from the angle of vergence (Foley, 1980). Fourthly, Cutting and Vishton (1995) have argued that the relative utility of different cues varies with the viewing environment. Specifically, they argued that the relative effectiveness of different cues varies with the absolute viewing distance. For example, Pentland (1987) has demonstrated that accurate estimates of depth and distance can be recovered by measuring the amount of blur within an image. However, Mather (1997) noted that due to imperfections in the human lens there was a physical limitation on the minimum depth difference that could be measured and that this varied as a function of distance. Additionally, certain cues are ineffective under specific viewing conditions, e.g. in a darkened viewing environment accommodation will tend to be in a resting state (Liebowitz & Owens, 1975). For all of these reasons the different sources of depth information can be regarded as complementary and an observer’s percept of a 3-d scene is derived from the interaction of the different cues when they are processed by the visual system.

Howard and Rogers (1995) outline five ways in which depth cues might, in principal, interact: cue averaging (the explicit combination of two cues by taking the, possibly weighted, mean of the depth magnitude specified by each cue), cue dominance (one cue suppresses the information of another cue), cue dissociation (each cue is attributed
to a distinct object), cue reinterpretation (information from one cue is reinterpreted in order to make it compatible with information from another cue) and cue disambiguation (one cue provides information which is absent from another cue).

Several researchers (Dosher, Sperling & Wurst, 1986; Bruno & Cutting; 1988; Bülthoff & Mallott, 1988; Nakayama & Shimojo, 1992; Landy, Maloney, Johnston & Young, 1995) have proposed possible models of depth cue interaction which implicitly, or explicitly, utilise processes equivalent to those outlined by Howard & Rogers. Clark & Yuille (1990) argued that depth cue combination models could be classified on a continuum between two extreme categories: (i) weak and (ii) strong fusion. In models tending towards weak fusion, depth cues are initially processed by independent mechanisms, which do not interact, and the output of these modules is then combined (generally by cue averaging). In strong fusion models, cues are processed simultaneously and provide a set of constraints, which specify the most probable interpretation of the depth structure of a scene. When the information specified by two, or more cues, is both qualitatively and quantitatively similar each type of fusion model tends to make similar predictions, e.g. linear averaging. However, strong fusion models may also predict non-linear interactions (e.g. cue dominance) for discrepant information. Though, Howard and Rogers (1995) have questioned the relevance of modelling situations that do not generally occur in natural viewing situations.

Landy et al (1995) criticised the extreme forms of both of these models of depth cue combination. They argued that weak fusion models ignored the difference between the type of information that the various cues specify. For example, because the nature of the information specified by binocular disparities and occlusion is qualitatively
different they cannot be merely averaged. They criticised certain strong fusion
models (e.g. Nakayama & Shimojo, 1992) for putting no constraints upon the
potential interactions between the different depth cues. As a consequence, they claim
that such models become arbitrarily complex and impossible to test. As an example
they noted that Cutting, Bruno, Brady & Moore (1992) have demonstrated that
Massaro’s (1988) fuzzy logic model of perception would fit even randomly generated
data. However, it should be noted they do not criticise all strong fusion models
merely those which do not clearly specify the possible interactions between different
cues.

Therefore, Landy et al advanced a hybrid model, which they termed modified weak
fusion. They proposed that there could be only be a limited interaction within the
early processing of depth cues. Specifically, they hypothesised a process called cue
promotion in which cues interact but only to access missing, or clarify ambiguous,
information. In effect promotion includes both the processes of cue reinterpretation
and cue disambiguation outlined earlier. This promoted information allows each
depth cue processing module to generate a map of the depths of different points
within a scene specified in common units. These maps can then be combined by
averaging the cues weighted according to an assessment of their reliability within the
current viewing situation. The model also includes the process of cue dominance as it
is presumed that any cue highly discrepant from the other cues would by weighted to
indicate its unreliability and so have little, or no effect, upon the final interpretation of
the perceived depth structure of the scene.
1.3 The Geometry of Binocular Vision and Structure from Motion

In order to understand the degree to which binocular disparity and motion information (alone and in combination with other cues) can specify the depth structure of a 3-d scene it is necessary to appreciate the geometric properties of the information available from these cues. Therefore, this section outlines the information that can theoretically be derived from each of these cues, the potential limitations of this information in specifying the depth structure of a scene and the possible sources of additional information that might be used to overcome these limitations.

1.3.1 Binocular Vision and the Derivation of 3-d Structure

When observers look at a 3-d scene binocularly, the physical separation of their two eyes means they view the scene from two slightly different vantage points. As a consequence, there are variations in the images projected to each retina. Part of the visual field will be projected to the images on both retina (the binocular region) and part to only one retina (the monocular region). Each point on the binocular region of one retina has a corresponding point on the other retina. If part of an image falls on the corresponding points on the two retina then it is said to have zero binocular disparity and portions of the image which fall at slightly different points on the two images are disparate. All points with a zero disparity occupy a location in space on a surface referred to as the horopter.

It is possible to derive a theoretical structure for the horopter in the horizontal plane of the eyes by making certain assumptions regarding the location of corresponding
points on the retina and using the geometric constraints of the viewing environment. Firstly, we assume that corresponding points on the retina lie at the same horizontal distance from the fovea. Secondly, Vieth (1818 cited in Ogle, 1950) and Müller (1826 cited in Ogle, 1950) noted that the angles between any two points on a circle and any other two points were always equal. Therefore when considering points lying in the same horizontal plane as an observer's eyes, the horopter would theoretically fall on a circle (known as the Vieth-Müller circle) which passes through the nodal points of the eyes and the point of fixation. However, Ogle (1950) reviewed a number of studies that demonstrated when the horopter was derived empirically (from observers' perception of the layout of a visual scene) it departs systematically from the Vieth-Müller circle in a manner dependant upon the viewing distance. For example, in one task observers were asked to set the position of a series of rods lying symmetrically around a central fixation point to form a fronto-parallel plane. Their settings formed a concave arrangement at close viewing distances, a fronto-parallel planar surface for a limited range of intermediate distances and a convex arrangement at further distances. Ogle (1932, 1950) also argued that the geometry of the empirical horopter could be specified using two parameters, $H$ the curvature at the point of fixation and $R_0$ the skew of this curve. Ogle's work inspired considerable interest in providing a formal mathematical model of the geometry of visual space (e.g. Luneburg, 1947; Blank, 1953).

Points in space that lie at different points in depth to the horopter create horizontal differences between the two retinal images that are referred to as horizontal binocular disparities. Horizontal disparities are sufficient, even when presented in isolation, to
evoke a powerful sensation of the depth structure of an object, or scene (Julesz, 1971).

Figure 1.1 illustrates the geometrical definition of horizontal binocular disparity, for symmetrical convergence, which is expressed in angular terms. The eyes are fixated on point F that means they converge such that the image of this point falls on the fovea on both retinas. P is a point lying in depth beyond the fixation point. The visual angle between P and F in the left image is denoted by \( \phi_l \) and in the right image by \( \phi_r \). The angular disparity (\( \eta \)) between F and P is equal to \( \phi_l - \phi_r \), which is equivalent to \( \theta_F - \theta_P \).

The values of \( \theta_F \) and \( \theta_P \) can be derived using basic rules of geometry as follows:

\[
\frac{\theta_F}{2} = \tan\left(\frac{I}{2D}\right) \tag{1}
\]

\[
\frac{\theta_P}{2} = \tan\left(\frac{I}{2(D + \delta)}\right) \tag{2}
\]

For small angles, the tangent of an angle is approximately equal to the angle in radians. Hence, the disparity is equal to:

\[
\eta \approx \theta_F - \theta_P \approx \frac{I}{D} - \frac{I}{D + \delta} \approx \frac{I\delta}{D(D + \delta)} \tag{3}
\]

If, however, the separation between the points is small relative to the viewing distance the angular disparity between two points is very closely approximated by:

\[
\eta \approx \frac{I\delta}{D^2} \tag{4}
\]
Figure 1.1 The angles and distances used to define horizontal binocular disparity of a point \( P \) with respect to the fixation point \( F \). The rest of the notation is as follows: \( D \) is the distance from the midpoint of the eyes to \( F \), \( \delta \) is the distance from \( F \) to \( P \), \( I \) is the interocular separation, \( \theta_P \) is the angle point \( P \) subtends with the nodal points of the eyes and \( \theta_F \) is the angle point \( F \) subtends with the nodal points of the eyes (after Howard & Rogers, 1995).

Hence, the angular disparity created by two points separated in depth by a fixed physical interval is inversely proportional to the square of the distance from the observer's eyes to the point of fixation (see equation 4). Therefore, any fixed 3-d shape defined by a series of points in depth will produce a different pattern of angular disparities at different viewing distances (see figure 1.2). Conversely, a particular pattern of angular disparities on the retina are consistent with a family of shapes at different viewing distances, e.g. a shallow triangle close to the observer or a deeper triangle further away. So, in order to recover the metric (Euclidean) structure, which specifies the absolute magnitude of depth between different points, of a scene it is necessary to supplement the information that is provided by binocular disparities.

One way to recover metric structure, in principal, is to scale binocular disparities by an estimate of the egocentric distance to the fixation point. In normal viewing situations this information might potentially be derived from a number of sources; for example, (i) the vergence angles of the eyes (Foley, 1980), (ii) accommodation (Fisher & Ciufredda, 1988), (iii) vertical binocular disparities (Mayhew & Longuett-
Higgins, 1982; Cumming, Johnston & Parker, 1993; Rogers & Bradshaw, 1993), (iv) linear perspective (O’Leary & Wallach, 1980; Durgin et al, 1995; Norman et al, 1996) or (v) the angular size of a familiar object (O’Leary & Wallach, 1980; Predebon, 1991, 1992, 1993). The extent to which observers can accurately recover estimates of relative distance at different absolute distances (depth constancy) on the basis of these, and other, cues has been the focus of many of experiments (see section 1.4 for a brief review).

Figure 1.2. An illustration of the different patterns of angular disparities created by the same triangular figure at two different distances: (a) $d_1$, and (b) $d_2$.

A further possibility for accurately recovering the depth structure of a scene is that disparities could be calibrated, or constrained, by a source of information other than the absolute distance. For example, Richards (1985) has demonstrated that in principal motion information (see section 1.3) might be used to constrain the possible interpretations of binocular disparity information. A suggestion that is tested empirically within the thesis (see chapters 3, 4 and 5) and is discussed more fully below (see section 1.5).
1.3.2 Motion Parallax and the Derivation of 3-d Structure

When an observer translates with respect to the objects within a 3-d scene, or an object translates with respect to the observer and other objects within the scene, a sequence of images are projected to each retina. The images projected from different points in depth within the scene move at different rates across the retina and form a pattern of relative angular velocities referred to as motion parallax (see Helmholtz, 1909 for an early account). Motion parallax has been demonstrated to be sufficient, even presented in isolation from other cues, to evoke a strong impression of depth structure (Rogers & Graham, 1979).

The geometry of the relative angular velocity between the successive retinal images of distinct objects can be derived in a similar manner to angular disparity (see Graham et al, 1948). If we assume the two retinal images in figure 1.1 mark the end points of the translation of a single eye over time \(t\) and the angle of rotation is small (so \(\theta_r \approx \tan(\theta_r)\)):

\[
\theta_r \approx \frac{I}{D}
\]  

Differentiating and rearranging gives:

\[
d(\theta_r) \approx -\frac{I\delta}{D^2}
\]  

As \(d(\theta_r)\) is equivalent to \(\theta_r - \theta_p\) then:

\[
\theta_r - \theta_p \approx -\frac{I\delta}{D^2}
\]  

Therefore, the difference in angular velocity \(\omega\) is:

\[
\omega \approx \frac{\theta_r - \theta_p}{i} \approx -\frac{I\delta}{D^2 t}
\]  

Consequently, the velocity of the relative angular motion between the objects’ retinal images is inversely proportional to the square of their absolute distance from the
observer (see equation 8). Hence, the geometrical relationships between the relative angular motion of images on a single retina generated by motion parallax is similar to the disparities generated between two retinal images when viewing a scene binocularly (Graham & Rogers, 1982). This is illustrated in figure 1.3a which shows that if a single eye moves through the interocular distance the first and last frame of the motion sequence are identical to the simultaneous images present, on both eyes, when binocularly viewing the same object (see figure 1.1). There are, however, important differences between the phenomena. Motion parallax is a (i) monocular cue and (ii) it generates a continuous sequence of image frames over time, whereas static disparity is an inherently two-frame phenomenon. Richards (1985) argued that for the later reason binocular disparity information is in principal less determinate of 3-d structure than motion information. Different points on a rigid translating surface slanted in depth relative to the fronto-parallel plane of the observer generate relative motion in proportion to their difference in depth and the change in slant of the surface. The change in slant can in principal be derived from perspective changes in the vertical height and horizontal width or the acceleration component of the flow field. The latter can be derived from the instantaneous angular velocities of points on the surface, which are available in a continuous series of views of a translating surface but not from two discrete views (i.e. stereo viewing or two-frame motion). For a surface of known slant the ratio of the relative motion of points to their angular size will specify the depth to size ratio of the points and hence the surface structure. Richards noted that most algorithms for deriving structure from motion were reliant upon the acceleration component of the flow field. However, although motion is theoretically more determinate of 3-d structure Rogers & Collett (1988) reported empirical
evidence that for surfaces specified by disparity and parallax the primary determinant of perceived shape was the disparity information (see section 1.3.3).

Figure 1.3 The final views of a motion sequence for; (a) motion parallax when a single eye, fixating point F, moves through the interocular distance from time T1 to T2 and (b) an object rotates around point R. The images on the retina for the first and last frames of the motion sequence are identical to those present when viewing the image binocularly (see figure 1.1).

The determination of the 3-d structure of an object from a motion sequence is referred to as structure from motion (SFM). This term refers to both motion parallax and object rotation. The latter case is often referred to as the kinetic depth effect (KDE: Wallach & O'Connell, 1953). Figure 1.3b illustrates that two views from a KDE sequence can also be equivalent to the two views available to binocular disparity (see figure 1.1) and that the two SFM cases are almost identical for small rotations (Durgin et al, 1995). Under isolated SFM conditions, Ullman (1979) has proved that three frames provide sufficient information to recover 3-d shape up to an isotropic size scaling. That is relative motion, alone, can specify the depth-to-width ratio of a shape but not its veridical size, e.g. it can determine that a triangular shape is equilateral but not whether it is small, and close to the observer, or large, and further away from the observer.

In the experiments within this thesis, head movements are used to generate the retinal
motion. This situation differs in principle from the case of SFM discussed by Ullman because the angle of rotation of the line of sight (which is equivalent to vergence angle in the case of binocular disparity) between the observer and the object is given by the visual angle subtended by the rotation. Therefore, just as vergence angle can be used to scale the pattern of horizontal disparities to recover full Euclidean structure, this rotation angle (if computed) can be used to scale the retinal motions to recover the same type of information. However, the degree to which this is possible is questionable on both theoretical and practical grounds. First, there is an intrinsic ambiguity about whether retinal motion should be attributed to head or object rotation, and second, the recovery of ego-motion may be impossible in the face of the large number of degrees of freedom involved. Moreover, the same information is available to stereopsis but empirical information suggests that the visual system does not make efficient use of it (see section 1.4). The same may be true in the case of motion parallax particularly when head movements are small (Braunstein et al, 1993; Durgin et al, 1995).

1.3.3 Binocular Disparity and Motion Parallax Compared

The principles underlying motion parallax have been understood for many years (Helmholtz, 1909) but the results of empirical studies into its effectiveness in isolation were often equivocal (e.g. Epstein & Park, 1964; Gogel, 1977). However, Rogers and Graham (1979) noted that within most of the earlier studies observers remained stationary whilst the stimuli translated. As a consequence, they examined the perception of relative depth for simulated surfaces in which the only depth
information present was from the motion of points relative to the translation of the observer’s head. The movement of the dots within a random pattern were linked to the horizontal motion of the observer’s head and simulated the movement of points on a three-dimensional surface. They found that their observers experienced a vivid percept of a solid three-dimensional structure that was qualitatively similar to that produced when viewing random dot stereograms (which present disparity information in isolation from other depth cues). A number of subsequent studies have examined further the relationship between binocular disparity and motion parallax by comparing the detection and processing of the two information sources (e.g. Rogers & Graham, 1982, 1983; Graham & Rogers, 1982; Bradshaw & Rogers, 1996).

Rogers and Graham (1982) examined observers’ depth detection thresholds for horizontally orientated sinusoidal corrugations specified by either motion parallax or binocular disparity information. They found their results were qualitatively similar for both of the information sources across the range of spatial frequencies tested (0.05-1.6 c/deg). Observers’ thresholds varied as a function of the spatial frequency of the surface with the lowest thresholds lying somewhere between 0.3 and 0.5 c/deg with increasing thresholds at both higher and lower spatial frequencies. Rogers and Graham interpreted these results as indicating a close relationship between the processing of the two information sources. Their claim is further supported by the empirical evidence for the existence of multiple channels tuned to the spatial frequency of depth modulations defined by disparity (Schumer & Ganz, 1979; Cobo-Lewis & Yeh, 1994) and motion parallax (Hogervorst, Bradshaw & Eagle, 1999).

Graham & Rogers (1982) found further similarities in the two information sources from an examination of simultaneous and successive contrast effects. They reported
that after a prolonged exposure to a sinusoidal depth corrugation, specified by either of the cues, a flat surface appeared to be a corrugated surface of the opposite phase to the adapting stimuli (successive contrast effects). In a second experiment they found a vertical surface appeared to slope in the opposite direction to the slope of its surround (simultaneous contrast effects). Again, the results were qualitatively similar for both information sources. Graham and Rogers argued that the successive contrast effects indicated that the visual system was sensitive to local contours and gradients of disparity and motion (see also Gillam, Flagg & Finlay, 1984; Gillam, Chambers & Lawergen, 1988). They further argued that the simultaneous contrast effects suggested that both cues demonstrate inhibitory interactions over space.

Bradshaw and Rogers (1996) reported evidence that binocular disparity and motion parallax interact in the computation of relative depth. In their first experiment they found observers were less sensitive to depth within surfaces defined by either cue after adapting to a supra-threshold surface defined by the same or a different cue. Furthermore, in a second experiment they found that the detection thresholds for surfaces defined by both cues were lower than for those specified by either cue in isolation. The improvement in thresholds they found was higher than would be expected from a linear additive model and so was consistent with a strong fusion model of cue combination. They argued that the most parsimonious interpretation of the data was the existence of a pool of neurones sensitive to both sources of information.

Despite considerable similarities in the processing of disparity and parallax information there is evidence that subtle differences in the information the visual system derives from the two cues might exist. Rogers and Collett (1989) investigated
the appearance of surfaces specified by both disparity and parallax information. Specifically, they examined the perception of surfaces in which the two cues specified discrepant amounts of depth. They found that the depth was largely determined by disparity information but that the surface was perceived to rotate in depth around a vertical axis (for parallax generated from horizontal head movements). They argued that this was a result of the visual system selecting an interpretation of the information that minimised the discrepancy between the two cues. Consequently, they suggested that motion parallax should be considered as a source of information about the structure and movement of objects not just their apparent depth.

However, all of these studies concentrated upon surfaces curved, or slanted, around horizontal axes. Rogers and Graham (1983) reported an anisotropy in the perception of 3-d surfaces curved, or slanted, around horizontally and vertically orientated axes defined by either binocular disparity or motion parallax information (see chapter 2 for a detailed description). Consequently, one of the aims of the experiments in chapter 2 is to compare the perception of disparity and parallax defined surfaces at a range of orientations.

The studies discussed within this section have reported empirical evidence for interaction of motion parallax and disparity information. However, they have not directly addressed the extent to which these cues support a veridical perception of depth when presented either alone or in combination. This question is addressed empirically by a number of the experiments within this thesis and is discussed in more detail below.
1.4 The Structure of Visual Space on the Basis of Binocular Disparity and Motion Information

The 3-d structure of a physical scene viewed by an observer can be accurately specified using Euclidean geometry. However, the underlying nature of the representation of physical space formed by an observer (assuming it exists) is less clearly understood. A large number of psychophysical studies have attempted to quantify and specify the properties of visual space. Many studies have focused on the degree to which observers can derive, and access, a veridical representation of the different metric properties of 3-d physical space on the basis of static binocular viewing. Gibson (1979) has criticised the approach of studying static binocular viewing as being unrepresentative of normal dynamic viewing. However, Loomis et al (1992) argued that it was valid for two reasons. Firstly, many visually guided actions are performed on the basis of a static, or nearly static, view of a scene. Secondly, these studies can give an insight into the underlying mechanisms of binocular vision, which can form the basis of models of both static and dynamic viewing.

Norman et al (1996) argued there were two ways in which visual perception could meaningfully be described as Euclidean. Firstly, observers’ underlying representation of visual space (the intrinsic structure) could perfectly preserve all of the spatial relationships present in physical space. Secondly, the mapping between visual and physical space (the extrinsic structure) could relate all of the spatial relationships between the two without any resulting distortion. They note that if a line is rotated, or translated, in Euclidean space that its length will remain invariant. So, it is possible to
assess if the extrinsic structure of visual space is Euclidean by testing observers' ability to correctly match the length of a line at a specific orientation and absolute distance to that of a line at either a different orientation, a different distance or both (Norman & Todd, 1993; Norman et al, 1996). Although some studies have attempted to assess the intrinsic structure of visual space without making reference to the external environment (Blank, 1960; Foley, 1972; Norman et al, 1996) most experiments, including those within the current thesis, have concentrated on the extrinsic structure.

A large number of studies have assessed observers' perception of size (the width, or height, of a stimulus in the fronto-parallel plane) and depth (intervals lying in depth along the observers' line of sight). These experiments have generally involved matching the length of a probe stimulus at one distance to that of a test interval at several different absolute distances to assess the degree of constancy in observers' perceptual judgements. Most researchers have reported that observers' perception of the frontal size of a stimulus is fairly accurate but tends to increase slightly with absolute distance for isolated binocular viewing conditions (Gogel, 1960; Norman et al, 1996), binocular viewed well illuminated multi-cue environments (Holway & Boring, 1941; Carlson, 1962; Baird & Biersdorf, 1967; Wagner, 1985; Norman et al, 1996) and on the basis of SFM information (Norman et al, 1996). In contrast, many investigators have found that depth intervals, of fixed magnitude, appear to become increasingly compressed as their absolute distance increases in isolated binocular viewing conditions (Gogel, 1960; Norman et al, 1996), binocularly viewed well lit multi-cue environments (Gilinsky, 1951; Wagner, 1985; Toye, 1986; Loomis et al, 1992; Norman et al, 1996) and on the basis of SFM information (Todd & Bressan,
1990; Norman et al, 1996). Similar mis-estimates of depth have been found using a
shape judgement task for both disparity (Johnston, 1991; Tittle et al, 1995;
Glennerster et al, 1996) and SFM specified shapes (Durgin et al, 1995, Tittle et al,
1995).

Attempts to explain these finding have generally focused on the scaling of binocular
disparity, or relative motion information, as it has been demonstrated that human
observers are highly sensitive when detecting binocular disparity (Howard, 1919) and
motion parallax (Graham et al, 1948) information. Gogel (1960) modelled the
differences between the perception of size and depth intervals by suggesting that
binocular disparity was scaled by the perceived physical width of the fixated object
divided by the visual angle it subtends at the retina. However, Gogel fails to suggest
how in isolated viewing conditions the perceived size of an object can be derived
without an estimate of absolute distance and, if this is available, why it would not be
used to scale disparities directly.

Foley (1980) reviewed evidence from a range of studies assessing observers' percep-
tion of relative depth at a range of viewing distances within isolated binocular
viewing conditions. He argued that the most parsimonious interpretation of all of
these results was that errors in the perception of relative depth resulted from the
calibration of binocular disparities by a mis-estimation of absolute distance. He
attributed the error in the estimation of distance to a mis-registration of the vergence
angle of the eyes. Following Foley, a number of recent studies have argued that the
mis-estimation of absolute distance, or motion scaling parameters, explain the failure
of constancy in stereoscopic and motion defined shape judgement (Johnston, 1991;

All of these experiments, and those reviewed by Foley (1980), directly assessed observers’ perception of depth at different distances by asking observers to set the relative depth of points according to either an internal or external standard, i.e. setting points to be co-planar with a fixed reference (Foley, 1980) or setting the depth of a cylinder such that it was circular in cross section (Johnston, 1991; Glennerster et al, 1996). Observers’ estimates of the viewing distance were then inferred from these relative settings. However, studies using other experimental techniques for estimating distance and depth have found considerable variations in observers' accuracy. For example, when using direct scaling methods (e.g. magnitude estimation and verbal reports) observers produce more accurate, though frequently still systematically biased, estimates of absolute distance (Baird & Biersdorf, 1967; Da Silva, 1985 Sedgwick, 1986). Da Silva (1985) examined three techniques for estimating distance. He found that observers' estimates varied systematically with the viewing distance and were well described by a power function. However, the exponent of the power function was closer to one for the magnitude estimation technique than for the other methods. The interpretation of the results of studies using verbal reports of distance have been criticised on the grounds that observers may be aware that depth intervals appear increasingly compressed with distance and so make cognitive adjustments to their estimates (Gogel, 1974). Additionally, Loomis et al (1992) presented evidence that observers can accurately perform visually guided actions that require a veridical estimate of egocentric distance and relative depth on the basis of binocular disparity information. Observers performed three action tasks with their eyes closed after
statically viewing a target in an open field; (i) walking to a target between 4 and 12 metres away with their eyes closed, (ii) walking to target as for (i) then walking to a second point either straight ahead or perpendicular to the original and (iii) walking along one of three straight lines whilst continuously pointing their arm in the direction of a fixed target at the side. Performance within all of these tasks was extremely accurate with no systematic biases. However, when observers were required to perform a perceptual judgement (matching the length of a fronto-parallel interval to an interval in depth at several different absolute distances) within the same viewing environment their results indicated a compression of perceived depth, which was proportional to the absolute distance. Loomis et al outlined two possible explanations for their results. Firstly, observers could be correcting for the distortion in perceptual space. As we constantly move around the world we may unconsciously (or consciously) learn to compensate for perceptual distortions in order to perform actions efficiently. Secondly, it has been hypothesised that the control of the motor system and perception of visual space are performed by partially, or wholly, distinct processing systems (Goodale & Milner, 1992). Consequently, the motor system may have access to a more veridical representation.

The experimental tasks used within this thesis employ perceptual judgements of relative depth, shape and intersection angle at different absolute distances. Many previous studies have shown that observers demonstrate systematic biases within these types of tasks (e.g. Todd & Bressan, 1990; Johnston, 1991; Collett et al, 1991; Todd & Norman, 1991; Liter et al, 1994; Todd et al, 1995; Tittle et al, 1995; Rogers & Bradshaw, 1993, 1995a; Bradshaw et al, 1996, 1997; Glennerster et al, 1996; Norman et al, 1996). There are, however, considerable variations in the degree of
constancy demonstrated between these experiments. For example, Johnston (1991) found an almost complete failure of constancy and Glennerster et al (1996), using the same task, found that although observers departed systematically from a veridical percept they did appear to take viewing distance into account. However, Johnston's experiments took place in a dark viewing environment whereas Glennerster et al performed their experiment in a dimly illuminated room with a textured surface in front of their monitors. The variation in constancy between different experiments has led several recent authors (Loomis et al, 1992, Durgin et al, 1995, Glennerster et al, 1996, Frisby et al, 1996) to question the generality of the conclusion that observers have difficulty recovering an Euclidean representation of visual space. They have claimed that constancy is effected by the range of depth cues available within the viewing conditions (Durgin et al, 1995, Frisby et al, 1996), the nature of the experimental task (Glennerster et al, 1996), and artefacts of computer generated stimuli (Frisby et al, 1996). The impact of all of these factors upon observers' assessment of depth is further examined within the experiments in the current thesis.

1.4.1 Mathematical models of visual space

There have been many attempts to provide a mathematical model of visual space. For example, Gibson (1950) claimed visual space was Euclidean, Angell (1974) suggested it was spherical, Luneburg (1947) and Blank (1953) suggested that it was hyperbolic and that it can be described by a Reimanian geometry of constant Gaussian curvature, Indow (1991) argued that the horizontal plane containing the eyes had a Reimanian geometry of non-zero curvature and fronto-parallel planes are structured according to
Euclidean geometry, and Wagner (1985) and Norman et al (1995) that it is an affine-transformed Euclidean space. Wagner argued that the different models were not necessarily contradictory and merely represented the operation of the visual system under different environmental conditions. Consequently, visual space may necessarily have to be described by a range of different geometric models. It also remains possible that the nature of the representation available reflects both the available information and the demands of the task. For example, the survival of an organism may depend on precision in the performance of actions, which themselves require an accurate Euclidean judgement. In contrast, certain distortions in the perceptual system may either be unimportant for survival, or even advantageous if they highlight features of significance. There is, however, no currently generally accepted mathematical model of the geometry of visual space.

1.5 Combining Binocular Disparity and Motion Information to Retrieve 3-d Scene Structure

1.5.1 A Theoretical Model of Disparity-Motion Combination

Richards (1985) demonstrated that the presence of both disparity and motion could be exploited to recover veridical information about surface shape without recourse to additional sources of information (e.g. perspective). This suggestion takes advantage of the fact that although each depth cue is affected by distance, each is affected in a different way. Richards considered small-field stimuli, where orthographic projection is a good approximation to perspective projection, and took the case of SFM, where
the object rotates relative to a stationary observer. As outlined earlier (see section 1.3.2), a sequence of relative retinal motion information can specify that an object belongs to a family of shapes with a known width-to-depth ratio. Similarly (see section 1.3.1), for the case of stereopsis, the pattern of horizontal disparities is consistent with a family of shapes, parameterised by the viewing distance. However, the visual angle of the width of each of the shapes is inversely proportional to the viewing distance and not, as for disparities, the square of the viewing distance. Consequently, the width-to-depth ratio of each the potential shapes specified in binocular viewing decreases with the absolute distance. Therefore, when considered together, these two sources of information provide an intersection of constraints with only one member of the disparity-defined family of shapes that is consistent with the shape specified by the motion information. So, both cues in combination can theoretically specify both the shape and size of the viewed object.

These theoretical considerations have implications for when depth information from motion and disparity are brought together. There are several possibilities. First, in binocular SFM conditions, shape estimates could be computed entirely from motion cues (given at least three views), although size estimation would still require determination either via the intersection of constraints with the disparity information or through the incorporation of additional information. Second, if only two time-separated views are utilised, perceived shape and size should be strongly related (even if incorrect) as both are given by the intersection of constraints from the two depth cues. This is an example of cue promotion suggested by Landy et al (1995) (see section 1.2). Finally, if vergence angle (or its equivalent) is used to scale disparities or retinal motions in the relevant experimental conditions then there is potentially
little need for any interaction to occur between the two systems as Euclidean structure could be computed from either cue alone.

1.5.2 Empirical Evidence of Disparity - Motion Combination

Johnston, Cumming and Landy (1994) assessed Richards’ computational theory empirically and found evidence in support. In their critical experiment (experiment 3) they showed that neither a 2-frame motion sequence nor binocular disparity (2-frame) could support the recovery of veridical shape in an apparently circular cylinder task — so called because subjects were required to select from a sequence of hemicylinders, which differed in elongation along the depth axis, the one with an apparently circular profile. However when both the cues were presented together, in a 2-frame sequence, the perceived shape was close to veridical. This case of strong fusion is consistent with Richards’ theory because in this situation there are two one-parameter families of shapes provided by the two (time-separated) binocular views that again have a single, Euclidean ‘intersection of constraints’. Thus, even when SFM cues are insufficient to support the recovery of shape information accurately, the combination of disparity and motion can lead to the recovery of the correct shape (Johnston et al, did not investigate the perception of size). Therefore, despite the findings that suggest that motion alone rarely determines shape correctly (Liter et al, 1993; and Norman and Todd, 1993) some benefit may still accrue through the combination of disparity and motion. Tittle et al, (1995), however, in a series of experiments did not find any apparent advantage when disparity and motion cues were presented simultaneously. Rather, they found that the combined-cue condition
seemed susceptible to the same distortions as the disparity-alone condition across a range of viewing conditions and tasks. Similarly, Brenner and van Damme (1997) concluded that the visual system operated in a modular way when they investigated whether it could recover metric structure and viewing distance through the combination of disparity, motion and other cues. This issue is investigated within the current thesis in the experiments within chapters 3-5.

1.6 Outline of the Thesis

Binocular disparity and motion parallax have both been demonstrated to be highly effective cues to the 3-d structure of a visual scene (Julesz, 1971; Rogers & Graham, 1979). There are similarities in the geometric nature of the information that the two cues specify (see section 1.3) and the way in which they are processed by the visual system (Graham & Rogers, 1982; Rogers & Graham, 1982; 1983; Bradshaw & Rogers, 1996). However, studies have also shown subtle differences in the way that the visual system employs the information specified by each of these cues (e.g. Rogers & Collett, 1989). As discussed in sections 1.4 and 1.5, many studies have examined the degree to which these cues can alone, or in combination, accurately specify the depth structure of a scene but there are considerable variations in their findings (e.g. Todd & Bressan, 1990; Johnston, 1991; Collett et al, 1991; Todd & Norman, 1991; Rogers & Bradshaw, 1993, 1995a; Liter et al, 1994; Todd et al, 1995; Tittle et al, 1995; Glennerster et al, 1996; Bradshaw et al, 1996, 1997; Norman et al, 1996). The precise reasons for, and significance of, the variations between these studies remains, however, an open question. The aim of the current thesis is to
examine a number of these issues further.

The experiments in chapter 2 address further the similarities and differences in the derivation of depth structure on the basis of binocular disparity and motion parallax information (see section 1.3.3). A considerable number of studies have examined which information sources are used by the visual system as stereoscopic primitives (e.g. Gillam et al, 1984, 1988; Mitchison & McKee, 1990; Gillam & Ryan, 1992; Cagenello & Rogers, 1993; Bradshaw et al, 1999). However, there is little work comparing the fundamental information sources used to derive depth structure from disparity and parallax. Previous studies (e.g. Rogers & Graham, 1982; Rogers & Graham, 1983) have examined the perception of depth across limited range for surfaces types, i.e. surfaces with horizontally and vertically orientated depth discontinuities, in which the geometric information relating the transformations between the images varied. This manipulation allows a quantitative comparison of the relative benefits that each of the two cues gain from the information available within these different geometric transformations. The experiments in chapter 2 extend this work by examining the perception of depth in surfaces specified by both cues across a range of orientations between horizontal and vertical. This manipulation allows a greater range of comparisons in the derivation of depth on the basis of each cue. As a consequence they reveal insights into the similarity and differences in the various information sources to which each cue is sensitive.

The subsequent chapters examine the geometric structure of the representations within visual space that can be specified by disparity and parallax information. They look at the degree to which an Euclidean metric representation can be derived on the basis of these cues alone or in combination. Additionally, they assess a range of potential
strategies that the visual system might use to supplement disparity and parallax information without having to obtain a full metric representation (see chapter 3, 4 and 5). Specifically, they examine whether the nature of the experimental task might influence performance for both cues (Glennerster et al, 1996), Ullman's (1979) theorem for the derivation of shape from motion information and Richards' (1985) proposal for the combination of disparity and motion information. These suggestions are assessed by comparing performance on a range of tasks that theoretically vary in the geometric information required to complete them. Variations in accuracy across these tasks allow an assessment of the potential strategies used to complete them and hence the nature of the representations that are available to the visual system on the basis of these cues. The results of these experiments are partially discussed in relation to the differences in observers' accuracy found between previous studies (see section 1.4).

The studies also examine the effect of the viewing environment upon the accuracy of observers' performance across a range of experimental tasks (see chapters 3, 4, 5 and 6). Firstly, they examine the use of real stimuli rather than computer generated stimuli (see chapters 3 to 6). This enables an assessment of the plausibility of Frisby et al's (1996) suggestion that the departures from depth constancy are caused by conflicting accommodation cues within computer generated stimuli. Secondly, they examine the effect of increasing the range of available cues upon observers' performance (see chapter 6) and assess the specific effects of different sources of additional information, i.e. perspective and surface texture information. Thirdly, the effect of using a remote tele-viewing system (see chapter 4) is assessed across a range of tasks to see whether changes in the viewing environment differentially effects the
strategies employed by the visual system to derive depth, i.e. does it still rely to the
same degree on the cues utilised within direct viewing. This question has important
implications for the transfer of research findings within one viewing environment
(e.g. real world viewing) to another (e.g. telepresence viewing). The implication of
all of these experiments for models of depth cue integration is discussed.

Finally, numerous studies outlined earlier have examined the scaling of size and depth
(e.g. Gogel, 1960; Baird & Biersdorf, 1967; Norman et al, 1996) but there has been
little work directly examining whether these two factors are scaled by the same
estimate of distance (see however, Brenner and van Damme, 1997). This question is
assessed within chapter 5 for both disparity and parallax viewing conditions. In the
parallax domain, this is of interest because it provides a way of further assessing
whether parallax information is scaled by the derivation of scaling parameters or
exclusively by Ullman's scheme. As noted in section 1.5, the later enables shape to be
derived but not size. So, if this scheme were used in isolation a greater variation in
size settings on the basis of motion information alone would be expected than on the
basis of disparity information.
CHAPTER 2

The Effect of Surface Orientation on the Perception on Corrugated Surfaces Specified by Binocular Disparity or Motion Parallax Information.

2.1 Introduction

The perception of depth in disparity-defined surfaces which slope or curve about a horizontal axis has been shown to differ from that for corresponding surfaces which slope or curve about a vertical axis (e.g. Rogers & Graham, 1983). In general, for surfaces orientated around a horizontal axis observers have lower detection thresholds (Rogers & Cagenello, 1989; Mitchison & McKee, 1990; Cagenello & Rogers, 1990, 1993; Bradshaw & Rogers, 1999), perceive more apparent depth in supra-threshold stimuli (Rogers & Graham, 1983; Buckley & Frisby, 1992; Gillam & Rogers, 1991) and have shorter latencies to perceive depth (Gillam, Finlay & Flagg, 1984; Gillam, Chambers & Russo, 1988; van Ee & Erkelens, 1996, Bradshaw, Hibbard & Gillam, 1999; Hibbard & Bradshaw, 1999). Furthermore, Rogers and Graham (1983) demonstrated that a similar anisotropy also exists for surfaces defined by motion parallax information. However, subsequent studies have concentrated on examining the phenomenon within the disparity domain. Although the results of these studies have allowed a greater insight into the mechanisms that might underlie the stereoscopic anisotropy there is still no complete explanation of the phenomenon. This may, in part, reflect an insufficient understanding of the circumstances under which the phenomenon occurs, e.g. nearly all studies have concentrated upon surfaces sloped around horizontal and vertical axes (see however, Bradshaw et al, 1999). As a
consequence the purpose of the following experiments was to examine empirically a number of aspects of the anisotropy, which have not been addressed by the previous studies. Specifically, the experiments investigated the perception of (i) binocular disparity and (ii) motion parallax defined corrugated surfaces curved around axes orientated at a regular series of oblique angles from horizontal to vertical for both threshold and supra-threshold tasks.

2.1.1 Previous Studies of the Depth Anisotropy

Rogers and Graham (1983) discovered that the depth anisotropy occurred for both disparity and parallax defined surfaces whilst examining the depth equivalent of the Craik-O'Brien-Cornsweet illusion. This illusion, reported by Antis, Howard & Rogers (1978), occurs when two equi-distant planar regions of gradually changing depth are separated by a sharp depth discontinuity (see figure 2.1) and as a consequence observers perceive one of the planar regions to be closer than the other. However, Rogers and Graham found that for surfaces specified by binocular disparities, or motion parallax generated from horizontal movement, the illusion only occurred when the depth discontinuity was orientated vertically (see figure 2.1a). In surfaces with a horizontally orientated depth discontinuity the flanking regions appeared equi-distant. They further found that if the surface was specified by parallax generated from vertical movements the effect was reversed, i.e. the illusion occurred for a horizontally not vertically orientated depth discontinuity. As a consequence they argued that the anisotropy was a result of the geometric properties generated by the transformations between the retinal images rather than the retinal orientation of the images per se. The magnitudes of the disparities, or relative motion, of individual
points within the horizontally and vertically orientated stimuli were the same and so cannot account for the differences in perception of the two surface types. Therefore, one implication of these results was to support the suggestion that the derivation of depth in these surfaces is dependant upon higher order and configurational aspects of the disparity, or motion, information (Wallach & Bacon, 1976; Wallach & Lindauer, 1976; Gillam et al, 1984, 1988; Gillam & Ryan, 1991; Cagenello & Rogers, 1993; van Ee & Erkelens, 1996; Bradshaw et al, 1999).

![Figure 2.1](image)

**Figure 2.1.** Perspective illustrations of the surfaces used by Rogers and Graham (1983). Consider surfaces specified by binocular disparities or motion parallax created by horizontal motion. For surface (a) observers will perceive point (ii) to be closer than point (i) and for surface (b) observers will perceive both planar regions to be equi-distant. This anisotropy reverses for surfaces specified by motion parallax created by vertical motion.

The suggestion that the visual system does not use point disparities alone to derive the depth structure of a scene was assessed empirically by Gillam et al (1988). They examined both (i) the time it took observers to fuse disparity defined horizontally and vertically orientated planar surfaces and (ii) the time required to perceive depth within them. Observers' latencies to fuse the two images (which required them to match all of the individual points) were similar for both surface orientations. However, the latencies to perceive depth were considerably greater for vertically orientated surfaces. Consequently, they argued from their results that the derivation of
stereoscopic depth was dependent on the relative disparities of groups of elements extracted from a point disparity map. They noted that the advantage of relative disparity information was that it remained unchanged by eye movements unlike point disparities that are absolute values. Furthermore, relative disparities are not restricted to pairs of congruent elements but can apply to groups of elements. Consequently, there are many potential sources of information that the visual system might utilise to derive depth and so many studies have attempted to identify the sources of information which are used in practice (Gillam et al, 1984, 1988; Mitchison & McKee, 1990; Gillam & Ryan, 1992; Cagenello & Rogers, 1993; van Ee & Erkelens, 1996; Bradshaw et al, 1999).

Koenderink and van Doorn (1976) demonstrated that one way to characterise differences generated between simultaneous binocular images, and the sequence of images generated by motion parallax, is in terms of three types of transformation: (i) *dilation* which is a uniform expansion or contraction, (ii) *curl* which is a rigid rotation and (iii) the two components of *deformation* which is an expansion in one meridian and a contraction in an orthogonal meridian (see figure 2.2). Koenderink and van Doorn also noted that all the information about surface slant could be derived from the components of deformation. There are a number of other ways that the transformations between the retinal images may be decomposed. For surfaces defined by binocular disparity (or parallax from horizontal motion), the images of horizontally orientated surfaces are primarily related by a *shear transformation* (a deformation and a rotation) and the images of vertically orientated surfaces are primarily related by an *expansion/compression* transformation (a deformation and a dilation).

\footnote{For brevity the term expansion/compression will be shortened to compression throughout the thesis.}
Rogers & Graham (1983) argued that the anisotropy was attributable to the greater sensitivity of the visual system to information generated by shear transformations in low frequency depth changes (see also Bradshaw et al, 1999). However, Cagenello and Rogers (1990, 1993) noted that even if true this explanation was incomplete as it failed to specify which components of the transformation were significant and why the visual system was more sensitive to this information source. They demonstrated that disparity defined shear and compression surfaces generate different magnitudes of orientation disparity and then argued that the visual system's sensitivity to this property was responsible for the anisotropy in threshold detection tasks.

Orientation disparity is the difference in orientation between corresponding features in the two simultaneous retinal images. Koenderink (1986) showed that the deformation between binocular images could, in principal, be calculated from the orientation difference between corresponding line elements at different absolute orientations. The magnitude of orientation disparity between the corresponding
images of a line on a surface is a function of (i) the magnitude of the surface slope and (ii) the absolute orientation of the line. For a planar surface sloped around a horizontal axis (also referred to as an inclined surface) with a fixed magnitude, the orientation disparity generated by an arbitrarily orientated line element is a monotonic function of its orientation with respect to the observer, starting at zero for horizontal lines and rising to a maximum for vertically orientated elements. However for a planar surface sloped around a vertical axis (also referred to as a slanted surface) with a fixed magnitude, the orientation disparity of an arbitrary surface line is a non-monotonic function of its absolute orientation starting at zero for horizontal lines, rising to a maximum for ±45° lines and then declining back to zero for vertically orientated lines. Cagenello and Rogers (1993) noted that for surfaces of equal slope magnitude, the maximum orientation disparity that could be generated by an element on a shear (horizontal) surface was twice as large as that for a compression (vertical) surface. They also pointed out that a texture comprising ±45° line elements generated the same magnitude of orientation disparity on both surface types. Furthermore, when both types of surface were covered with randomly orientated texture elements the average orientation disparity will be 57% greater within the shear surface.

To assess their theory Cagenello and Rogers measured the depth detection thresholds for disparity defined planar horizontally and vertically orientated surfaces whilst varying the available orientation disparity information. This was achieved by using surfaces of three types. Their first stimuli were random dot stereograms (RDSs) that did not contain any explicit line elements. These surfaces did, however, contain a distribution of Fourier energy across all orientations in the luminance domain to which the visual system is sensitive. They found a difference of 65% between the thresholds for horizontal and vertical surfaces. This was closer to that predicted by a
dependence upon the average orientation disparity (57%) rather than the maximum orientation disparity (100%). The second stimuli were stereoscopic surfaces covered with a 0/90° grid of line elements that maximised the orientation disparities present in horizontal surfaces whilst vertical surfaces still contained no explicit orientation disparities. They found that the thresholds were 2-3 times higher for vertical surfaces, which indicated that although orientation disparities were important the visual system could use other information sources to derive the surface inclination. The third stimuli were stereoscopic surfaces ruled with a ±45° grid of line elements that generated the same magnitude of orientation disparity at both surface orientations. In this condition observers' thresholds were the same for both surface types and of a slightly greater magnitude than for horizontal surfaces ruled with a 0/90° grid, which Cagenello and Rogers argued was also consistent with the use of average orientation disparity.

Nonetheless, studies using supra-threshold stimuli have shown that orientation disparity cannot by itself provide a sufficient explanation of the anisotropy (Mitchison & McKee, 1990; Gillam & Ryan, 1992). Gillam and Ryan argued that the perspective information available from the lines on ruled surfaces specified a flat fronto-parallel surface and therefore conflicted with the surface slope specified by disparity information. Gillam (1968) had previously demonstrated that the compression of texture elements produced by surfaces sloped around a horizontal axis was a less effective cue to slope magnitude than the convergence of texture elements produced by surfaces sloped around a vertical axis. As a consequence, they argued conflicts generated by perspective in Cagenello and Rogers experiment could have caused the differences in slope magnitude between surfaces slanted around a vertical axis ruled with ±45° lines and those ruled with ±90° lines. Gillam and Ryan claimed
that the perspective conflict rather than variations in the available orientation disparity information was the primary cause of the anisotropy. Their finding that the anisotropy persisted within stereoscopic supra-threshold horizontally and vertically orientated surfaces ruled with ±45° line elements supported this suggestion. The orientation disparity present within these surfaces was equal but the effectiveness of conflicting perspective information was greater for the vertical surfaces. They did, however, note that neither orientation disparity nor perspective conflicts were sufficient to explain all of their results. In particular, they could not explain why a similar perspective conflict within horizontally orientated surfaces had little effect. As a consequence, they argued that other (unspecified) configural aspects must play a role in the anisotropy.

A possible explanation for the differences between the results of Cagenello and Rogers (1990; 1993) and Gillam and Ryan (1992) was that there were variations in the information exploited by the visual system within the threshold and supra-threshold tasks. Cagenello and Rogers (1993) argued that the perspective information available within surfaces at disparity threshold was too small to be detected by the visual system. As a consequence, they suggest that orientation disparity was important for the perception of depth within disparity threshold tasks but that it interacts with other factors, including perspective, in supra-threshold tasks.

Subsequent studies have produced evidence of further factors that are implicated in the perception of depth within horizontally and vertically orientated surfaces. Hibbard and Langley (1998) examined the perception of slanted and inclined plaid surfaces defined by two sinusoidal gratings (orientated between 5° and 40° to the horizontal). They found that their results were best fit by a model that was sensitive to both orientation and spatial frequency differences in the luminance domain. There
is also evidence that in addition to orientation disparity the visual system might utilise one, or more, of the components of a compression transformation as a stereoscopic primitive (Bradshaw et al, 1999). The effect of the anisotropy tends to be removed by increasing the viewing time for a stimulus, i.e. detection thresholds for horizontal and vertical surfaces become closer to each other as viewing time increases (van Ee & Erkelens, 1996; Hegarty, Bradshaw, DeBruyn & Parton, 1997; Hibbard & Bradshaw, 1999). Therefore, the anisotropy may partially result from the use of stimulus presentation times that are too short for the visual system to fully benefit from the available compression information.

2.1.2 Objectives of the Current Experiments

A major purpose of the current experiments was to examine the effect of varying the magnitude of shear and compression information generated within threshold and supra-threshold surfaces. All of the above studies (except Bradshaw et al, 1999) manipulated the magnitude of orientation disparity within a surface by either adding lines at different orientations (Cagenello & Rogers, 1990, 1993; Mitchison & McKee, 1990; Gillam & Ryan, 1992) or varying the magnitude of the slope of a surface at a particular orientation (Gillam & Ryan, 1992). The former has the disadvantage of introducing explicit perspective information and the latter can only be used within supra-threshold tasks. In the experiments within this chapter shear and compression information were manipulated by changes in the orientation of the axis of curvature of a stimuli covered with random dots. Shear information will be at a maximum for surfaces curved around a horizontal axis and will decline monotonically as the orientation of the axis of curvature is increased to vertical. Conversely, compression
is at a minimum for surfaces around a horizontal axis and at a maximum for surfaces around a vertical axis. Additionally, the average and maximum orientation disparity will co-vary with the magnitude of shear and compression information. The advantage of this method of manipulating the information content of the stimuli is that the random dot surface contains no explicit conflicting perspective information and performance can be compared in threshold and supra-threshold tasks with closely matched stimuli. This allowed an assessment of the degree to which the differences between threshold and supra-threshold disparity tasks can be attributed to conflicts from perspective information. As previous studies have indicated that the visual system makes little use of monocular pictorial information on randomly patterned surfaces (Gillam, 1968; also see section 6.3).

A further aim of the experiments was to compare performance for disparity and parallax defined stimuli. Although the studies outlined earlier have made considerable progress in understanding the causes of the anisotropy within the disparity domain there has been little published data examining the depth anisotropy for motion defined surfaces since Rogers and Graham (1983) first reported its existence. Bradshaw and Rogers (1999) have reported extensive data quantifying the disparity sensitivity function for horizontally and vertically orientated surfaces across a large range of corrugation frequencies. However, the only available data for the parallax sensitivity function examined horizontally orientated surfaces (Rogers & Graham, 1982). No published studies have examined the effect of the anisotropy across a range of corrugation frequencies for parallax defined surfaces. Therefore, the experiments within this chapter examined observers' performance for parallax defined stimuli across a range of corrugation frequencies for threshold detection and supra-threshold depth matching tasks.
Additionally, although Rogers and Graham (1983) argued that the existence of the anisotropy within both disparity and parallax domains was indicative of a close relationship between the processing mechanisms of the two information sources this has not previously been examined in detail. The closely matched stimuli and procedures used within the disparity and parallax tasks enabled direct comparisons to be made between observers' settings on the basis of each of these cues.

2.2 Experiment 2.1 - Binocular Disparity Detection Threshold Task

Previous studies have published data outlining the disparity sensitivity function (DSF) for horizontally and vertically orientated corrugated surfaces across a range of corrugation frequencies (Bradshaw & Rogers, 1999), depth detection thresholds for planar surfaces sloped around a 45° axis (Bradshaw et al, 1999) and thresholds for surfaces defined by sinusoidal gratings at a range of orientations (Hibbard & Langley, 1998). However, there is no published data describing the DSF for disparity defined surfaces orientated around other axes. The current experiment assessed observers' disparity detection thresholds for sinusoidal corrugations curved around a regular series of axes orientated between 0° (horizontal) and 90° (vertical). The transformations between the images of the surfaces vary systematically in the amount of shear and compression information that they generate. Shear information is at a maximum for surfaces curved around a horizontal axis and steadily declines as the orientation of the axis of curvature increases to vertical. Conversely, the magnitude of compression information is inversely proportional to shear and so increases with the surface orientation. As a consequence, if thresholds were critically dependent on shear information (or a property that co-varies with it such as orientation disparity)
they would increase with the orientation of the axis of curvature. However, an improvement of thresholds at intermediate orientations could indicate an interaction in the processing of shear and compression information (Bradshaw et al, 1999).

2.2.1 Method

2.2.1.1 Observers

Six observers who had normal, or corrected to normal, visual acuity and stereo acuity of <20 arc sec participated in the experiment. All observers (except HDT) participated as unpaid volunteers. One observer (HDT) had not previously participated within a psychophysical experiment. Two observers participated only for the individual pixel stimuli, two participated for the Gaussian stimuli and two participated for both sets of stimuli.

2.2.1.2 Stimuli

The stimuli were RDSs that depicted sinusoidal corrugated surfaces in a 10° circular aperture. The 'central' corrugation was marked by two lines 1.5° in length offset by 1° from either side of the centre of the stimulus. A cross was presented in the centre of the screen between each trial to enable the observers to maintain fixation. Two different types of RDS were used in which the dots were defined by either (i) Gaussian blobs with a spatial standard deviation of 1.78 arc min or (ii) individual pixels. The density of the dots in these two stimuli was 37.7 dots/deg² and 50% respectively. A standard grey-level-interpolation algorithm was used to create sub-pixel disparity-shifts between the stereo pairs (e.g. Georgeson, Freeman & Scott-Samuel, 1996). These stereo pairs were pre-computed and stored to prevent delays.
between presentations.

The non-linear relationship between intensity of light reproduced on the screen and the internal coding of the 256 pixel grey levels was corrected using a look-up table. A pattern of maximum contrast black and white squares was displayed in an otherwise darkened room. The contrast of the monitor was adjusted until the luminance of a white square (measured using a photometer) was 23 cd m\(^{-2}\) and that of a black square was 0.15 cd m\(^{-2}\). The luminance of the black squares was then increased in 26 steps (1 to 256 in steps of 5) from the minimum to maximum (whilst the luminance of the white squares was simultaneously reduced) and was measured after each increase. A quadratic expression was fitted to the resulting data to produce the look-up table.

2.2.1.3 Apparatus

The stimuli were generated on an Apple Macintosh 7500 and presented using two 12" monochrome Apple monitors arranged in a Wheatstone style configuration and viewed through two front silvered mirrors set ±45° to the median plane. The viewing distance was 95 cm at which distance a pixel subtended 1.19 by 1.23 arc min.

2.2.1.4 Design and Procedure

The method of constant stimuli was used to determine detection thresholds. The observer made a two alternative forced choice classifying the 'central corrugation' of a stimulus as either 'concave' or 'convex'. The stimuli were viewed in a dimly lit room. Thresholds were measured for four corrugation frequencies (0.1, 0.2, 0.4 and 0.8 c/deg) and five surface orientations (0°, 22.5°, 45°, 67.5° and 90°). Each observer performed four blocks of seventy trials for all corrugation frequency/orientation
combinations. The experiment was run over several days with the order of these blocks randomised. Prior to commencing each block the observer viewed a supra-threshold surface (simulated peak-to-trough depth of between 200 and 600 arc sec) of the same corrugation frequency and orientation as the test stimuli. On each trial, the simulated depth amplitude was selected randomly from seven possible values corresponding to -3, -2, -1, 0, 1, 2 or 3 times the smallest 'step size'. Each surface was presented for two seconds. Probit analysis (Finney, 1947) was used to fit a cumulative Gaussian curve to the percentage correct data set. Thresholds were taken to be the 75% point on the psychometric function.

2.2.2 Results and Discussion

Only data for the Gaussian stimuli was analysed statistically and included in the summary plot (see figure 2.3) for two reasons: (i) Gaussian stimuli were used within all four experiments and (ii) once the two observers who had participated in both conditions were excluded there were insufficient observers to perform inferential statistics. Individual performance data (see figure 2.4) for the random dot conditions is included to confirm the generality of the effects reported and for comparison with earlier studies that have used these types of stimuli. The lowest corrugation frequency (0.1 c/deg) was excluded from the analysis, as half of the observers (ADP, MFB and SJW) could not produce reliable thresholds for surfaces curved around a vertical axis within this condition. The thresholds for each observer were converted into standard z-scores using the mean and standard deviation of their thresholds across all three analysed corrugation frequencies. This removed the effects of the absolute differences in sensitivity to disparity information between the individuals.
The normalised scores were analysed using a two-way ANOVA with the corrugation frequency and the orientation of the axis of curvature (or surface orientation) as factors.

Figure 2.3 plots the mean of the observers' normalised thresholds against corrugation frequency and surface orientation for the Gaussian stimuli. There was a monotonic increase in observers' thresholds as the orientation of the axis of curvature increased from 0° (horizontal) to 90° (vertical), which was reflected in a highly significant main effect of surface orientation ($F_{(4,12)}=18.55$, $p<0.001$). Post-hoc analysis using Page's monotonic trend test confirmed that there was a significant increase ($p<0.01$) in thresholds for all corrugation frequencies (see table 2.1). There was a highly significant interaction between surface orientation and corrugation frequency ($F_{(8,24)}=10.77$, $p<0.001$). Figure 2.3 indicates that this was due to a particularly pronounced increase in thresholds with surface orientation at lower corrugation frequencies ($<0.2$ c/deg). Therefore, observers clearly demonstrated the effects of the stereoscopic anisotropy.

![Figure 2.3](image)

**Figure 2.3** Plots the normalised mean detection thresholds (z-scores) for all 4 observers against surface orientation and corrugation frequency for the Gaussian stimuli. NB. The 0.1 c/deg condition was excluded as only 1 of the 4 observers produced thresholds at 90° in this condition.
<table>
<thead>
<tr>
<th>Corr. Freq.</th>
<th>L</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>211</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>0.4</td>
<td>236</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>0.8</td>
<td>226</td>
<td>&lt;0.01*</td>
</tr>
</tbody>
</table>

Table 2.1. The results of post-hoc Page's monotonic trend tests of surface orientation for three of the tested corrugation frequencies. Significant results are marked with an asterisk.

Figure 2.3 also shows that, for all surface orientations, observers' lowest thresholds occurred in the middle of the tested corrugation frequencies (0.4 c/deg) and increased at lower and higher frequencies. The analysis indicated that the main effect of corrugation frequency was highly significant ($F_{(2,6)} = 40.35$, $p < 0.001$). Post-hoc comparisons with Tukeys Honestly Significant Difference (HSD) test revealed that the 0.2 c/deg corrugation frequency differed significantly from the other tested frequencies ($p < 0.05$) and that difference between the other frequencies (0.4 and 0.8 c/deg) approached significance ($p < 0.1$). Furthermore, the magnitude of the effect of frequency is almost certainly underestimated due to the exclusion of the 0.1 c/deg condition in which all observers exhibited their highest thresholds (see figures 2.4a and 2.4b). The results confirmed that the U-shaped disparity sensitivity function, previously reported for horizontally (Tyler, 1974; Rogers & Graham, 1982) and vertically orientated surfaces (Bradshaw & Rogers, 1999), occurred for oblique surface orientations.
Figure 2.4. Plots the detection thresholds, in arc sec, as a function of the surface orientation for the (a) Gaussian and (b) individual pixel stimuli. The error bars represent ±1SE. The different symbols denote different corrugation frequencies: 0.1 (O), 0.2 (▵), 0.4 (Δ) and 0.8 (▼) c/deg. Lines of best linear fit join the points.
Figures 2.4a and 2.4b depict the depth detection thresholds for each observer for both the Gaussian and the individual pixel stimuli. These figures show there was considerable inter-observer variation in performance, which was consistent with the results of earlier studies (e.g. Rogers & Graham, 1982; Bradshaw & Rogers, 1999). This variation was demonstrated by differences in (i) the absolute magnitude of observers' detection thresholds and (ii) observers' susceptibility to the anisotropy. Although there was a considerable variation in sensitivity to disparity information across individuals (e.g. ADP's thresholds were consistently 2-3 times higher than those of the other observers) the best acuity demonstrated within the study (~2.5 arc for observers MFB and PBH at 0.4 c/deg) was comparable with the best performance of previous studies (see for example Howard, 1919; Rogers & Graham, 1982; Bradshaw & Rogers, 1999). The variation across individuals in the effects of the anisotropy can be seen in two main ways. Firstly, three observers failed to obtain thresholds for vertically orientated surfaces at 0.1 c/deg. Interestingly, their thresholds for surfaces orientated around a 67.5° axis were generally of a similar magnitude to those for observers who did produce thresholds. Secondly, one observer HDT showed only a small increase in her thresholds with orientation even at the lowest corrugation frequency.

In conclusion, the results of this experiment were consistent with the suggestion that observers' thresholds for detecting the slope of a disparity defined surface was critically dependent upon the magnitude of horizontal shear (Rogers & Graham, 1983), or a property that co-varies with it such as the magnitude of orientation disparity (Cagenello & Rogers, 1993), generated by the surface.
2.3 Experiment 2.2 - Binocular Disparity Depth Matching (Supra-threshold)

Task

This experiment assessed the magnitude of depth that observers perceived in a fixed amplitude disparity defined corrugated surface that was curved around axes of different orientations. Previous studies have suggested that there was a difference in the processing of depth for disparity defined surfaces within threshold and supra-threshold tasks (Cagenello & Rogers, 1990; 1993; Mitchison & McKee, 1990; Gillam & Ryan, 1992). The current experiment extends the previously published work in two ways. Firstly, the magnitude of shear information (and orientation disparity) was manipulated independently of the surface amplitude and without the presence of conflicting perspective information generated by oriented line elements. Although, there may have been other conflicting pictorial information present, e.g. variations in the mean texture density across the surface, previous studies have indicated that observers cannot make effective use of such information (see chapter 6 for a brief review). Secondly, the task was performed for two corrugation frequencies (0.2 and 0.8 c/deg). There is no previously published data examining the effect of corrugation frequency upon the anisotropy for supra-threshold stimuli. Although the anisotropy is generally considered to be a low corrugation frequency phenomenon the results of the threshold task indicated that it does also occur at higher corrugation frequencies.

2.3.1 Method

2.3.1.1 Observers

Sixteen observers who had normal, or corrected to normal, visual acuity and stereo
acuity of <20 arc sec participated in the second experiment. These observers included all of those who had participated within the threshold task (experiment 2.1). Six observers participated for the individual pixel stimuli alone, seven participated for the Gaussian stimuli alone and three participated for both sets of stimuli.

2.3.1.2 Stimuli and Apparatus
The stimuli were produced in the way outlined for experiment 2.1 except the central corrugation was no longer marked with two lines. They were displayed on exactly the same configuration of experimental equipment.

2.3.1.3 Design and Procedure
The observer's task was to adjust the apparent peak-to-trough depth of a corrugated probe surface until it appeared to match that of a reference surface. The reference surface had fixed amplitude of 300 arc sec and an orientation selected at random from one of the five angles (0°, 22.5°, 45°, 67.5° and 90°) used within the threshold experiments. The probe surface always curved around a horizontal axis and had a randomly assigned initial amplitude (with a mean of 300 arc sec). The two surfaces were presented in alternation with each displayed for two seconds. The probe surface was proceeded by a beep and the observer then adjusted its amplitude. There was no time limit for a trial and the observer was instructed to perform the task as accurately as possible. The stimuli were viewed in a dimly lit room. Four settings were made for each orientation at two corrugation frequencies (0.2 and 0.8 c/deg). The experiment was run in two blocks (defined by corrugation frequency) with their order counterbalanced across observers.
2.3.2 Results and Discussion

The data for observers viewing the Gaussian and individual pixel stimuli were analysed separately using two-way, ANOVAs (Orientation X Frequency). The three observers (ADP, MFB and DTF) that participated for both stimulus types were exclude from the analysis of the individual pixel stimuli.

Figure 2.5 shows the mean depth settings for all observers at each surface orientation and both corrugation frequencies. Figure 2.5 indicates that there was a non-monotonic relationship between surface orientation and the mean depth that observers perceived. The greatest magnitude of perceived depth occurred for surfaces curved around a 45° axis within which observers perceived considerably more depth than for the horizontal probe surface. Additionally, observers considerably underestimated the depth within a surface orientated around a vertical axis at both corrugation frequencies. There was a highly significant main effect of surface orientation (Gaussian: $F_{(4,36)} = 22.01, p<0.001$; Dot: $F_{(4,20)} = 10.39, p<0.001$). The main effect of corrugation frequency was not significant (Gaussian: $F_{(1,9)} = 1.38, p>0.05$; Dot: $F_{(1,5)} = 0.1, p>0.05$) reflecting the generally similar patterns in the observers' performance at both corrugation frequencies. The magnitude of the increase in perceived depth at 45° and the decline for vertically orientated surfaces was less for the 0.8 c/deg surface. Hence, there was a highly significant interaction between surface orientation and corrugation frequency (Gaussian: $F_{(4,36)} = 10.39, p<0.001$; Dot: $F_{(4,20)} = 4.63, p<0.01$). Taken together these findings clearly demonstrated the effect of the stereoscopic anisotropy within supra-threshold stimuli (Gillam et al, 1988, Mitchison & McKee, 1990; Gillam & Ryan, 1992; Cagenello & Rogers, 1993).
Mean Depth Setting (arc sec)

![Graph](image)

Figure 2.5. Plots the mean perceived surface magnitude for all observers as a function of the surface orientation for the (a) Gaussian and (b) individual pixel stimuli. The error bars represent ±1SE. The symbols denote the two different corrugation frequencies: 0.2 (⊙) and 0.8 (▼) c/deg.

Figures 2.6a and 2.6b show the mean settings for each of the observers who had participated within the threshold experiment. Although there was considerable inter-observer variation in the magnitude of the effects described earlier there was generally a very good qualitative agreement between individuals. All observers showed a peak in perceived depth for 45° surfaces and most showed an underestimate of depth for vertical surfaces.

There were, however, two exceptions to this latter pattern: HDT demonstrated no biases at either corrugation frequency when viewing the individual pixel stimuli, and MFB demonstrated no bias for the 0.8 c/deg Gaussian surface. These results question the validity of labelling visual processing isotropic, or anisotropic, on the basis of observations of two points on a continuum. Although HDT and MFB produced veridical responses for horizontal and vertical surfaces their highest estimates of perceived depth occurred for obliquely orientated surfaces. This indicated that previous studies that concentrated upon two orientations might have missed
anisotropic patterns in performance. The peak in perceived depth occurred for *all* 16 observers at both corrugation frequencies and so was subject to less individual variation than the difference between horizontal and vertical surfaces.

Figures 2.6a and 2.6b indicated there were differences in the magnitude of observers' settings for the two stimuli. Firstly, the magnitude of observers' depth settings for obliquely orientated surfaces was greater for the Gaussian stimuli. Secondly, the relationship between the high and low corrugation frequency conditions was different for the two stimuli. However, the purpose of the two sets of stimuli (as noted earlier) was to ensure the generality of the findings. Some degree of variation in the magnitude of the effects is to be expected between the stimuli due to the differences in their dot density. The important point is that the patterns of the observers’ performance are qualitatively very similar for the two types of stimuli and so the effects are not merely an artefact of a particular stimulus.
Figure 2.6. Plots the perceived surface magnitude, in arc sec, as a function of the surface orientation for the (a) Gaussian and (b) individual pixel stimuli for the observers who participated in experiment 2.1. The error bars represent ±1SE. The different symbols denote different corrugation frequencies: 0.2 (❖), and 0.8 (▼) c/deg. The horizontal bar indicates a veridical response. The curves represent best fitting quadratic expressions.
The results of this experiment were not explicable purely in terms of observers' greater sensitivity to shear information or conflicts from perspective cues (Cagenello & Rogers, 1990, 1993; Gillam & Ryan, 1992). The magnitude of shear information declined monotonically as the surface orientation increased. Consequently, if the magnitude of perceived depth was predominantly influenced by this factor it should have also declined. Neither were the results consistent with Gillam and Ryan's (1992) suggestion that perspective conflicts could explain the anisotropy as the stimuli did not contain line elements to generate such perspective information. Bradshaw et al (1999) have suggested an explanation that could partially account for the current data. They proposed that when both shear and compression information is simultaneously present then one might facilitate the perception of the other. This suggestion was supported empirically by their finding that observers' perceptual latencies to detect depth were lower for planar surfaces sloped around a 45° axis than for similar surfaces around either horizontal or vertical axes.

However, Bradshaw et als' suggestion cannot by itself account for the differences between the patterns of the results within the disparity threshold (experiment 2.1) and supra-threshold tasks, as their suggestion does not explain why the interaction between the two information sources does not also occur at threshold. One possible explanation is that the stimulus presentation times were not sufficiently long for the visual system to make effective use of the compression information within the threshold task. However, the greater magnitude of compression information available within supra-threshold surfaces may have enabled the visual system to exploit this information in a shorter time. This suggestion might predict that if the presentation times for threshold stimuli were increased observers' lowest thresholds would occur for obliquely oriented surfaces due to the facilitation effect from the availability of
both information sources.

In summary, the perceived magnitude of depth within a supra-threshold surface in this experiment varied as a non-monotonic function of the orientation of its axis of curvature for all corrugation frequencies. Observers perceived most depth at oblique orientations peaking at 45°. These results cannot be explained exclusively in terms of orientation disparity information (Cagenello & Rogers, 1993) or conflicts from linear perspective (Gillam & Ryan, 1992) which have previously been advanced to explain the anisotropy. They may be due to mechanisms used in the derivation of stereoscopic depth being sensitive to the simultaneous presence of components of shear and compression transformations.

2.4 Experiment 2.3 - Motion Parallax Detection Threshold Task

This experiment assessed observers’ depth detection thresholds for parallax defined corrugated surfaces curved around a series of axes orientated between 0° (horizontal) and 90° (vertical). Rogers and Graham (1982) have reported data showing the depth sensitivity function for horizontal parallax surfaces that demonstrated it was qualitatively similar to that for disparity, i.e. peak sensitivity tended to occur for intermediate corrugation frequencies (~0.4 c/deg) with a decrease in sensitivity at low (<0.1 c/deg) and high corrugation frequencies (>1.6 c/deg). The major aim of this experiment was to extend these findings to quantify depth sensitivity functions for surfaces across a range of different orientations and to examine how the depth anisotropy interacted with corrugation frequency. Additionally, the close matching of the stimuli and procedure to experiment 2.1 allowed an extensive comparison of the effects of the anisotropy within both disparity and parallax domains. Although the
geometric nature of the information specified by the two cues is very similar (see section 1.3) there might be variations in the information that the visual system utilised from each source. Specifically, these results allowed a comparison of the relative sensitivity of the visual system to information generated by shear and compression transformations within the disparity and motion domains.

2.4.1 Method

Except where noted the procedure is as for Experiment 2.1.

2.4.1.1 Observers

Five observers who had normal, or corrected to normal, visual acuity participated in the experiment. Four observers participated within the first two experiments although two of these observers (SJW and PBH) performed this experiment before the disparity experiments.

2.4.1.2 Stimuli

The stimuli were random dot kinematograms (RDK) which depicted sinusoidal corrugated surfaces in a 20° circular aperture. Each surface was defined by a series of frames between which the relative displacement of individual elements, as the observer's head moved from side-to-side, simulated the movement of point on a real 3-d surface. The individual elements within the stimuli were Gaussian blobs with a standard deviation of 1.78 arc min and a density of 37.7 dots/deg². A standard grey-level-interpolation algorithm was used to create sub-pixel shifts in position between successive frames in the motion sequence. The motion frames were pre-computed
and stored to prevent delays between presentations.

2.4.1.3 Apparatus

The stimuli were generated on an Apple Macintosh G3 and presented using a 21" monochrome Radius monitor. The luminance output of the monitor was linearised and matched to that of the monitors in the disparity experiments using the procedure outlined previously (see section 2.2.1.3). The observer viewed the stimuli monocularly via a headrest that was free to move from side-to-side (± 6.5 cm) parallel to the front of the monitor. When the headrest was in its central position the centre line of the stimuli was aligned with, and 66 cm from, the observers' dominant eye at which distance one pixel subtended 1.77 by 1.79 arc min of visual angle. The horizontal position of the observer's head was monitored using a potentiometer and they moved their heads at a rate of 1 Hz, which was synchronised by a metronome. Tyler and Torres (1972) measured observers' sensitivity to sinusoidal movement and found peak sensitivity lay between 1 and 5 Hz. The lower part of this range was chosen due to the practical difficulty of getting observers to move their heads at a faster rate than 1 Hz. All observers reported a solid impression of depth when viewing supra-threshold stimuli at this rate of movement.

2.4.2 Results and Discussion

The observers' thresholds were converted into normalised z-scores before being analysed in a two-way ANOVA (see experiment 2.1). Figure 2.1 plots the mean of the observers' normalised thresholds against corrugation frequency and surface orientation. From figure 2.1 it can be seen that for the two lowest tested corrugation
frequencies (0.1 c/deg and 0.05 c/deg) thresholds increased monotonically as the orientation of the axis of curvature increased (from horizontal to vertical). However at higher corrugation frequencies, thresholds either remained relatively constant or tended to decrease as the surface orientation increased. There was a significant main effect of surface orientation ($F_{(4,16)}=4.168, p<0.05$). Table 2 summarises the results of post-hoc analyses of the orientation of the axis of curvature using Page's monotonic trend tests. Due to the change in direction of the anisotropy at higher corrugation frequencies the sign of a line of best fit was used to determine the direction of the trend to be tested. There was significant increase in thresholds ($p<0.05$) for the 0.05 c/deg corrugation frequency and 0.1 c/deg corrugation frequency approached significance ($p<0.1$), which confirm the trends apparent in figure 2.1. This result was further reflected in the significant interaction between the surface orientation and the corrugation frequency ($F_{(16,64)}=2.373, p<0.01$). These results clearly demonstrated the effects of the anisotropy for motion parallax defined stimuli (Rogers & Graham, 1983).

![Figure 2.7. Plots the mean detection thresholds for all 5 observers against surface orientation and corrugation frequency.](image-url)
Table 2.2. The results of post-hoc Page's monotonic trend tests for three of the tested corrugation frequencies. Significant results were marked with an asterisk. Negative results denote the test was for a downward trend.

<table>
<thead>
<tr>
<th>Corrugation</th>
<th>L</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>251</td>
<td>&lt;0.05*</td>
</tr>
<tr>
<td>0.1</td>
<td>241</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>0.2</td>
<td>230</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>0.4</td>
<td>227</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>0.8</td>
<td>-236</td>
<td>&gt;0.05</td>
</tr>
</tbody>
</table>

Figure 2.7 shows that observers were most sensitive to the corrugation frequencies in the middle of the tested range (0.2-0.4 c/deg) and increasingly less sensitive to outlying frequencies. As a consequence, there was a highly significant main effect of corrugation frequency ($F_{(4,16)}=22.3, p<0.001$). A post hoc Tukey HSD test revealed a significant difference ($p<0.05$) between all but two of the combinations of corrugation frequencies (0.1 c/deg vs. 0.8 c/deg and 0.2 c/deg vs. 0.4 c/deg). These results confirmed that the U-shaped sensitivity function reported by Rogers and Graham (1983) for horizontally orientated parallax defined surfaces also applied to surfaces curved around axes at all orientations.

Figure 2.8 depicts the depth detection thresholds for each individual observer. There was considerable inter-observer variation in performance. For example at 0.8 c/deg, two observers' thresholds (MFB and SJW) increased monotonically whilst the other two observers' thresholds (ADP and PBH) decreased monotonically. Additionally, none of the observers showed clear trends across all surface orientations and corrugation frequencies. For example, consider the observers' thresholds at 0.05 c/deg, there was considerable individual variation in the exact surface orientation at which observers demonstrated their lowest and highest thresholds. However, it should be noted that at this corrugation frequency all observers' thresholds for vertically oriented surfaces were higher than for horizontally orientated surfaces. Hence, the anisotropy was evident within each individual's data. Additionally, all observers showed U-shaped depth sensitivity functions for all surface orientations. (i.e. Peak sensitivity for all observers lay between 0.2 and 0.4 c/deg).
Figure 2.8. Plots the parallax depth detection thresholds, in arc sec, as a function of the surface orientation. The error bars representing ±1SE. The different symbols represent the different corrugation frequencies: 0.05 (•), 0.1 (O), 0.2 (○), 0.4 (Δ) and 0.8 (▼) c/deg. A line of best linear fit joins individual points.
Three observers (ADP, MFB and PBH) exhibited a greater sensitivity to depth from disparity information in horizontal surfaces whilst the fourth observer (SJW) was slightly more sensitive to motion information. This was very similar to the patterns of inter-observer variability reported by Rogers and Graham (1982). However, for all observers the differences between the two cues disappeared as the surface orientation increased. This was particularly true at low corrugation frequencies. Observers who were unable to produce reliable disparity thresholds at 0.1 c/deg experienced little difficulty in producing parallax thresholds at 0.05 c/deg. This is indicative of a greater ability to utilise compression information within the parallax domain.

These results were consistent with earlier results reported by Nakayama, Silverman, MacLeod and Mulligan (1985). They assessed the sensitivity of observers to shear and compression motion in random dot displays across a range of corrugation frequencies (0.125 c/deg-8 c/deg). They found that at the lowest tested corrugation frequencies (~0.125 c/deg) sensitivity to both sources of information was similar and that at higher corrugation frequencies (>0.2 c/deg) observers were more sensitive to compression information. This latter increase in relative sensitivity to compression information might account for the reverse anisotropy that occurred for two of the observers (PBH and SJW). The relatively flat performance at the mid-range of corrugation frequencies across all surface orientations could reflect the observers' higher sensitivity to compression information, i.e. either source of information allowed them to reach a ceiling in performance. Nakayama et al did not test observers with surface corrugation frequencies as low as those used within this experiment. The results of this experiment would predict a higher sensitivity to shear information in such surfaces. This might also explain the difference between the results of Nakayama et al and those of Rogers and Graham (1983), who found that
observers were more sensitive to stimuli that generated a shear transformation. Nakayama et al had attributed the difference to the different psychophysical criterion used in the two studies (e.g. assessing the perception of motion rather than depth).

In summary, the results of this experiment indicate that for surfaces defined by parallax observers demonstrate an increase in thresholds between horizontally and vertically orientated surfaces but only at low corrugation frequencies. Additionally, the magnitude of the increase in observers' thresholds for surfaces at different orientations was much smaller than for disparity defined surfaces at all corrugation frequencies. The experiment indicated that observers have a greater ability to utilise compression information generated by parallax defined surfaces than for the disparity defined surfaces.

2.5 Experiment 2.4 - Motion Parallax Depth Matching (Supra-threshold) Task

This experiment assessed the magnitude of depth that observers perceived in a parallax defined corrugated surface that curved around a series of axes of different orientations. Therefore, the main purpose of the experiment was to allow a comparison of the effects of the depth anisotropy at different corrugation frequencies between (i) parallax threshold and supra-threshold tasks and (ii) parallax and disparity supra-threshold tasks. This allowed an assessment of the similarities and differences in the effectiveness of shear and compression information across these conditions. Experiments 2.1 and 2.2 had already shown a difference in the pattern of observers' performance in disparity threshold and supra-threshold tasks. This was interpreted as indicating that the visual system was able to utilise the presence of both shear and compression information within supra-threshold tasks but that it was primarily
dependent upon shear information at threshold. Additionally, a comparison between experiments 2.1 and 2.3 demonstrated that differences in derivation of depth on the basis of disparity or parallax occurred within threshold tasks, i.e. the anisotropy was only apparent for parallax defined stimuli at low corrugation frequencies (<0.1 c/deg). This indicated that differences in the relative effectiveness of shear and compression information existed at threshold for the two cues.

The task was carried out for three corrugation frequencies (0.1, 0.2 and 0.8 c/deg) rather than the two used within the disparity supra-threshold experiment. This ensured it was performed at a corrugation frequency (0.1 c/deg) at which observers were showing evidence of anisotropic performance within the parallax threshold task.

2.5.1 Method

The stimuli and apparatus were the same as used for Experiment 2.3 (the motion parallax threshold task) and the procedure is as for Experiment 2.4 (the binocular disparity depth-matching task).

2.5.1.1 Observers

Ten observers who had normal, or corrected to normal, visual acuity participated. These observers included four of those who had participated within the motion parallax threshold task (experiment 2.3).

2.5.2 Results and Discussion

Figure 2.9 shows the mean depth perceived by all observers at each surface
orientation and corrugation frequency. For all three corrugation frequencies perceived depth increased monotonically with surface orientation. Consequently, there was a highly significant main effect of surface orientation ($F_{(4,30)} = 12.251$, $p < 0.0001$). However, due to the similarity in observers' performance across all three corrugation frequencies there was no effect of corrugation frequency ($F_{(2,18)} = 1.286$, $p > 0.05$) nor any interaction between the two factors ($F_{(8,72)} = 0.905$, $p > 0.05$).

Figure 2.9. Plots the mean perceived surface magnitude for all observers as a function of the surface orientation. The error bars represent ±1SE. The three symbols represent the different corrugation frequencies: 0.1 (O), 0.2 (♦) and 0.8 c/deg (▼).

Figure 2.10 shows the results for each of the observers who participated within the parallax threshold experiment. Although there was considerable inter-observer variation in the magnitude of the effects described earlier there was generally a very good agreement in the qualitative pattern of the results. All of the observers showed a monotonic increase for most corrugation frequencies. There were some fluctuations in the intermediate orientations (possibly reflecting an interaction between shear and compression information) but in every instance more depth was perceived in the vertically orientated surface than the horizontal orientated surface. Therefore, the results showed that observers exhibited a reversed anisotropy across all of the tested
One interpretation of these results was that they again reflect a greater ability to utilise compression information within motion parallax defined stimuli. Although, all of the graphs appear to show an increased overestimation of depth it should be remembered that the probe was a horizontal surface. Therefore, all estimates of perceived depth were relative. It is perfectly possible that the depth of horizontally orientated surfaces was underestimated and that the increase in perceived depth for vertically orientated surfaces actually represented a more veridical perception. Similarly, the peak for the mid-corrugation frequencies within the disparity defined stimuli might also have represented more veridical perception of depth. As noted earlier, studies have shown that the perceived depth in disparity defined surfaces appears to build up over time (see for example Gillam et al., 1984, 1988; Hibbard & Bradshaw, 1999). As a consequence, the conditions at which peak depth was perceived within the two supra-threshold tasks may have represented the optimal conditions for deriving depth magnitude of a surface in a fixed time period on the basis of these two information sources.

The differences between the results for the parallax threshold and supra-threshold surfaces may also be explicable in terms of the visual system’s reliance upon compression information. The results at threshold show that observers benefit equally from the presence of shear and compression information (at most corrugation frequencies). However for supra-threshold surfaces, the compression information appears to dominate the determination of depth. This difference may be due to the visual system being unable to fully exploit the lower magnitudes of compression information available in threshold surfaces.
Figure 2.10. Plots the perceived surface magnitude, in arc sec, as a function of the surface orientation for 4 observers who participated in the parallax threshold experiment. The different symbols represent the different corrugation frequencies: 0.1(O), 0.2 (⊛) and 0.8 (▼) c/deg.
In summary, the magnitude of perceived depth within a parallax defined supra-threshold surface tends to increase as the orientation of the axis of curvature increases from horizontal to vertical. This suggests that the major determinant of depth within these surfaces is compression information as it also increased with orientation.

2.6 General Discussion

The major aim of the experiments in this chapter was to examine and compare the effects of the anisotropy for disparity and parallax defined threshold and supra-threshold surfaces across a range of corrugation frequencies. The results revealed a number of variations in observers' performance that appeared dependant on both the source of information defining the stimulus and the nature of the task.

For stereoscopic surfaces, observers' thresholds increased monotonically as the orientation of the surface changed from horizontal (0°) to vertical (90°) for all tested corrugation frequencies. This was consistent with the visual system being reliant upon either information generated by a horizontal shear transformation or orientation disparities (Rogers & Graham, 1982; Cagenello & Rogers, 1993; Bradshaw et al, 1999). However, in the stereoscopic supra-threshold task, for both corrugation frequencies, perceived depth was a non-monotonic function of the surface orientation, which peaked for surfaces curved around a 45° axis and declined considerably for surfaces around a 90° axis. This was consistent with the suggestion that in supra-threshold tasks observers were able to exploit the simultaneous presence of information generated by shear and compression transformations (Bradshaw et al, 1999). If this suggestion were true it might be expected that there would be a greater magnitude of neuronal activity associated with obliquely orientated surfaces than
horizontal or vertically orientated surfaces. Regan, Hong and Regan (2000) have reported data measuring evoked potentials for cyclopean square wave gratings at a variety of orientations. They found that maximum response amplitude was obtained for obliquely orientated gratings and the minimum response occurred for horizontal and vertical gratings, which is consistent with Bradshaw et al's proposal.

In contrast, for parallax defined surfaces, observers' thresholds increased monotonically with the surface orientation at low corrugation frequencies (<0.1 c/deg) but at higher corrugation frequencies thresholds were either not systematically effected by the surface orientation or tended to decrease. Additionally, the magnitude of the anisotropy, as indicated by the difference in the thresholds for horizontally and vertically orientated low corrugation frequency surfaces, was much smaller for parallax than disparity defined surfaces. Taken together these results indicate that observers were less dependent upon information generated by a shear transformation and more capable of exploiting information generated by a compression transformation for parallax defined threshold surfaces. Finally, in the parallax supra-threshold tasks perceived depth increased as a monotonic function of the surface orientation. This indicated that although information from both shear and compression could be utilised by the visual system the latter had a greater effect on the magnitude of perceived depth within this task.

The differences in observers' performance in threshold and supra-threshold tasks cannot be explained by the presence of conflicting perspective information in the supra-threshold stimuli (Gillam & Ryan, 1991; Cageneello & Rogers, 1993). Although a weak inherent perspective cue was arguably present within the stimuli Gillam (1968) demonstrated that the visual system can make little, or no, use of such information (see chapter 6). Therefore, the most parsimonious explanation seems to
be that variations in performance are caused by differences in sensitivity to the available primitive information sources across the two tasks, e.g. shear and compression. Assuming the visual system extracts a variety of different components from the transformations generated between retinal images then it entirely possible that the detection threshold for each component will differ. As a consequence, observers' depth thresholds may be determined by a more constrained range of information than their percept of depth magnitude in supra-threshold surfaces.

The findings of these experiments have implications for the interpretation of previously reported interactions between parallax and disparity information, and are suggestive of the mechanisms that may underlie these interactions. Bradshaw and Rogers (1996) have shown that disparity and motion information can be combined to increase sensitivity to the detection of depth beyond that which is possible on the basis of either cue in isolation. However, in the Bradshaw and Rogers study thresholds were considerably higher for parallax defined surfaces than those for disparity defined surfaces. As a consequence, in the surfaces defined by both cues the depth specified by the motion cue was increased to equate the thresholds for both information sources. The current study indicates that the existence of mechanisms for pooling information from the two cues may be due to the existence of situations in real world viewing where the depth detection thresholds for the two cues are of a similar magnitude, i.e. when processing low frequency differences in depth or surfaces sloped around axis at orientations other than horizontal. There is evidence within both the disparity and motion domains that the visual system filters the depth structure of an image based on the frequency of depth changes and that it has distinct mechanisms for processing this filtered information (Schumer & Ganz, 1979; Hogervorst et al, 1999). Therefore, in scenes with a complex depth structure various
different primitive information sources might be crucial in extracting depth at different frequencies.

Bradshaw and Rogers (1996) also argued that non-linear increases in observers' thresholds for surfaces specified by both cues, compared to those specified by an individual cue, was evidence for the combination of information from these cues during the early stages of visual processing rather than from combining the output of cue specific depth modules. The differential sensitivity of each cue to different components of the retinal images, suggested by the results of the current study, would make such a system an especially efficient method for extracting depth. This is because in such a system the derivation of depth could be based on the combination of the most sensitive components from each information source rather than the output of separate depth modules that are limited by the least sensitive components. An additional possibility is that the visual system might combine information from the two cues to enhance thresholds for specific primitive information sources. These suggestions would both be more consistent with a strong fusion model of early cue combination rather than the separate depth modules suggested by Landy et al's (1995) modified weak fusion module.

A major question for further studies is to identify the specific information sources that are crucial for the determination of depth from parallax and disparity information both alone and in combination at threshold and supra-threshold. Although, there is a considerable amount of evidence supporting the importance of shear information and/or orientation disparities for rapidly determining depth in disparity threshold surfaces (e.g. Rogers & Graham, 1983; Gillam & Rogers, 1991; Cagenello & Rogers, 1993; Hibbard & Langley, 1998; Bradshaw et al, 1999) there has been less examination of the important factors in the motion domain. Harris, Freeman and
Hughes (1992) attempted to assess the relative effects of dilation, curl and deformation on the magnitude of tilt for vertically and horizontally orientated motion specified planar surfaces. They found no differences between vertically and horizontally orientated surfaces and so suggested that deformation was the crucial factor (as it is present in both surface types). However, this conclusion is problematic because the presentation time of their stimuli was unlimited and this has been shown to have a crucial effect on the magnitude of the anisotropy within the disparity domain (van Ee & Erkelens, 1996; Hegarty et al, 1998). Consequently, this would probably also explain why Harris et al found no difference in perceived depth between horizontal and vertical surfaces.

The precise nature of how information from these different cues is combined might be tested by assessing observers' sensitivity to depth for surfaces defined by individual cues and the combination of both cues either (i) after lowering sensitivity to individual components of the retinal images on which the cues may be dependent using techniques of adaptation or (ii) by separately manipulating the magnitude of these individual primitive information sources (Gillam & Rogers, 1991; Bradshaw et al, 1999). This would allow an assessment of the degree to which the individual components of the retinal images critically affected thresholds for surfaces specified by both individual cues and the combination of the two cues. Additionally, it would be possible to assess whether cross-modal interactions occurred for the individual primitive information sources. This should enable an identification of which information sources are critical to the perception of depth for surfaces specified by both cues, which components interact in combined surfaces and components that are not utilised.
This account still leaves open to question why there would be a difference between the sensitivity of the visual system to different components of the retinal image generated within each domain given the geometric similarities in the information that they generate. One possibility is that the succession of views contained in the motion information provided observers with supplementary information, e.g. the acceleration component of motion. Examining performance on the basis of two-frame motion, which is more directly analogous to disparity viewing, might assess this latter factor. If higher order motion information was utilised then performance for two-frame motion should be similar to that for disparity as the instantaneous velocity of elements on the surface is not available in such displays. A further possibility is that the variation sensitivity reflects the differential utility of these information sources within the different domains. For example, dilation information (a component of a compression transformation) is of critical importance within the motion domain as it can signal the rate at which an object is approaching. Therefore, the survival of an organism depends on its ability to quickly process and utilise such information. However, horizontal gradients of disparity could either signal the degree of inclination of an object or be generated by eccentric viewing (Mitchison & McKee, 1990). These judgements are generally likely to be less time critical. As a consequence, the visual system may devote more resources to the extraction of dilation information from the motion flow field than within the disparity domain. This could generate both a greater sensitivity and lower latencies to perceive compression information generated by the motion flow field compared to that generated in the disparity domain.

In summary, previous studies have tended to emphasise similarities in the processing of disparity and parallax information (e.g. Rogers & Graham, 1982, 1983; Graham &
Rogers). However, the results of the current experiments showed that the visual system could be differentially sensitive to the various components of the geometric transformations generated within each domain. This in turn may cause a variation in sensitivity to depth derived from each cue for surfaces that generate different types of geometric transformations. Additionally, the experiments show that for either parallax or disparity defined surfaces different patterns of performance occur within threshold and supra-threshold surfaces. This also probably reflects the differences in sensitivity of the visual system to different components of the transformations between the retinal images.
CHAPTER 3

The Task Dependant Use of Binocular Disparity and Motion Parallax Information with Physical Stimuli.¹

3.1 Introduction

The experiments presented within this chapter address the importance of the nature of the perceptual task in determining the degree of depth constancy that observers demonstrate on the basis of motion parallax and binocular disparity information. They examine extent to which the visual system can, and does, derive a metric (Euclidean) representation of depth structure when making visually guided depth judgements. This was assessed by investigating the variations in observers’ performance for three different depth judgement tasks when viewing the same physical stimuli in a dark (restricted cue) viewing environment.

Glennerster et al (1996) drew attention to the possibility that the nature of the perceptual task might influence the degree of depth constancy that observers demonstrate within experiments. They pointed out the visual system could theoretically complete many tasks without deriving a full specification of the magnitude of depth of different features within a scene (i.e. the metric scene structure). For example, they noted knowledge of the ratio of the depths of the different features (i.e. the relief structure) is sufficient to distinguish most natural surfaces (e.g. faces). These ratios could be derived directly from the magnitude of the

¹ A paper based on the material within this chapter is in press in Vision Research.
binocular disparities, or relative motion, between the features. Therefore, in theory different combinations of information might allow the visual system to form representations that describe the layout of a visual scene to different degrees of geometric specificity. Furthermore, there may be variations in the reliability and accuracy of different information sources and hence distortions in the representations that they are used to specify, e.g. errors in the measurement of vergence may lead to systematic distortions in the perception of distance and relative depth in conditions of isolated binocular viewing (Foley, 1980; see section 1.4). Glennerster et al argued the visual system might complete a task using the most efficient available strategy. The most efficient strategy is the one that utilises the available information within the scene to produce the most accurate representation sufficient to perform the task. The extent to which different strategies are available to the visual system could be indicated by variations by observers' performance across a range of tasks that vary in the geometric representations that are theoretically required to complete them. Glennerster et al compared performance for two different tasks under near identical experimental conditions. In a shape judgement task (similar to experiment 3.2) depth constancy was near 75% whereas in a depth matching task (similar to experiment 3.3) constancy was close to 100%. They argued that the difference between these results was due to the fact that the former task requires absolute distance to be recovered whereas the latter can be completed with knowledge of the relative distance of the two displays only.

For the present experiments three tasks were developed that would each benefit from the presence of binocular disparity or motion parallax information. However, these tasks varied in (i) the degree to which they required further scaling information to
complete them and/or (ii) the potential information that they made available to the observer. The tasks included: (i) a depth nulling task that could be completed by detecting and minimising disparities and/or relative motion, (ii) a depth matching task that could be completed from relative disparities or motion interpreted up to a scaling factor derived from local image measurements, and (iii) a shape judgement task that required either the recovery of scaling parameters (in disparity and motion conditions), the use of the rotational component of the object movement to recover shape (Ullman, 1979) or the constraints imposed by simultaneous presence of both cues (Richards, 1985). The inclusion of the depth nulling task extended the range of tasks beyond those used by Glennerster et al. It allowed an assessment of whether the visual system could use uncalibrated disparities, or relative retinal motion, to complete a task and if that were true provided an estimate of observers’ sensitivity to those two sources of information within the experimental viewing conditions.

Glennerster et al’s conclusions were made on solely the basis of binocular disparity information. So, one aim of these experiments was to examine the extent to which their suggestion also applied to judgements based on motion parallax information either presented individually or in conjunction with disparity information. A comparison of the observers’ performance within a specific viewing condition across the three tasks allows an assessment of the extent to which that cue combination could support a range of different strategies. For example, if observers were biased within the static binocular disparity and monocular motion conditions but accurate in the presence of both cues this would imply some form of interaction between them (all other things being equal). However, by comparing the performance between viewing conditions for a specific task it is possible to determine the extent to which similar
strategies were used to perform the task across the different combinations of depth cues. For example, if performance for the monocular motion parallax condition was highly biased within the shape task but close to veridical for the matching task this would imply the visual system was employing a different strategy for completing each task. The combination of the two cues is of theoretical interest because, as outlined in chapter 1, it has been shown in principal that their simultaneous presence can be exploited to recover veridical information about surface shape without recourse to additional sources of information (Richards, 1985). Previous empirical studies of the combination of these two information sources have found different results with some finding benefits (Johnston et al, 1994; Bradshaw et al, 1998) and others finding no advantage (Tittle et al, 1995; Brenner & van Damme, 1997).

The present experiments employed physical stimuli. Frisby et al (1996) argued that the cause of departures from constancy in many previous studies might be artefacts associated with computer-simulated stimuli (but see Bradshaw et al, 1998). Studies that have used real stimuli have found divergent results with some reporting a misperception of shape similar to that for computer simulated stimuli (Blank, 1953; Baird & Biersdorf, 1967; Loomis et al, 1992; Wagner, 1985) and others finding considerably improved (Frisby et al, 1996), or even near perfect (Durgin et al, 1995), performance. However, in real world situations with well-illuminated environments, it is often only possible to have a relatively poor control over the sources of depth information available within a scene. Therefore, the improvement in observers’ performance found by Durgin et al (1995) and Frisby et al (1996) may be due to the introduction of additional cues (e.g. perspective information) rather than problems with computer simulated stimuli per se. So the current experiments used physical
stimuli that were presented in darkness to remove the extraneous perspective cues while obviating many potential artefacts introduced by the use of computer displays.

3.2 Experiment 3.1 - Nulling Task

The first experiment employed a depth nulling procedure based on the Howard-Dolman stereo acuity test (Howard, 1919). The observers' task was to move a central LED so that it was co-linear with two fixed reference LEDs (see figure 3.1). It is well established that motion parallax is an effective cue to depth for both physical (Graham et al, 1948) and simulated stimuli (Rogers & Graham, 1979, 1982) and that it can be combined with disparity information to lower depth detection thresholds (Bradshaw & Rogers, 1996). However, no studies have explicitly compared the interaction of motion parallax and disparity in the detection depth in physical stimuli.

In addition the task may provide a quantification of the observers' detection limits for disparities and/or parallax (depending on the viewing condition) within the experimental viewing environment. Theoretically, this task only required observers to detect and minimise the binocular disparity, or retinal motion, between the LEDs and did not necessitate the recovery of scaling parameters. Consequently, if observers did use disparity or relative motion without scaling them then performance in the task would reflect the sensitivity limits of the cues. However, it remains possible that in practice the visual system completes this task by recovering scaling parameters and fully specifying the depth structure of the scene. If this is true then observers' performance will also reflect the accuracy and reliability with which the visual system can recover these scaling parameters and so should have a similar pattern of biases to
the shape task (experiment 3.2) that necessitates the recovery of these parameters (in
the static binocular viewing condition at least).

3.2.1 Method

3.2.1.1 Observers
Four observers participated in this experiment that all had normal or corrected-to-
normal visual acuity and stereo acuity of <20 arc sec. They were experienced
psychophysical observers and participated as unpaid volunteers.

3.2.1.2 Apparatus
Three bright yellow LEDs were presented in the dark with any additional sources of
light, and surface reflections, removed (see figure 3.1). When the LEDs were co-
linear, and viewed from above, they formed a flat line, orthogonal to the centre line.
The LEDs were arranged to lie in the same horizontal plane as the observer's eyes.
Two of the LEDs were fixed in position for the duration of each trial 3.8° either side
of, and orthogonal to, the centre line and their midpoint lay either 150 or 300 cm from
the origin (observer). The third LED could be moved at a constant speed backwards
or forwards along the centre line under observer control. Great care was taken
positioning the lights to prevent the observer being able to employ a strategy based on
2-d information to complete the task. The brightness of the LEDs was changed
randomly between each trial and the moving light flashed at a rate of 5 Hz to
eliminate the sense of perceived motion. To avoid signalling the viewing distance via
auditory cues, the observer wore ear-defenders throughout the trials and conversation
between the experimenter and observer was avoided. The experiment was run in total darkness but between each trial the room lights were switched on to enable the experimenter to record the results and set-up the next trial.

A headrest was used which was either fixed, or free to move from side-to-side. When viewing stimuli statically, the headrest was fixed such that the centre line of the stimuli was aligned with either the dominant eye (monocular viewing) or midway between the eyes (binocular viewing). In motion conditions the observers were required to move their head 6.5 cm either side of the centre line (twice the interocular distance). They moved their heads at a rate of 1 Hz, which was synchronised by a metronome.

Task: adjust flashing LED so 3 lights co-linear

![Diagram](image)

**Figure 3.1.** Illustrates the procedure for the nulling task. In the diagram, the task is being performed at 150 cm. The dotted circle represents an adjustable flashing LED which the observer moves along the centre line until it appears fronto-parallel with the reference LEDs (represented by open circles). The closed circles represent the position of the reference LEDs when the task is performed at 300 cm and the dotted lines projecting from the observer denote a visual angle of 3.8°.

3.2.1.3 Design and Procedure

The observer moved the adjustable LED so that it appeared to lie at the same distance as two fixed LEDs, i.e. making them all co-linear (see figure 3.1). The adjustable LED started at a random position in front or behind the baseline (the fronto-parallel plane containing the fixed LEDs). Settings were recorded as deviations from the
baseline. The experiments were performed under the following four viewing conditions; (i) static monocular, (ii) static binocular, (iii) monocular with head motion and (iv) binocular with head motion. Each observer made 10 settings within all viewing conditions at both viewing distances (150 and 300 cm). The experiments were blocked by condition and viewing distance, which were ordered randomly.

3.2.2 Results and Discussion

Figure 3.2 shows the signed mean deviation from zero, in cm, of all four observers’ settings with veridical performance indicated by the horizontal dashed line. To quantify the accuracy and reliability of observers’ settings the mean setting error and SD were computed across viewing distances and are presented (as absolute values) in table 3.1. If observers perform this task by detecting and minimising disparities or relative motion then their reliability scores can be taken as an indicator of their sensitivity for each tested information source, or combination of information sources.

![Figure 3.2](image-url)

**Figure 3.2.** Plots the mean (n=4) depth settings, in cm, as a function of viewing distance with the error bars representing ±1SE. The different symbols represent the individual viewing conditions: static monocular viewing (●), monocular motion condition (○), static binocular viewing (▲) and binocular motion viewing (Δ). The dashed horizontal line indicates perfect (veridical) settings.
Table 3.1. The mean (across viewing distance) group performance indices, in absolute terms, for the four experimental viewing conditions within the nulling task.

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>18.4</td>
<td>11.7</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>2.2</td>
<td>6.3</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>0.8</td>
<td>3.4</td>
</tr>
<tr>
<td>Binocular Motion</td>
<td>-0.6</td>
<td>2.8</td>
</tr>
</tbody>
</table>

The results, in figure 3.2 and table 3.1, for the observers’ settings within the static monocular condition indicate a relatively poor level of performance compared to the other conditions. Table 3.1 shows that even compared to the worst of the 'depth cue' condition (static motion parallax) observers were considerably less accurate (the mean was ~1000% greater) and reliable (the mean SD was ~185% greater). The substantial difference between this condition and the other three conditions suggested that any residual information available was insufficient to perform the task.

Table 3.2. Summary of the post-hoc Tukey HSD test comparing the mean settings within the four viewing conditions. Significant results (p<.05) are marked with an asterisk.

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Monocular Motion</th>
<th>Static Binocular</th>
<th>Binocular Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>2.223</td>
<td>.835</td>
<td>.554</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>.026*</td>
<td>.087*</td>
<td>.004*</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>---</td>
<td>.899</td>
<td>.324</td>
</tr>
</tbody>
</table>

Figure 3.2 also shows performance within the three ‘depth cue’ conditions: monocular motion, static binocular and binocular motion. In all three conditions observers’ settings were more accurate and reliable than those for the static monocular condition. There was a significant difference in performance between the four viewing conditions ($F_{3,9}=9.066; p<0.005$). A post-hoc Tukey HSD test (see table 3.2) showed that this was attributable to significant differences (p<0.05) in settings between the
static monocular condition and all three ‘depth cue’ conditions. However, there were no significant differences between any of the ‘depth cue’ conditions.

Performance within the ‘depth cue’ conditions was close to veridical at both viewing distances, which is reflected in the very similar magnitude of the mean setting error and SD scores (in table 3.1). Additionally, one way t-tests revealed that none of the ‘depth cue’ conditions differed significantly (p>0.05) from zero (a veridical response) and an ANOVA indicated there was no significant difference between the two viewing distances (F_{1,3}=0.597; p>0.05).

The results demonstrate that observers could produce accurate and reliable performance under these viewing conditions on the basis of both disparity and/or motion parallax information. Assuming, that this accurate performance was attributable to the use of unscaled disparities or relative motion information then the SDs provide an estimate of the observers' sensitivity to these information sources. Furthermore, the relative sensitivity of the cues can be estimated by taking the ratios of the SDs and as the performance in both binocular conditions appears to be determined by disparity information these conditions can be combined. Hence, the ratio of the average SD for the two binocular conditions to the monocular motion parallax condition is approximately two. This was consistent with previously reported experiments using CRT displays which found that many observers' parallax depth detection thresholds are approximately double their disparity thresholds (Rogers & Graham, 1982; Bradshaw & Rogers, 1996). In contrast to Bradshaw and Rogers’ results there was no improvement in detection thresholds from the simultaneous presence of both cues. Although, as they pointed out, thresholds for detection of depth from binocular disparity information were typically below that for motion
parallax and would, as a consequence, tend to be reached sooner and so determine performance.

3.3 Experiment 3.2 - Shape Task

This task required observers to set the position of an adjustable LED in front of two fixed LEDs (located either side of, and orthogonal to, the centre line) by an amount equal to their separation (i.e. b=h in figure 3.3). Theoretically, observers could adopt various strategies, dependant on the viewing conditions, to perform this task. For the static binocular condition, it was necessary to recover an estimate of the absolute distance to the point of fixation. The dark viewing environment removed pictorial cues, and positioning the stimuli in the same horizontal plane as the observers' eyes eliminated vertical disparities. As a consequence, under these viewing conditions the viewing distance could only be recovered from the vergence angle of the eyes (Foley, 1980; Cumming et al, 1991; Rogers & Bradshaw, 1995a; Bradshaw et al, 1999). For the monocular motion parallax condition there are two possible ways to derive accurate estimates of object shape; (i) using Ullman's (1979) theorem to recover the 3-d shape up to an isotropic size scaling from the rotation of the stimuli configuration relative to the observer (see section 1.3.2), and (ii) the relative motion of different features within a scene could be scaled by the angular extent of eye rotation and ego-motion. However for the combined viewing condition, in addition to the option of recovering object shape on the basis of disparity or parallax in isolation a further possibility exists. As noted in section 1.5, Richards (1985) has demonstrated that the binocular disparity and SFM cues can theoretically be combined to recover veridical information about surface shape without requiring additional scaling parameters.
3.3.1 Method

All details of the method are as for experiment 3.1 except where noted.

3.3.1.1 Design and Procedure

The observer’s task was to move an adjustable LED in front of two fixed LEDs, on either side of the centre line, such that the interval \((h)\) between it and their centre point was the same as their separation \((b)\) (see figure 3.3). On each trial, the initial location of the LED was set at random with an equal number of trials starting either side of the veridical position. Settings were recorded as the deviation from the veridical position, in cm, with a positive score denoting a position nearer the observer.

**Task:** adjust flashing LED so \(h = b\)

![Figure 3.3](image)

*Figure 3.3.* Illustrates the procedure for the shape task. In the diagram, the task is being performed at 150 cm. The dotted circle represents an adjustable flashing LED which the observer moves along the centre line until it appears to be distance \(h\) from the mid-point of the reference LEDs (represented by open circles), where \(h\) is equal to the separation of the reference LEDs \((b)\). The closed circles represent the position of the reference LEDs when the task is performed at 300 cm and the dotted lines projecting from the observer denote a visual angle of 3.8°.

3.3.2 Results And Discussion

The angular size of the fixed LEDs was held constant (3.8°) across the two viewing distances (150 and 300 cm) by doubling the physical size of their separation between the two distances (from 20 to 40 cm). Consequently, the base-to-apex height that observers were aiming to set (i.e. the veridical height) also doubled. The sensitivity of
observers' settings would be, by Weber's law, proportional to the magnitude of a veridical setting. Consequently, their settings at 300 cm were halved to enable direct comparison with those at 150 cm. Figure 3.4 presents the signed means, in cm, for the four observers' settings with the error bars representing the standard error of the mean. The dashed line indicates veridical performance and so perfect, unbiased, performance would lie on this horizontal line. Any deviation from this line indicates a bias in the perceived depth and a slope that diverges from horizontal marks an interaction with viewing distance. The accuracy and reliability of observers' settings, shown in table 3.3, were again quantified by computing the mean setting error and SD at each viewing distance but are expressed as percentages of the veridical depth setting.

![Figure 3.4. Plots the mean (n=4) depth settings in cm as a function of viewing distance with the error bars representing ±1SE. The different symbols represent the individual viewing conditions: static monocular viewing (●), monocular motion condition (○), static binocular viewing (△) and binocular motion viewing (Δ). The dashed horizontal line indicates ideal (veridical) performance at both viewing distances.](image)

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>122</td>
<td>80</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>64</td>
<td>49</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>37</td>
<td>40</td>
</tr>
<tr>
<td>Binocular Motion</td>
<td>36</td>
<td>42</td>
</tr>
</tbody>
</table>

Table 3.3. Group (n=4) depth judgement performance indices (expressed as a percentage of a veridical setting). The mean is the unsigned average deviation from the veridical point.
The results in figure 3.4 and table 3.3 indicate observers’ settings were extremely inaccurate (mean: 122%) and unreliable (SD: 80%) in the static monocular condition. There was a significant difference in observers’ settings between viewing conditions ($F_{3,9}=8.2; p<0.01$). The results of a post-hoc Tukey HSD (see table 3.4) showed that performance within the static monocular condition differed significantly ($p<0.05$) from that in the binocular (static and motion) conditions and that its difference with the monocular motion condition approached significance ($p<0.1$). Although table 3.4 also revealed there were no significant differences between any of the three ‘depth cue’ conditions the performance indices (see table 3.3) showed that the mean setting error was on average 75% greater within the monocular motion condition. The relatively poor performance within the monocular motion condition was attributable in part to the observers’ lower sensitivity for this cue (indicated by the results for the nulling task). Additionally, Durgin et al (1995) demonstrated that the derivation of motion parallax scaling parameters required more information, and that they are more difficult for observers to obtain, than those for binocular disparities. If observers are less sensitive to relative motion and motion scaling parameters than disparities and their associated scaling parameters then this should effect the relative performance within the monocular motion and binocular viewing conditions. However, the ratio of the SDs between the binocular conditions and the motion parallax was 1.2, which was considerably less than found for the nulling task. It is possible the visual system might not be deriving motion scaling parameters but using Ullman’s (1979) proposal to derive shape up to a size scaling factor with deviations from a veridical percept being caused by noise in the measurement of motion. Hogervorst and Eagle (1998) proposed a Bayesian model for interpreting depth from structure from motion information that predicts systematic distortions in perceived depth that are caused by, but not linearly related to, noise in early motion measurements.

It can be seen from figure 3.4 that the observers' bias remained approximately constant across the two viewing distances in the monocular motion condition, which
implied that the viewing distance was taken into account to some degree. Although figure 3.1 indicated that for both binocular conditions the magnitude of the observers' depth settings were larger at the far (300 cm) viewing distance the difference failed to reach significance ($F_{1,3}=2.74; p>0.05$). There was little, if any, evidence of settings within the binocular conditions being improved by the addition of motion information, i.e. the performance indicators in table 3.3 and figure 3.4 are of a similar magnitude for both binocular conditions.

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Monocular Motion</th>
<th>Static Binocular</th>
<th>Binocular Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>32.839</td>
<td>27.42</td>
<td>27.105</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>.071</td>
<td>.01*</td>
<td>.009*</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>---</td>
<td>.552</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.4. Summary of the post-hoc Tukey HSD test comparing the four viewing conditions. Significant results ($p<.05$) are marked with an asterisk.

There were large systematic biases from a veridical percept within all viewing conditions. One way t-tests revealed a significant deviation ($p<0.05$) from zero (veridical) for all three ‘depth cue’ conditions\(^2\). Figure 3.4 clearly shows observers’ depth settings were generally greater than the target size, which indicated a compression of perceived depth similar to that reported in previous studies (Gogel, 1960; Foley, 1980; Todd & Bressan, 1990; Title et al, 1995; Glennerster et al, 1996; Norman et al, 1996).

In order to quantify the degree to which viewing distance was taken into account in the observers’ depth settings the raw results were converted into ‘effective scaling distances’ (Foley, 1980). The mathematical techniques used were equivalent to those described by Bradshaw et al (1998) and van Damme & Brenner (1997). The

\(^2\) Normally the multiple use of t-tests on a set of data requires the adjustment of the p-value of the tests using the highly conservative Bonferroni correction. This is because the probability of each t-test being significant is additive and so the probability of a single type I error (one test being significant) is increased. However, the probability of all three t-tests being significant is less than .05.
computation was based on the assumption that any error in the observers’ settings is due to a mis-estimate of the viewing distance. Consequently, the effective scaling distance\(^3\) was the distance at which the observers’ setting in units of disparity would correspond to a veridical response (i.e. a depth interval of 20 or 40 cm). From section 1.3.1, the observed disparity (\(\eta\)) is approximated by:

\[
\eta = \frac{I\delta}{D^2}
\]  

(9)

This equation was then rearranged with the veridical response (\(v\)) substituted for the observed depth (\(\delta\)) and the scaling distance (\(D_s\)) for the viewing distance to give the following expression:

\[
0 = \eta D_s^2 - Iv
\]  

(10)

This was then solved as a quadratic equation to give the ‘effective scaling distance’. The transformed data was then plotted as ‘viewing distance’ against ‘effective scaling distance’. This is presented in figure 3.5a with the diagonal hatched line indicating a perfect correspondence between the estimated and actual viewing distance. The ratio of the ‘effective scaling distance’ to the ‘viewing distance’ (expressed as a percentage) is illustrated in figure 3.5b as this directly shows the degree of depth constancy between the two viewing distances. The hatched lines represent a perfect correspondence between estimated and actual viewing distance. Any horizontal line in figure 3.5b indicates complete depth constancy (e.g. depth estimates are the same proportion of the veridical setting at both distances) but only the hatched line denotes a veridical estimate of depth (a proportion of 1).

\(^3\) Although the monocular motion condition does not generate binocular disparities they can be converted into units of equivalent disparity (Rogers & Graham, 1982). The observer travels through twice the interocular distance and so the same angular displacements were created in the spatio-temporal domain as in the spatial domain (binocular conditions). See section 1.3.2.
The graphs plot (a) the 'estimated' scaling distance at which the angular extent of the depth settings would have been veridical as a function of the physical viewing distance, and (b) the ratio of the 'estimated scaling' and actual viewing distance. The different symbols represent the individual viewing conditions: monocular motion condition (O), static binocular viewing (▲) and binocular motion viewing (▼). The dashed line indicates a perfect correspondence between estimated and actual distance.

The results in figures 3.5a and 3.5b clearly show that although the observers' settings depart from a veridical percept they appear to take the different viewing distances into account. For both binocular conditions (static and motion), observers' estimates of scaling distance declined slightly with the physical viewing distance denoted by slopes of less than 1 within figure 3.5b (see Foley, 1980; Johnston, 1991). There was little apparent benefit for observers' performance when disparity information was supplemented by the presence of motion parallax information. In the monocular motion condition, the scaling distance was underestimated by a relatively constant proportion (estimates were ~75% of veridical) across both viewing distances and was considerably less accurate than within the binocular conditions.

Overall, the results of the experiment show neither the near perfect (veridical) performance reported by Durgin et al (1995) and Frisby et al (1996) nor the extremely

---

4 Throughout the thesis when referring to constancy ratios from the graph are converted to percentages by multiplying by 100.
poor levels of depth constancy reported by Johnston (1991). However, many other studies (Gogel, 1960; Foley, 1980; Wagner, 1985; Loomis et al, 1992; Tittle et al, 1995; Norman, et al. 1996; Glennerster et al, 1996) have found qualitatively similar results to those within this study, i.e. a systematic compression of absolute distance. Though the exact degree of compression and constancy within these experiments appears to be a function of the prevailing viewing conditions and available cues. For example, Norman et al (1996) compared observers' assessments of the separation of pairs of LEDs at a variety of viewing distances and orientated at different angles to the fronto-parallel plane. They found performance on the basis of disparity or parallax information was inaccurate (a mean setting error of 38%) and was of similar magnitude to that found within the binocular conditions of the current experiment (a mean setting error of 36.5%) but it was considerably better than within the monocular motion conditions (a mean setting error of 64%).

However, their experiment was performed in a dimly illuminated environment with a visible textured ground-plane providing cues unavailable within this experimental set-up. The extremely accurate performance found by Durgin et al (1995) might also be partially explained by their use of an illuminated structured viewing environment and additionally to the use of textured surfaces as experimental stimuli (these issues are considered further within chapter 6). Frisby et al (1996) compared observers' assessments of the length of a knarled stick orientated at different angles to the fronto-parallel plane under the four similar viewing cue conditions to those used in this experiment (see their experiment 5). The only major difference was that motion parallax was generated by object rather than observer motion. They found a mean setting error (across the three viewing cue conditions) of 14.1% and a mean SD of
5.7%, which are approximately a third and a seventh of their respective values within the current experiment. However, observers’ performance within the static monocular viewing condition within this experiment was surprisingly good (mean: 10% and SD: 5.5%) and actually more accurate and reliable than settings within the three 'depth cue' conditions. Frisby et al did find that observers were least accurate within the static monocular condition in some of their other experiments (see their experiments 4 and 6). However, observers’ settings within static monocular viewing conditions within these experiments were still more accurate and reliable than within the current experiment, or those of Norman et al (1996), for conditions with binocular disparities or motion parallax cues present. This strongly indicates that there were residual cues, or artefacts, within Frisby et als' experimental set-up that enhanced observers’ performance. Therefore, the results of their experiment appear consistent with previous studies that suggested the prevailing viewing conditions have a large effect on the degree of constancy that an observer demonstrates. A suggestion considered further in experiment 6.1 (see chapter 6) that assessed observers’ depth constancy in a fully illuminated environment.

In conclusion, the perception of depth under these viewing conditions can be systematically distorted. Observers could perform the nulling task, under the same viewing conditions, precisely and without bias. Consequently, their difficulty within this task was consistent with an inability to accurately recover appropriate depth scaling parameters. This was attributed to the impoverished viewing conditions, which removed additional information that may have facilitated observers’ performance in previous studies. However it still remains possible within the monocular motion condition, that observers performed the task by judging the
stimulus shape directly using the rotational component of its motion (Ullman, 1979) and not the scaling parameters, a possibility that is considered further in chapter 5.

3.4 Experiment 3.3 - Depth Matching Task

The first two experiments had demonstrated that observers were capable of accurately performing a task involving a judgement of relative depth (experiment 3.1) but that they were not capable of recovering an accurate metric representation of the scene structure under these viewing conditions (experiment 3.2). These experiments were interpreted as indicating that observers could utilise measures of disparity, or relative motion, but not accurately recover scaling parameters to calibrate them. Additionally, they demonstrated that observers could not accurately recover shape up to an isotropic size scaling factor using Ullman’s (1979) proposal. The primary aim of the current experiment was to assess whether observers could make use of additional information source to calibrate disparities or parallax information, when that information is theoretically sufficient to perform the task but not to derive a metric (Euclidean) representation of the scene structure. It is possible that whilst the visual system is capable of utilising disparities or parallax directly (experiment 1) it will always attempt to recover the Euclidean structure for other tasks. The deviations from veridical perception that were found in experiment 2 may merely reflect difficulty recovering the necessary additional information to do this. However, another possibility is that the visual system is able to employ a range of different strategies depending on the requirements experimental task and on the information available (Glennerster et al, 1996). In this experiment the observers’ task was to adjust the
base-to-apex height of a probe triangle to match the base-to-apex height of a fixed reference triangle of LEDs located at a different viewing distance. For a similar depth matching task Glennerster et al (1996) noted disparities, and possibly relative motion information, could be scaled using the ratio of the angular sizes of the stimuli to estimate the ratio of their absolute distances from the observer.

3.4.1 Method

All details of the method are as for experiment 3.1 unless otherwise noted.

3.4.1.1 Apparatus

In addition to the apparatus used in the previous two experiments three fixed reference LEDs, which formed a triangle when viewed from above, were present throughout the experiment. The fixed reference triangle was positioned with its base orthogonal to the centre line at a distance of 212 cm from the observer. The triangle’s base LEDs were separated by 40 cm (which projected 5.4° either side of the centre line), and the base-to-apex height was fixed for each trial at either 20 or 60 cm (22 arc min and 84 arc min in terms binocular, or equivalent, disparity). The reference was positioned 1.8° above the horizontal plane of the remaining LEDs. The probe LEDs were identical to that used in experiments 3.1 and 3.2 except the distance between the base LEDs was fixed at 40 cm. These sizes were chosen such that the angular extent (~5 arc min) of the relative displacement between the fixed and adjustable LEDs, for the smallest setting which observers made (20 cm at the 300 cm viewing distance), was
well above the typical detection thresholds for relative and absolute displacement (<1 arc min) (McKee, Welch, Taylor & Bowne, 1990).

Task: adjust flashing LED to match reference configuration ( \( h = b \) )

Figure 3.6. Illustrates the procedure for the matching task. In the diagram, the task is being performed at 150 cm. The dotted circle represents an adjustable flashing LED which the observer moves along the centre line until it appears to be distance \( h \) from the mid-point base LEDs (represented by open circles), where \( h \) is equal to the height (\( b \)) of the target LEDs (denoted by the striped circles). The closed circles represent the position of the base LEDs when the task is performed at 300 cm and the dotted lines projecting from the observer denote a visual angle of 7.6°.

3.4.1.2 Design and Procedure

The observer was required to adjust the base-to-apex height of a probe triangle, to equal that of a reference triangle. The base of the probe triangle lay either 150 or 300 cm from them and the base of the reference was fixed at a distance of 212 cm (see figure 3.6). The physical separations of the base lights of the reference and probe triangles were the same. The base-to-apex height of the reference triangle was set at random for each trial (to either 20 or 60 cm). The initial position of the apex light on the probe triangle was randomised either side of the veridical position, and its position could then be adjusted by the observer. All other aspects of the procedure were as specified in experiment 3.1.
3.4.2 Results and Discussion

As within the shape task the sensitivity of observers' settings were proportional to the depth of the reference, i.e. either 20 or 60 cm. Consequently, observers' settings for the large (60 cm) triangle were divided by three so that they were directly comparable to those for the small (20 cm) triangle. As expected their was no significant difference between settings for the two triangles ($F_{1,3}=5.579$ p > 0.05) and so settings for both triangles were combined for all further analysis. Table 3.5 shows the mean setting error (accuracy) and SD (reliability) of the observers' settings (averaged across viewing distances and reference size) expressed as percentages of a veridical setting.

![Figure 3.7. Plots the combined mean (n=4) depth settings for the small and large triangles, in cm, as a function of viewing distance with the error bars representing ±1SE. The different symbols represent the individual viewing conditions: static monocular viewing (○), monocular motion condition (O), static binocular viewing (▲) and binocular motion viewing (▲). The dashed horizontal line indicates veridical settings.](image)

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>-25</td>
<td>100</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>-8</td>
<td>34</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>-15</td>
<td>12</td>
</tr>
<tr>
<td>Binocular Motion</td>
<td>-10</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 3.5. The mean (across distance and reference size) performance indices (expressed as percentages of veridical values).
Both table 3.5 and figure 3.7 show that observers made their least accurate and reliable settings within the *static monocular condition*. Both the mean setting error and SD (see table 3.5) were considerably greater (between 66% and 733%) than those found within any of the ‘depth cue’ conditions. However, there were no significant differences in observers’ mean setting errors between any of the three ‘depth cue’ viewing conditions ($F_{3,9} = 0.102, p > 0.05$).

For the ‘depth cue’ conditions, figure 3.7 and table 3.5 indicate that observers' settings demonstrated a very different pattern to those for the shape task (see figure 3.4 and table 3.3). Firstly for all viewing conditions, observers' settings were more accurate and reliable than within the shape task (averaged across 'depth cue' conditions observers had a mean of ~11% compared to ~46% and an average SD of ~22% compared to ~44%). Additionally, unlike the shape task one-way t-tests revealed observers demonstrated no significant bias ($p > 0.05$) within any of the ‘depth cue’ conditions. Secondly, figure 3.7 indicates a slight tendency for observers to underestimate the magnitude of the depth interval in contrast to the overestimates produced with the shape task (see figure 3.4). Finally, performance was apparently most accurate within the motion parallax condition (see table 3.5). However, figure 3.7 shows that parallax settings were highly accurate at the near viewing distance (150 cm) but less accurate than the other depth conditions at the farthest distance (300 cm). This variability in the parallax settings is reflected in the comparatively high SD in comparison with the other 'depth cue' conditions.

This raises the question of how much of the additional variability in the monocular motion condition is attributable to the visual system’s lower sensitivity to relative motion. The ratio of the average SDs of the binocular conditions to the monocular
motion parallax condition in experiment 1 (2.03) is extremely close to that for experiment 3 (2.13). This suggests that difference in variability between the monocular motion and binocular conditions may be largely due to differences in their relative sensitivity. Furthermore, it suggests that the results in the monocular motion condition are being calibrated using a parameter with the same degree of variability as that used in the binocular conditions. The simplest interpretation of this is that the same scaling parameter is used within all conditions, i.e. the ratio of the angular size of the target and probe triangles.

Figure 3.7 indicates there was little difference in settings for the two binocular conditions (static and motion). Therefore, it appears that as for the two earlier tasks (experiments 3.1 and 3.2) observers’ sensitivity to disparity information within these conditions is so much better than for motion parallax that they can get no additional benefit from the presence of both cues.

To further compare performance with the shape task the mean settings were transformed into ‘estimated scaling distances’. This procedure was based on that used by Glennerster et al (1996) but it was modified as in their procedure the probe surface was at a fixed viewing distance whilst the reference surface was presented at different distances (the opposite of the arrangement of the stimuli in this experiment). Glennerster et al assumed the distance to the fixed probe surface was known and so used the observers’ setting of this probe surface as an indication of the perceived size of the reference surface. The scaling distance was taken as the distance at which the disparity of the reference surface would correspond to a physical depth interval of the size of the observers’ setting. For the purposes of this study the calculation was made assuming that the distance to the fixed reference surface was known (rather than the
probe surface). Therefore, the scaling distance was the distance at which the disparity of the probe triangle’s depth, as set by the observer, corresponded to a physical depth interval of the size of the reference surface (which was fixed at 20 or 60 cm). As a consequence the resulting calculation was mathematically equivalent to that used for the shape task. The converted results were again plotted as (i) ‘viewing distance’ against ‘estimated scaling distance’ (figure 3.8a), and (ii) the ratios of the ‘estimated scaling distance’ to the ‘viewing distance’ (figure 3.8b).

![Graphs](image)

**Figure 3.8.** The combined data for both triangles is plotted as (a) the scaling distance at which the angular extent of the depth settings would have been veridical as a function of the physical viewing distance, and (b) the ratio of the scaling distance to the viewing distance as a function of the viewing distance. The different symbols represent the individual viewing conditions: static monocular viewing (●), monocular motion condition (O), static binocular viewing (▲) and binocular motion viewing (Δ). The dashed line indicates perfect correspondence between estimated and actual viewing distance.

The results in figures 3.8a and 3.8b show that observers’ settings exhibit only a small departure from veridical and that they appear to take the different viewing distances into account. Observers tended to overestimate the scaling distance although this tendency declined slightly with absolute distance for both binocular conditions and increased for the monocular motion condition.
The scaling distance conversion allowed the observers' settings to be compared directly with those from the shape task to see if there was evidence of observers employing a different strategy to complete the task. This might have been demonstrated in three ways: (i) a higher precision in the estimation of the 'effective scaling distance', (ii) greater constancy (though given the relatively high levels of constancy within the shape task this was unlikely) and (iii) a different pattern of biases. Figures 3.8a and 3.8b indicate that differences in the patterns of these results do exist between the shape and matching tasks and so indicate that observers employ different strategies to complete the two tasks. For the matching task, estimated scaling distances are generally more accurate and demonstrate a smaller bias although in the opposite direction to that for the shape task (see figures 3.5a and 3.5 b). As within the shape task there was little difference in performance between the two binocular conditions, which was again probably attributable observers' performance in the static binocular condition being so accurate that no improvement could be gained from using the additional motion information.

In conclusion, the results of this study indicate that the nature of the experimental task may have an important effect on the observers' performance in a depth assessment task (Glennerster et al, 1996). Specifically, they are consistent with the visual system switching strategy on the basis of available information and the demands of a particular task (Rogers & Bradshaw, 1995a; Glennerster et al, 1996). The improvement in observers' accuracy in comparison with the shape task suggests that observers use an estimate of the ratio of the viewing distances between the two objects rather than derive the absolute egocentric distance of the probe. Further, these results show that observers can calibrate motion parallax information, as well as
disparity information, using measurements of local parameters from within the visual scene.

3.5 General Discussion

A major aim of the experiments within this chapter was to investigate whether the visual system always derives a full metric (Euclidean) representation of scene structure when performing a visually guided depth judgement task. The sources of information from which observers could potentially derive a metric representation were constant across all three tasks. Hence, if the visual system were reliant upon a single type of representation the pattern of biases should have remained the same across these tasks. However, there were considerable variations between the accuracy and reliability of observers' settings across the different tasks, which was consistent with them being completed on the basis of different representations (Glennerster et al, 1996).

The results demonstrated that the visual system switched strategies and used different combinations of information depending upon both the demands of the task and the range of available information (Glennerster et al, 1996). Firstly, in the nulling task, observers' settings for all tested combinations of depth cues were extremely accurate and reliable in comparison with those for the shape task (see tables 3.1 and 3.3). These results were consistent with observers completing the nulling task by merely detecting and minimising relative motion or disparities. A suggestion that was further supported by the difference in sensitivity between disparity and monocular motion
conditions, which was of a similar magnitude to the difference in detection thresholds reported in previous studies (Rogers & Graham, 1982; Bradshaw & Rogers, 1996).

Secondly, observers’ settings within all 'depth cue' conditions of the matching task were much more accurate and reliable than for the shape task. The matching task could not, unlike the nulling task, be completed by detecting and minimising disparities, or relative motion. Furthermore, the information within the scene that could be used to specify the metric structure was equated for the matching and shape tasks. Therefore, if the same representation were used to complete any of the tasks it would be expected that the same pattern of biases would occur for each of those tasks. As this was clearly not the case the observers’ performance in the matching task was attributable to the use of a different strategy (and internal representation) than that used for either of the other tasks. The results confirm Glennerster et al.’s (1996) suggestion that disparities could be calibrated by a local scene measurement, i.e. using the ratio of the angular sizes of the reference and probe stimuli to estimate the ratio of their absolute distances from the observer. They also extend the Glennerster et al.’s findings by demonstrating that the same strategy can also be implemented to calibrate motion parallax information. The change in pattern of the observers’ performance between the shape and matching tasks was qualitatively very similar in the monocular motion and disparity conditions (see section 3.4.2).

The results of the shape task demonstrated that under these impoverished viewing conditions the visual system cannot recover an accurate metric representation of scene structure on the basis of disparity and/or relative motion information. In all 'depth cue' conditions of the shape task observers demonstrated a significant bias away from a veridical setting, which was consistent with a mis-estimations of scaling parameters
(Foley, 1980; Johnston, 1991). This confirms the results of a number of other studies that have found large deviations from constancy using computer simulated stimuli (e.g. Todd & Bressan, 1990; Johnston, 1991; Tittle et al, 1995; Glennerster et al, 1996; Norman et al, 1996). However, the replication of these results with physical stimuli in the shape task demonstrated that the deviations from constancy cannot be merely attributable to conflicting depth cues inherent within stimuli displayed on CRTs (Frisby et al, 1996). Furthermore, when considered alongside the results of the other tasks they suggest that the degree of constancy demonstrated in a particular experiment is determined by the range and nature of available information sources within the viewing environment and the demands of the task. The results are consistent with the suggestion that the accurate performance reported in previous studies using physical stimuli were due to the presence of pictorial cues within their well-structured viewing environment (Durgin et al, 1995), surface texture cues (Frisby & Buckley, 1992) and artefacts of the experimental viewing situation (Frisby et al, 1996). The extent to which performance improves within viewing environments with a richer array of depth cues is considered further with chapter 6.

A question left open by these experiments is whether the visual system ever computes a metric representation of the scene structure. As noted above, the results of the three experiments within this chapter have demonstrated that the visual system is capable of utilising a range of strategies to complete tasks. Therefore, in a viewing environment within many sources of information available it is difficult to assess precisely which sources are being utilised to complete a task. It is possible that the visual system has many strategies, like that apparently employed with the matching task, which enable it to complete tasks accurately without calculating a metric representation. Alternately,
when a range of cues are available it may be comparatively easy to derive accurate distance estimates and as a consequence a metric representation of the scene structure. These issues are considered further in the discussion of the two experiments in chapter 6.

The experiments within this chapter also examined whether the visual system could make use of two other strategies to avoid deriving metric scene structure: (i) Ullman’s (1979) proposal that rotational component of object motion can be used to derive its depth structure up to an isotropic size scaling and (ii) Richards (1985) suggestion that the simultaneous presence of disparity and motion information could used to derive veridical size and depth by mutually constraining the range of possible interpretations. The large biases found within the shape task for both clearly showed that neither proposal could be utilised by the visual system within the current viewing conditions to derive an accurate representation of stimulus depth. However, it is possible that either, or both, of these strategies are utilised and that the biases might be introduced by noise within the measurement of retinal motion. Hogervorst & Eagle (1998) have proposed a Bayesian model for the interpretation of SFM information which produces similar biases caused by, but not linearly related to, noise in early motion measurements. Nonetheless, if either proposal could produce an effective assessment of depth it might be expected that observers would use them within both the shape and matching tasks. This was clearly not the case the variations in observers' settings between the shape and matching tasks demonstrated that different strategies were used to complete them. Additionally, for every task observers' performance in the combined cue condition was comparable to that for disparity alone and so the visual system did not seem to exploit the two sources of information in the manner proposed
by Richards. The interaction between parallax and disparity was consistent with both a ‘weak fusion’ model of cue combination (see section 1.2) in which the cues are processed separately the combined by weighted averaging or a strong fusion interaction like cue vetoing (Landy et al, 1995). Both of these approaches are dependent upon on an assessment of the relative reliability of the two information sources. This may mean that in situations where the comparative reliability and accuracy of the two information sources is different then the interactions between them may also vary. This possibility is considered further in chapters 4 and 6, which consider observers' performance in a telepresence and well-illuminated multi-cue environments.

In summary, the visual system does not appear, at least under these viewing conditions, to be capable of the derivation of an accurate metric representation of 3-d scene structure. However, it can make use of a range of other representations depending on the requirements of the perceptual tasks and the available information. The existence of this range of strategies means that performance may vary tremendously depending upon both the prevailing viewing conditions and the specific demands of the experimental task.
CHAPTER 4

The Design of Telepresence Systems: An Application of the Task-Dependant Use of Binocular Disparity and Motion Parallax Information.¹

4.1 Introduction

Telepresence systems are observer controlled machines which attempt to convey sufficient sensory information from a remote environment to allow interaction and performance of tasks with the same level of efficiency as if actually present at that location (Browse & McDonald, 1991). They are being used for an increasingly wide range of applications, e.g. inspection in nuclear plants, bomb disposal, work in space, applications undersea and minimal access surgery (McGreevy & Stoker, 1990; Sheridan, 1992; Yamoto, 1992; Drascic & Grodski, 1993; Schebor, 1995; Huber, Taffinder, Jansen, Russell & Darzi, 1998). Furthermore, they potentially provide a useful tool within vision and other sensory research as they allow researchers a greater ability to control and manipulate sensory information from dynamic real world scenes (e.g. Rushton & Wann, 1999). As a consequence there is considerable empirical interest in the optimisation of these systems, especially quantifying the degree to which observers benefit from the transfer of specific sources of sensory information. The experiments within this chapter addressed the extent to which the results presented in chapter 3 applied to remote viewing using a telepresence system, the implications of changes in the viewing environment for models of depth cue

¹ A paper based on the material presented in this chapter has been published in the International Journal of Cognitive Ergonomics.
combination and whether the tasks could form the basis of a performance taxonomy for the assessment and design of these systems. Specifically, they examine whether the telepresence viewing condition has a differential effect upon the tested depth cues and as a consequence the strategies employed by the visual system to complete tasks. For example, if both binocular disparity and retinal motion thresholds were elevated to a similar level it would then be possible for the visual system to benefit from sub-threshold summation.

Telepresence systems require the accurate portrayal of visual information about the depth and spatial layout of objects within a scene. In systems that use traditional television (monoscopic) displays conveying depth and distance information can be problematic because they often carry impoverished or conflicting information. More modern systems have addressed some of these shortcomings by augmenting the available cues to include binocular disparities and motion information linked to the observer’s head movements (Asbery & Pretlove, 1995). However, it is often expensive in terms of human, physical and financial resources to increase the fidelity of the information relayed by these systems (Drascic & Grodski, 1993). Moreover, the potential benefits of this extra expense are sometimes unrealised (Lippert, Post & Beaton, 1982) and occasionally upgrading a system may actually increase problems (Wann, Rushton & Mon-Williams, 1995). For example, in normal viewing conditions vergence and accommodation vary in unison whereas in stereoscopic displays they must be decoupled — operators must accommodate on the 2-d display screen to keep the visual information resolved whilst making vergence movements to objects lying at different depths within the scene. This decoupling causes several unwanted effects including physical problems such as eyestrain and headaches in
addition to being a source of cue-conflict regarding the depths displayed. However, Wann et al argued rather than solving these problems improving the screen resolution might actually worsen them as the improved resolution could act as a more effective stimulus for accommodation. Therefore, to optimise the trade-off between observer performance, equipment limitations and the expense of providing improved fidelity, it is essential to examine empirically the extent to which providing additional depth cues improves (i) task performance and (ii) the impression of perceived depth. A number of factors have to be considered in this context: first, the cost of providing a specific source of information; second, its potential effect (either positive or negative) on the observer; and third, the nature of the information required for the successful completion of a particular task. The results from chapter 3 indicated the latter two factors would vary with the task to be performed and the sources of depth information available. As a consequence, the intended purpose of the system must be taken into account at the initial design stage when it is decided which visual depth cues to convey.

A number of recent studies have investigated the advantages of stereoscopic displays although they often focused on a single task and neglected to analyse the information requirements of the task, or available cues. These studies found divergent results with some experiments demonstrating improved performance for operators using stereoscopic displays (Pepper, Smith & Cole, 1981; Drascic, 1991; Drascic & Grodski, 1993; Huber et al, 1998) whilst others show little difference from using monoscopic displays (Lippert et al, 1982; Massimino & Sheridan, 1994). Therefore, the current experiments examine a range of carefully selected visually guided tasks (based on those used within chapter 3) and to assess the relative benefits afforded by
different sources of information when viewing a telepresence environment using a head mounted display (HMD).

As before the experiments examined observers' ability to use binocular disparity and motion parallax information, alone and in combination, with extraneous cues removed via a dark viewing environment. However, viewing the experimental scene using the HMD differs from direct viewing in two main ways. Firstly, as noted earlier observers accommodate upon the plane of the viewing screen and not that of the experimental stimuli. Frisby et al. (1996) have argued that conflicts from accommodation are partially responsible for departures from constancy observed in earlier studies of binocular depth perception. The use of telepresence equipment allows the same physical stimuli to be viewed whilst introducing accommodation conflicts that were directly analogous to those generated within computer simulated stimuli. This allows an assessment of the degree to which this factor might be effecting perception whilst holding other differences between stimuli constant. Secondly, the resolution of the cameras and HMD potentially limit observers' acuity (see section 4.2.3). The precise effect of both changes in the viewing environment will depend on the way in which depth information is integrated. Therefore, the results of the experiments are of both theoretical and practical significance.

In summary, the purpose of the present experiments was to determine the extent to which the visual system could exploit different sources of information to support a variety of tasks in a telepresence environment. In particular, they examine the degree to which the relative performance in the various viewing conditions differed from that found within the direct viewing experiments (see chapter 3). This information could, in turn, be used to inform the design of future telepresence systems, and to further
understand the mechanisms by which the disparity and motion information are processed alone and in combination.

4.2 The Telepresence Equipment

4.2.1 The Head Mounted Display

The stimuli were viewed using an ‘EyeGen 3 head mounted display’ manufactured by Virtual Research (USA). This has two CRT displays subtending 32° (horizontally) by 24° (vertically) with a resolution of 493 by 250 lines. The screens displayed a full colour image by decomposing the NTSC input signal into separate red, green and blue fields that were displayed at a rate of 180 Hz on the black and white CRTs. As each field was displayed a synchronised filter of the appropriate colour (e.g. a red filter for the red field) was positioned in front of the screen. This results in the percept of a flicker free colour image that is refreshed at an effective rate of 60 Hz. The interpupillary separation of the screens could be adjusted within the range 50-74 mm. The HMD and cable weighed 1 kg.

4.2.2 The Cameras

The images of the experimental scene were relayed to the HMD by two Panasonic WV-KS152 industrial CCD colour video cameras. These had a resolution of 625 (horizontally) by 430 lines (vertically) which is standard PAL video format. These

1 When testing 106 male adult observers Howard (1919) found a range of interpupillary distances of 58.5-72.5 mm with only three observers above 70 mm or below 60 mm. Hence, the range of adjustment available on the HMD covers the normal range of ocular separation found in the adult population.
images were converted to a NTSC signal before being relayed to the HMD. Each camera weighed 16 g, was 17 mm in diameter and 52.5 mm long.

4.2.3 Visual Acuity Using the Telepresence Equipment

The effective visual acuity supported by the cameras and HMD was determined using a resolution (or grating) acuity technique. The test stimuli consisted of two laser-printed gratings covering an area of 19.2 by 27.4 mm on an A4 sheet. Each grating contained a series of 1, or 2, mm wide horizontally orientated black strips separated by a white gap of the same width. Two observers (with corrected to normal vision) viewed each of the test gratings monocularly. For each trial one of the test gratings started at a random position in front of the camera and was then moved backwards or forwards by the experimenter under the direction of the observer. The trial stopped when the observer was satisfied it was at the final position at which they could resolve an individual line. This procedure was repeated six times for both stimuli and each camera. The distances for the 2 mm stimuli were divided by two and all of the results were then averaged together. The mean distance at which observers could still resolve a 1 mm line element was 36.59 cm (SD 1.85 cm). Hence, observers’ acuity within the telepresence system was 9.39 arc min (i.e. the visual angle of a 1 mm element at 36.59 cm). When tested with the stimuli under direct (real world) viewing conditions the same observers' acuity was 1.1 arc min.
4.3 Experiment 4.1 - Telepresence Nulling Task

The experimental task, as for experiment 3.1, was to adjust a central LED so it appeared co-linear with two fixed LEDs positioned either side of, and orthogonal to, the centre line (see figure 3.1). Experiment 3.1 had already demonstrated that within a reduced cue environment observers completed a depth nulling task by detecting and minimising binocular disparities or relative motion. Therefore in the current task the observers’ mean setting error was taken as their operational acuity for each viewing condition when looking at a scene through the telepresence equipment. This should be higher than directly viewing a scene for all of the cue conditions due to the limitations on the observers’ acuity imposed by the telepresence equipment. However, the experiment assessed whether this, and the other constraints imposed by the equipment, had an equal effect on observers’ sensitivity within all of the combinations of viewing cues.

4.3.1 Method

4.3.1.1 Observers

Five observers (three females and two males) participated in the experiment. Four were naïve to the purpose of the experiment and had only limited experience of psychophysical experiments. One observer had participated in the experiments within chapter 3. All had normal or corrected-to-normal, visual acuity, and a stereoacuity of <20 arc sec.
Figure 4.1. A front view of the experimental headrest. By moving their head from side-to-side observers moved the cameras mounted in the base of the headrest.

4.3.1.2 Apparatus

The stimuli consisted of three LEDs in exactly the same configuration as used within experiment 3.1 (see section 3.2.1.2) viewed in a dark environment using the cameras and HMD. The cameras were fixed to the base of a headrest unit (illustrated in figure 4.1) that was mounted on rollers and could move horizontally along a smooth metallic bar. These rollers could be tightened to fix the headrest in place and prevent camera movement in the static viewing conditions. During monocular conditions the camera viewed by the observer's non-dominant eye was disconnected, leaving a black screen, and the working camera was aligned with the centre line. For binocular conditions, the mid-point of the two cameras was aligned with the centre line. In the head-movement conditions the headrest ensured that the motion signal generated from camera movement was directly coupled to the observer’s head movements. The observer's head moved 6.5 cm both sides of the centre line (twice the interocular distance) and the extent of travel was limited by two end-stops. The observer moved at a rate of .5 Hz as the additional encumbrance of the HMD made movement at 1 Hz
(used previously) impractical. This rate of movement had previously been employed in several studies of parallax specified depth (Rogers & Graham, 1979, 1982, 1983; Graham & Rogers, 1982; Rogers & Collett, 1989).

4.3.1.3 Design and Procedure

The experimental procedure and viewing conditions were the same as those outlined for the real world nulling task (see experiment 3.1) except each observer made four settings within all viewing conditions. Great care was taken to ensure that observers could correctly fuse the images on the two CRT screens. Each observer adjusted the interpupillary distance of the screens until they could see a single image with solid depth. They were instructed to inform the experimenter if they saw a diplopic image (i.e. more than three light sources) at any point during the course of the experiment. The observers were also told they could rest, or withdraw, from the experiment at any time. It has previously been reported that many observers suffer from feelings of nausea when using a head mounted binocular viewing equipment, e.g. HMDs and night vision goggles (Mon-Williams, Wann & Rushton 1993; Hadani, 1991). None of the observers reported any ill effects during the first experiment.

4.3.2 Results and Discussion

Figure 4.2 presents the signed means, in cm, for the five observers’ settings with the error bars representing the standard error of the means. The dashed horizontal line indicates perfect, unbiased, performance. Table 4.1 shows the accuracy and reliability of observers’ settings (in cm) which were again quantified by computing
the mean setting error and SD averaged across viewing distances.

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>21.2</td>
<td>14.4</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>8.6</td>
<td>6.7</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>12.5</td>
<td>11.7</td>
</tr>
<tr>
<td>Binocular Motion</td>
<td>4</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Table 4.1. Group (n=5) performance indices, in cm, for the four experimental viewing conditions within the nulling task.

Figure 4.2. Plots the mean (n=5) depth settings, in cm, as a function of viewing distance for the telepresence nulling task with the error bars representing ±1SE. The different symbols represent the individual viewing conditions: static monocular viewing (●), monocular motion condition (O), static binocular viewing (▲) and binocular motion viewing (Δ). The dashed horizontal line indicates veridical settings.

As within the direct viewing nulling task (experiment 3.1), figure 4.2 and table 4.1 indicate that observers' performance was, in general, least accurate and reliable within the static monocular condition. This demonstrated the effectiveness of the experimental set-up when viewed using the HMD. Further, the results in figure 4.2 indicate, as for experiment 3.1, a positive bias within all viewing conditions although one way t-tests revealed that none of the 'depth cue' conditions differed significantly (p>0.05) from zero (a veridical response).

Observers' accuracy denoted by their mean setting errors (see table 4.1) were in all
viewing conditions worse than for direct viewing. This was to be expected due to the limitation on observers' visual acuity imposed by telepresence viewing. However, the accuracy of observers' performance was considerably more effected by the HMD within the static binocular condition (an increase in the deviation from zero by a factor of ~18) than either the monocular (a factor of ~4) or binocular (a factor of ~10) motion conditions. Additionally, the reliability of observers' settings (SDs) within the two viewing environments were of a similar magnitude for the motion conditions but the static binocular condition became considerably worse using the HMD (by a factor of ~3.5). There was a significant difference in the observers' mean settings between viewing conditions ($F_{3,12}=6.19; p<0.01$). Post-Hoc Tukey HSD tests (see table 4.2) revealed a similar pattern of results to those found when performing the task under direct viewing conditions (experiment 3.1). Both motion viewing conditions (monocular and binocular) were significantly different ($p<0.05$) from the static monocular condition. Additionally, the difference between the static binocular and monocular conditions approached significance ($p<0.1$). As within experiment 3.1, none of the depth cue conditions differed significantly.

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Monocular Motion</th>
<th>Static Binocular</th>
<th>Binocular Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>14.26</td>
<td>20.32</td>
<td>7.38</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>*0.009</td>
<td>0.061</td>
<td>*0.001</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>---</td>
<td>0.691</td>
<td>0.604</td>
</tr>
<tr>
<td></td>
<td>---</td>
<td>---</td>
<td>0.137</td>
</tr>
</tbody>
</table>

Table 4.2. Summary of the Post-Hoc Tukey HSD test comparing mean settings between the four viewing conditions. Significant results ($p<0.05$) are marked with an asterisk.

The relatively poor performance within the static binocular condition might be attributable to a number of factors. Firstly, the observers could have failed to fuse the two images reliably. However, as noted earlier, care was taken to ensure that the
equipment was set-up for correctly for each observer before starting and none of the observers reported seeing a diplopic image at any point during the experiment. Furthermore, observers’ performance did receive some benefit from the binocular information (see table 4.1). Secondly, the absence of accommodative blur in the stimuli may have had a differential effect across the three 'depth cue' conditions, i.e. a greater effect upon the static binocular viewing condition. However, even when viewed directly it is unlikely that the depth intervals and viewing distances used within the current experiments would create sufficient accommodative blur to exceed observers’ threshold limits for this cue (Mather, 1996, 1997). Thirdly, observers may have been mildly affected by the problems of fatigue and nausea associated with the use binocular HMDs (Mon-Williams et al, 1993). Although, no observers reported feeling unwell it is possible that they experienced greater physical discomfort within this condition which, though not sufficient to make them withdraw from the experiment, adversely effected their performance. This suggestion is partially supported by the fact that one of the observers (MEW) withdrew from a later experiment (see experiment 4.3) due to feelings of nausea.

One side effect of the elevation in binocular disparity thresholds relative to motion parallax was that observers might now potentially benefit from combining these two sources of information (Bradshaw & Rogers, 1996). In contrast to the direct viewing experiment observers were least biased and most reliable within the combined disparity and parallax condition (see table 4.1). Although, these differences failed to reach statistical significance (see table 4.2).

All observers repeated the task binocularly (i) viewing the stimuli directly and (ii) through the telepresence system but viewing the experimental set-up with the room
lights switched on (which allowed observers access to pictorial depth cues). When viewing the stimuli directly, the observers’ settings were slightly more accurate and sensitive (mean deviation from zero of 0.1 and a SD of 1.9) than those for the observers in experiment 3.1. Observers’ settings with the lights on were more accurate (mean error of 5.7) and reliable (SD of 2) than within telepresence viewing in the dark but were considerably worse than for direct viewing in the dark. So, although the HMD introduced some limitations on to observers’ performance they do not appear to have imposed an absolute constraint within any of the dark viewing ‘depth cue’ conditions.

In summary, the results from this experiment indicated that observers could perform experiments using a HMD within these viewing conditions. However, the patterns of performance on the basis of different cues may vary from those found when the stimuli are viewed directly. Specifically, the current experiment demonstrated that observers experienced greater difficulty recovering disparity information using the telepresence viewing equipment.

4.4 Experiment 4.2 - Telepresence Equilateral Triangle Task

The task in this experiment was to set the position of an adjustable LED in front of two fixed LEDs (located either side of the centre line) such that when viewed from above the triangle formed by the three lights was equilateral, i.e. to make the distance from the central LED to each of the rear LEDs equal to the separation of the rear LEDs (see figure 4.3). To perform this task successfully observers were obliged to estimate the depth or shape of the stimulus by either deriving appropriate scaling
parameters (the viewing distance or angular rotation), recovering an estimate of the 3-
d shape in the motion conditions (Ullman, 1979) or combining information from the
two cues in the binocular motion condition (Richards, 1985). The similarity to the
shape task (experiment 3.3) would confirm that the previous findings were not due to
specific difficulties associated with that particular task (e.g. a particular problem with
intervals orientated vertically in depth with respect to the observer) but a general
difficulty in performing metric depth tasks. In addition the task provided an
assessment of the relative accuracy with which observers could recover disparity and
motion scaling parameters using the HMD and whether this differs from viewing the
stimuli directly. For example, accommodation and vergence vary in unison in normal
viewing but within the telepresence system accommodation is fixed on the screen
(approximately 3 cm from the observer). As a consequence, this may especially affect
the recovery of estimated distance from the vergence angle (which is necessary for the
calibration of the static binocular disparity information). Finally a change in the
observers’ relative sensitivity to binocular disparity and motion parallax information
may affect their performance in the combined-cues condition. If observers’
sensitivity to the two cues is similar there may be advantages in pooling the two
sources of information (Bradshaw & Rogers, 1996). Alternately, if observers are
more sensitive to motion parallax it might solely determine performance as disparity
did within the direct viewing experiments. Although no evidence for this strategy
was found within the direct viewing tasks the decrease in observers’ sensitivity to
binocular disparities relative to motion information, evident within the telepresence
nulling task, may cause the visual system to employ a different strategy to complete
the task.
4.4.1 Method

All details of the method are as for experiment 4.1 except where noted.

4.4.1.1 Design and Procedure

The observer’s task was to move an adjustable LED in front of two fixed LEDs positioned either side of, and orthogonal to, the centre line such that when viewed from above the three LEDs formed an equilateral triangle. Figure 4.3 depicts the task and shows that when completed without error the diagonal intervals ($d_1$ and $d_2$) created between the central LED and each of the fixed LEDs were the equal to their separation ($b$). All other aspects of the experimental procedure were the same as experiment 4.1.

Task: adjust flashing LED so $b = d_1 = d_2$

Figure 4.3. Illustrates the procedure for the equilateral triangle task. In the diagram, the task is being performed at 150 cm. The dotted circle represents an adjustable flashing LED which the observer moves along the centre line until it appears to be distance $d_1$ and $d_2$ from the reference LEDs (represented by open circles), where $d_1$ and $d_2$ are both equal to the separation of the reference LEDs ($b$). The closed circles represent the position of the reference LEDs when the task is performed at 300 cm and the dotted lines projecting from the observer denotes a visual angle of 3.8°.
4.4.2 Results and Discussion

The settings which observers made at 300 cm were converted to the same scale as those at 150 cm (see experiment 3.2). Figure 4.4 presents the signed means, in cm, for the five observers' settings with the error bars representing the standard error of the means. Table 4.3 shows the accuracy (mean setting error) and reliability (SD) of the settings expressed as percentages of their veridical values.

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>299</td>
<td>127</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>204</td>
<td>61</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>209</td>
<td>126</td>
</tr>
<tr>
<td>Binocular Motion</td>
<td>193</td>
<td>84</td>
</tr>
</tbody>
</table>

Table 4.3. Group (n=5) depth setting performance indices (expressed as a percentage of the veridical setting).

Figure 4.4. Plots the normalised mean (n=5) depth settings in cm as a function of viewing distance for the equilateral triangle task with the error bars representing ±1SE. The different symbols represent the individual viewing conditions: static monocular viewing (○), monocular motion condition (○), static binocular viewing (▲) and binocular motion viewing (▲). The dashed horizontal line indicates veridical performance at both viewing distances.

Figure 4.4 and table 4.3 indicates that observers were least accurate and reliable within the static monocular condition. As within the shape task (experiment 3.3) there was a positive bias within all viewing conditions (see figure 4.4) although the
magnitude of the bias was greater within the current task. One way t-tests revealed that all 'depth cue' conditions differed significantly ($p<0.05$) from zero (a veridical response). There was also a significant difference in observers’ settings between the viewing conditions ($F_{3,12}=4.18; p<0.05$). A post-hoc Tukey HSD (see table 4.4) revealed only the static monocular and binocular motion conditions differed significantly. However, the differences between the static monocular condition and the two other 'depth cue' conditions (static binocular and monocular motion) approached significance ($p>0.1$). This provides a further indication that motion and disparity information may be interacting within the telepresence viewing conditions.

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Monocular Motion</th>
<th>Static Binocular</th>
<th>Binocular Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>51.462</td>
<td>52.693</td>
<td>50.819</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>.055</td>
<td>.078</td>
<td>.046*</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>---</td>
<td>.997</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 4.3. Summary of the post-hoc Tukey HSD test comparing the four viewing conditions. Significant results ($p<0.05$) are marked with an asterisk.

There was no significant difference between observers' settings at the two viewing distances ($F_{1,6}=0.041; p>0.05$). However, to quantify the degree to which the viewing distance was taken into account the results were converted into scaling distances (as outlined in section 3.3.2). The converted data is plotted as (i) ‘effective scaling distance’ against ‘viewing distance’ (figure 4.5a), and (ii) the ratios of the ‘effective scaling distance’ to the ‘viewing distance’ (figure 4.5b). Although there were greater departures from veridical distance assessments (represented by the hatched diagonal line) when using the HMD observers still appeared to take different viewing distances into account. For all three conditions, observers' estimates of scaling distance were of a similar magnitude and an almost constant proportion of the viewing distance (~52%). This would suggest that if scaling parameters were being derived then
viewing using the HMD increases the degree to which they were underestimated. This is almost certainly due, at least in part, to the reduction in the acuity using the HMD. Furthermore, it is also possible that the large conflict from absolute distance signalled by accommodation (the screens were only a few cm from the observer’s eyes) reduced depth estimates via some form of cue averaging. However, Landy et al (1995) have questioned the degree to which information is likely to be combined from hugely discrepant cues.

![Figure 4.5](image)

Figure 4.5. The two graphs plot (a) the ‘estimated’ scaling distance at which the angular extent of the depth settings would have been veridical as a function of the physical viewing distance, and (b) the ratio of the ‘estimated’ scaling distance to the viewing distance, as a function of the viewing distance. The different symbols represent the individual viewing conditions: monocular motion condition (O), static binocular viewing (▲) and binocular motion viewing (△). The dashed line indicates perfect correspondence between estimated and actual viewing distance.

One possible explanation for the increase in observers' bias compared to the shape task may be that observers experienced a greater difficulty in judging the shape of an equilateral triangle. This is unlikely as evidence from Norman et al (1996) indicated that observers’ accuracy actually decreased as the interval they were judging became increasingly orientated in depth with respect to them. Thus, observers should be worse within the shape task where the depth interval they were judging was
perpendicular to the axis of their eyes rather than at 60° as in the current task. However, to confirm that their increased difficulties were not an artefact of the particular task observers repeated the static binocular condition of the task viewing the stimuli directly. The mean setting error and SD, expressed as percentages of the veridical setting, were 34% and 55% respectively. The results indicated a very similar level of accuracy (mean setting error) to the shape task (see table 3.3) but a reduction in observers' sensitivity (i.e. an increased SD). The later is probably merely due to the smaller number of settings performed by each observer within the current experiment. As within the nulling task, observers repeated the static binocular condition with the lights on whilst viewing using the HMD. Observers' performance (a mean error of 48% and SD of 24%) was considerably less accurate than for the direct viewing condition but better than any of the three 'depth cue' conditions. So, it is possible to conclude that the poor levels of performance in the 'depth cue' conditions do not merely reflect a floor in performance imposed by the resolution of the equipment.

In summary, the systematic distortion in the perception of depth experienced by observers when viewing a scene through the telepresence system was similar to that found for directly viewing a scene. As observers could perform a depth nulling task using this equipment the results from the equilateral triangle task provided further evidence of their difficulty in accurately recovering a metric representation of scene structure.

4.5 Experiment 4.3 - Telepresence Depth Matching Task

The observer's task was to adjust the base-to-apex magnitude of a probe triangle to
match that of a fixed reference triangle located at a different viewing distance. In the previous experiment observers had shown extremely poor levels of accuracy and reliability but that task required an estimate of viewing distance, angular rotation or object shape. However, the results of the direct viewing matching task (experiment 3.3) suggest observers might be able, under certain circumstances, to calibrate these cues on the basis of a local measurement (Glennerster et al, 1996). The present experiment assesses whether observers could still benefit from employing this strategy despite the limitations imposed from viewing the scene through the telepresence equipment.

4.5.1 Method

All details of the telepresence equipment are as for experiment 4.1 and all details of the stimuli and procedure are as for experiment 3.3 unless otherwise noted.

4.5.1.1 Observers

Four observers who had all participated in experiments 4.1 and 4.2 completed the experiment. The fifth observer (MEW) withdrew from the experiment after experiencing adverse side effects from viewing stimuli through the HMD.

4.5.1.2 Design and Procedure

The observer’s task was to match the base-to-apex magnitude of an adjustable probe triangle to that of a fixed reference triangle. The procedure was almost identical to that used within experiment 3.3 (the matching task). However, within this experiment
the reference triangle was positioned at 150 cm when the probe triangle was at 300 cm and vice versa. Each observer made five settings within each viewing condition for both triangles and viewing distances.

4.5.2 Results and Discussion

The settings which observers made for the large (60 cm) triangle were converted to the same scale as those for the small (20 cm) triangle (see experiment 3.3). The data for the two triangles was combined as there was no theoretical reason to expect them to vary and no observed statistical difference between them ($F_{1,3}=7.92, p>0.05$).

Figure 4.6 presents the signed means, in cm, for the four observers’ settings. Table 4.4 shows the accuracy (mean setting error) and reliability (SD) of the settings expressed as percentages of the veridical value.

![Figure 4.6](image)

*Figure 4.6.* Plots the combined mean (n=4) depth settings for the small and large triangles, in cm, as a function of viewing distance with the error bars representing ±1SE. The different symbols represent the individual viewing conditions: static monocular viewing (●), monocular motion condition (O), static binocular viewing (▲) and binocular motion viewing (△). The dashed horizontal line indicates a perfect (veridical) setting.
Table 4.5. Group (n=4) performance indices (expressed as percentages of veridical values) with the matched depth settings combined for the 20 and 60 cm reference triangle.

Table 4.5 shows that overall the least accurate and reliable performance occurred within the static monocular condition. Furthermore, there was a highly significant difference in observers' mean settings between viewing conditions ($F_{3,9}=18.016$, $p<0.001$). Post-hoc Tukey HSD tests (see table 4.6) indicated that this was attributable to significant differences between the monocular static condition and all three 'depth cue' conditions ($p<0.01$). There were, however, no significant differences between any of the 'depth cue' conditions.

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>115</td>
<td>77</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>16</td>
<td>43</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>Binocular Motion</td>
<td>11</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 4.6. Summary of the Post-Hoc Tukey HSD test comparing the four viewing conditions. Significant results ($p<.05$) are marked with an asterisk.

The patterns of results in the three 'depth cue' conditions were different from both the telepresence equilateral triangle task (experiment 4.2) and the direct viewing matching task (experiment 3.3). There was a significant difference in observers' performance across the two viewing distances ($F_{1,3}=14.83$, $p<0.05$). This is reflected in figure 4.6 which indicates that at a viewing distance of 150 cm observers’ settings within the 'depth cue' conditions show a slight negative bias and at 300 cm they display a larger positive bias. However unlike the equilateral triangle task, one-way t-tests revealed
that settings at neither distance differed significantly (p>0.05) from zero. Table 5 shows that the observers' settings across all 'depth cue' conditions, as indicated by the mean setting error and SD, were considerably more accurate (a factor of ~4-6 times) and reliable (a factor of ~1.5-4.5 times) than for the equilateral triangle task (experiment 4.2). This, again, indicates that observers performed a depth matching task using a different strategy to a metric shape task.

Performance within the motion conditions (monocular and binocular) was only slightly worse than that for direct viewing matching task. The static binocular condition was, as for the nulling and equilateral triangle tasks, the least accurate and reliable of the 'depth cue' conditions and considerably worse than for direct viewing. The most accurate performance was within the combined-cue condition, which further indicated there was an interaction between the two information sources in this viewing environment.

**Figure 4.7.** The combined data for both triangles is plotted as (a) the 'estimated' scaling distance at which the angular extent of the depth settings would have been veridical as a function of the physical viewing distance, and (b) the ratio of the 'notional' scaling distance to the viewing distance as a function of the viewing distance. The different symbols represent the individual viewing conditions: monocular motion condition (O), static binocular viewing (A) and binocular motion viewing (Δ). The dashed lines indicate a perfect correspondence between estimated and actual viewing distance.
To quantify the degree to which the viewing distance was taken into account within the observers' settings their results were converted into 'notional scaling distances' (as outlined in section 3.3.3). The converted results are plotted as (i) 'notional scaling distance' against 'viewing distance' (figure 4.7a), and (ii) the ratios of the 'notional scaling distance' to the 'viewing distance' (figures 4.7b). This provides a convenient method for comparing the data with that for the equilateral triangle task, and quantifying the degree of depth constancy, even if observers do not actually use the scaling distance to make their setting.

Figures 4.7a and 4.7b indicate that the magnitude of observers' estimates of the 'scaling distances' were far closer to veridical than for the equilateral triangle task. However, figure 4.7b clearly shows that observers exhibited considerable departures from constancy within all 'depth cue' conditions. The telepresence nulling task indicated that there was no increase in the observers' bias for detecting binocular disparities or retinal motion with viewing distance. Furthermore, the equilateral triangle task showed that observers could demonstrate depth constancy when scaling disparity or parallax information on the basis of their respective scaling parameters.

As a consequence, the failures of constancy within this experiment are likely to be due to biases in the local measurement used to calibrate the disparity or parallax information.

In summary, the results of this task indicate that observers could perform a depth matching task with more precision and reliability than the equilateral triangle task. This provided further evidence for the importance of the experimental task in determining observers' performance (Glennerster et al, 1996) on the basis of both motion parallax and binocular disparity information.
4.6 General Discussion

The major aim of these experiments was to both confirm and extend the findings reported in chapter 3. Firstly, the results confirm that the visual system cannot derive a metric representation of the structure of a scene on the basis of binocular disparity and/or motion parallax information presented in isolation from other depth cues. The use of a different metric shape task showed that observers' difficulty was not due to a specific problem with the original shape task (experiment 3.2) but probably reflects a general limitation of the visual system. Furthermore, the displays used within these experiments contained conflicting accommodation information analogous to that present within studies that used computer-generated stimuli. The results show that observers have a qualitatively similar tendency to underestimate the magnitude of depth intervals whether or not this accommodation conflict is present. As a consequence, they confirm that this factor cannot be the sole explanation for departures from constancy in experiments using computer-simulated stimuli. Secondly, the experiments again demonstrate that observers are capable of making use of a variety of representations dependant upon the range of available information and the demands of the task. The results replicate the findings in chapter 3 by showing that observers can perform tasks accurately if they do not require a metric representation. Observers' settings were, as before, consistent with performing tasks on the basis of detecting retinal disparities/relative motion (experiment 4.1) or by calibrating the data on the basis of a local scene measurement (experiment 4.3). However, the results extend the previous findings by demonstrating that the strategies and representations used by the visual system are probably dependent on an
assessment of the reliability of information sources and not merely their presence. This was shown by a change in the relative patterns of performance within the three 'depth cue' conditions (see chapter 3). In all three telepresence experiments performance was most accurate (closest to veridical) within the combined-cue condition, which stands in contrast to the direct viewing experiments where performance was almost identical within the static binocular and combined-cue conditions. This difference was most notable in the nulling task in which observers' bias and sensitivity were approximately 50% of the closest other 'depth-cue' condition, monocular motion parallax (see table 4.1). These results showed that (unlike within direct viewing tasks) parallax and disparity information interacted to determine observers' performance when presented simultaneously within the telepresence environment. This change in results between direct and telepresence viewing is most probably attributable to the decrease in the relative reliability of binocular disparity as an information source. For each experiment, performance was worse in all viewing conditions than for direct viewing, which was to be expected because of the limitations upon acuity imposed by the use of the HMD. However, there was a considerable decline in performance in static binocular conditions relative to the other 'depth cue' conditions. In particular, the telepresence nulling task clearly showed that it was the least reliable 'depth cue' condition (see SDs in table 4.1).

The implication of these results is that when combining information from different depth cues the visual system assesses the reliability of each information source and uses this to determine if, and how, the information is to be combined. This combination might be from modifying the weighting of the information from different sources in a constant combinatorial strategy (e.g. weighted averaging), changing to a
different strategy (e.g. removing one or more information source using a strong fusion strategy like cue vetoing) or, when detecting the presence of depth, probability summation.

An examination of the telepresence nulling task can generate some degree of insight into the processes underlying the integration of information from the two different sources. The task involves detecting and minimising the depth signalled by motion and/or disparity. Bradshaw and Rogers (1996) noted, in their examination of the interaction of parallax and disparity thresholds, that if information from the two sources is processed independently when both cues are presented together the threshold will be determined by either the most sensitive individual cue or by combining information output from the two processes using probability summation. They further showed that probability summation could not increase performance above that predicted by linear summation, which predicts a magnitude of sub-threshold summation of 1.41. Assuming that the sensitivity limits (SDs) in this task represent the thresholds then the magnitude of sub-threshold summation is 2.5, which is considerably greater than expected on the basis of linear summation. Consequently, this data clearly supports Bradshaw and Rogers (1996) suggestion that these cues are integrated in a non-linear manner due to an interaction in the early mechanisms for extracting these two information sources. This would be consistent with a strong fusion model of cue interaction rather than the sort of modularization predicted by Landy et als’ (1995) modified weak fusion model.

The consistent departures from a veridical percept within the motion and combined-cue viewing conditions further showed that the visual system could not employ either the rotation of the stimuli configuration relative to the observer (Ullman, 1979) nor
the constraints of the two information sources (Richards, 1985) to accurately recover the stimulus shape. However, it is not surprising as both cues were more biased and variable than in previous experiments where observers had also failed to utilise these strategies effectively (see chapter 3). The ability of the visual system to switch strategies does, however, mean that in situations where more reliable measurements of the appropriate disparity and motion information can be made, the visual system might be able to implement either of these strategies (see chapter 5). This could explain the mixed evidence of previous studies which have reported benefits from the simultaneous presence of both cues (Johnston et al., 1994; Bradshaw et al., 1998) and no effect (Tittle et al., 1995; Brenner & van Damme, 1997).

In addition to the theoretical issues discussed these results were important because they provide designers of telepresence systems with guidance as to the information required by system operators to perform specific tasks in a particular situation. Drascic (1991) has argued that tasks will vary in the degree to which they benefit from binocular information but previous studies (Drascic, 1991; Massimino & Sheridan, 1994) have focused on the role of single task difficulty, manipulated by increasing the required accuracy, rather than the nature of the task. The results of these tasks clearly show that the utilisation of different information sources is dependent upon the nature of the experimental task. Therefore to assess the relative importance of different information sources when designing these systems it is necessary to examine tasks that require the same kinds of representations as those for which the system is to be used. Although apparently abstract, the tasks used within these experiments are analogous to many tasks that are performed regularly in everyday life. For example, placing the end of a screwdriver against the slot in a screw is a depth nulling task, and
trying to cut even sized pieces of a pie is a depth matching task. However, it is less common to do an absolute shape judgement - such as judging the roundness of a cup. Additionally, a comparison of these experiments and those in chapter 3 clearly shows that the viewing environment may have a differential effect upon different cues and as a consequence the way in which the visual system combines them. Therefore, the results of studies in the real world cannot necessarily be assumed to show how information sources are used within telepresence environments. So, it is necessary when assessing the practical utility of a range of information sources for the experimental viewing conditions to closely model those of the intended application.

In summary, the evidence seems to indicate that the observer can make use of motion and disparity information within the impoverished cue environment viewed through a HMD. However, the relative benefits of different cue combinations may differ from direct viewing conditions and as a consequence the manner in which they are combined. It is important to note, however, that although the absolute levels of performance depend on the resolution of the system, the relative performance depended on the task and the viewing conditions. Therefore, it is clearly important to assess performance across a range of tasks within telepresence systems in order to determine the information that should be displayed.
CHAPTER 5

Perceived Size and Depth on the Basis of Binocular Disparity and Motion Parallax Information.¹

5.1 Introduction

The experiments in chapters 3 and 4 examined observers' judgements of depth magnitude on the basis binocular disparity and/or motion parallax information. They assessed the degree to which on the basis of these cues observers could perform a variety of tasks that theoretically varied in the information required to successfully complete them. This was used to assess the degree to which the nature of the experimental task affected the observers' performance, the strategies that the visual system used to complete the different tasks and the nature of the representations that the cues (alone or in combination) facilitated. The major purpose of the experiment was to extend the range of tested tasks to include one that requires a full Euclidean representation within all of the experimental viewing conditions.

None of the tasks used within the previous chapters are a complete test of whether visual perception is Euclidean in every tested viewing cue condition. Consider the static binocular condition, a particular pattern of retinal disparities was consistent with a family of shapes (see section 1.3.1). For observers to perform both the shape and equilateral triangle tasks veridically, they were required to scale binocular disparities by an accurate estimate of absolute distance which, within the experimental viewing conditions, is not possible. This highlights the limitations of the tasks used and the need for a wider range of tested viewing conditions.

¹ A paper based on the material in this chapter has been published in Perception.
environment, could only be derived from the angle of convergence of the eyes (Foley, 1980; Johnston, 1991). Consequently, if the observer could set the depth of the shapes correctly then binocular vision would be shown to support the accurate recovery of both absolute and relative distance estimates at different orientations in depth (considering both tasks). Therefore, on the basis of those tasks it is possible to determine whether, under the experimental viewing condition, static binocular viewing allows the recovery of Euclidean properties of the scene. This is, however, not necessarily the case within the motion parallax conditions. As outlined in section 1.3.2, Ullman (1979) demonstrated that for SFM, based on three frames from the motion sequence, it is possible for an observer to derive the shape of an object up to a size scaling factor. Theoretically both the shape and equilateral triangle tasks could be completed on the basis of this information irrespective of whether the observer perceived the correct size or viewing distance. That is, although the correct shape (e.g. an equilateral triangle) may be perceived, the observer may be unaware whether it is a small triangle close to them, or a much larger triangle further away. This criticism would also apply to previous studies (Johnston et al, 1994; Tittle et al, 1995) that have used an apparently circular cylinder (ACC) task to assess Richards (1985) suggestion for the combination of disparity and motion information (see section 1.5.2 for more details). Consequently, combined motion and disparity conditions within those studies (and the shape and equilateral triangle tasks within this thesis) could have been completed purely on the basis of the motion information.

In the current experiment the observers were required to make explicit size judgements in the fronto-parallel plane as well as depth judgements along the centre line. The observers' task was to set the depth and size of a visual probe at a series of
different viewing distances to match that of a hand-held reference shape, which they felt but never viewed. This allowed an assessment of perceived size and depth to be made without depending on a purely memory based representation or providing additional visual information that might aid observers' performance of the task.

The following requirements have to be met for observers to perform this task: (i) they need to derive a reliable representation of the size and depth of the reference that is systematically related to its physical dimensions, (ii) their representations of size and depth must be related in the same way to the physical size and depth and (iii) the representations of features in haptic space must map systematically to those within visual space. There have been numerous previous studies of participants' judgements of the perception of spatial relationships on the basis of haptic information (e.g. Teghtsoonian & Teghtsoonian, 1965, 1970; Stanley, 1966, 1974; Jones, 1983; Lanca & Bryant, 1995; Kappers, Koenderink & Oudernaarden, 1997; Kappers, 1999). The degree to which participants' produced veridical estimates of different spatial relationships across these studies has tended to vary with the method of measurement employed, and so the findings share a qualitative similarity to results reported within the visual domain (see section 1.4). For example, studies using magnitude estimation techniques (Teghtsoonian & Teghtsoonian, 1965; Stanley, 1966) found that when participants judged the length of a rod (up to 33 inches long) by simultaneously touching the ends with their two index fingers the results could be described by a power function with an exponent approaching unity (1.05). Conversely, other studies using interval matching and other direct measures have found that there are systematic departures from veridicality within the haptic space (Lanca & Bryant, 1995; Kappers et al, 1997; Kappers, 1999). The precise relationship between estimates of spatial
relationships within visual and haptic space is still an open question but all studies have indicated that observers can produce reliable estimates of spatial extent from haptic information. This clearly indicates that observers within the current study should be able to produce representations of size and depth that are systematically related to the physical dimensions of the hand-held probe though this representation will not necessarily be veridical. Additionally, Lanca & Bryant (1995) reported that when blindfolded observers could feel lines of different lengths at various orientations in the horizontal plane their drawings of these lines could be represented by a power function which was independent of orientation. As a consequence, it would be expected that there should be no difference in observers’ representation of depths and sizes for intervals of equal physical magnitude within the current study. Previous studies of visual-haptic cross-modal judgements of length have found a systematic relationship between settings within one domain on the basis of access to information from the other domain (e.g. Gogel, Wist & Harker, 1963; Davidon & Mather, 1966). Therefore, taken together these studies indicated that observers would be able to perform the task reliably and that their responses would be systematically related to the physical dimensions of the reference.

The experiment also extends the results of chapters 3 and 4 by examining observers’ performance across a greater number of viewing distances. The generality of the conclusions of the previous chapters was based upon the assumption that there was a systematic continuous relationship between perceived distance and depth. The validity of this assumption can only be verified by examining performance across several viewing distances to ensure that the relationship between the two viewing distances used previously was not merely coincidental.
Additionally, the experiment addresses the question of whether the same estimate of the absolute distance is used to calibrate both size and distance estimates. For disparity defined shapes with simulated changes in viewing distance Rogers and Bradshaw (1995b) and van Damme and Brenner (1997) found that the same, although incorrect, estimate of absolute distance was used to correct for size and depth judgements. However, whether the same estimate of viewing distance is used and whether viewing distance is recovered when motion parallax and stereo-motion combinations are considered remains to be determined in the context of real-world stimuli.

5.2 Experiment 5 - Size and Depth Task

5.2.1 Method

5.2.1.1 Observers

The participants were four observers who had taken part in the experiments within chapters 3 and 6.

5.2.1.2 Apparatus and Stimuli

The stimuli consisted of a configuration of three yellow LEDs aligned in the horizontal meridian of the observers' eyes viewed in a dark environment. They formed, when viewed from above, a triangular shaped probe with its base lying orthogonal to the centre line, and its base-to-apex line lying along the line of sight, or along the centre line (depending on the viewing condition). In the current experiment,
all three probe LEDs moved independently. The two base LEDs were moved, by the experimenter under verbal instruction from the observer, along a path orthogonal to the centre line. The base lights were 150, 178, 212, 252 or 300 cm from the origin (observer). The central LED, as within the earlier experiments (chapters 3 and 4), could be moved backwards or forwards along the centre line by the observers. Hence, measures of perceived size and depth could be derived independently. Two ‘T’ shaped hand held reference figures were used in the experiments. The size and depth of each reference was equal, either 15 cm (small reference) or 30 cm (large reference).

Task: adjust probe LEDs to match hand held reference configuration

![Figure 5.1](image.png)

Figure 5.1. Illustrates the procedure for the size and depth task being performed at 150 cm. Observers held, but never viewed, a “T” shaped reference. Their task was to adjust a three LEDs until they formed a shape of the same height and width as the reference. The dotted circle represents an adjustable flashing LED that observers moved along the centreline until it was the same distance from the base LEDs (the white circles) as the height of the reference. The base LEDs could be moved orthogonally to the midline and observers moved them until they matched the width of the reference. The closed circles represent the position of the base LEDs for the four other viewing distances. The diagonal lines projecting from the observer denote a visual angle of either 5.7° (15 cm reference) or 11.4° (30 cm reference).

5.2.1.3 Design and Procedure

The observer’s task was to adjust a probe triangle of LEDs such that both its base width (size) and its base-to-apex height (depth) matched that of a hand held reference (see figure 5.1). As the experiment was conducted in the dark the reference could not be seen by the observers during the course of the experiment. As within chapters 3
and 4 the experiment was conducted under four viewing conditions: (i) static monocular, (ii) static binocular, (iii) monocular motion and (iv) binocular motion. The observer made at least four settings of size and depth for the 15 cm reference and two settings for the 30 cm reference, within all combinations of viewing condition and distance. The experiment was run over several days in twelve trials, which were blocked by viewing distance. The reference size and viewing conditions were randomised within each block.

5.2.2 Results

Figure 5.2 shows the signed mean of all four observers’ size and depth settings for both the 15 and 30 cm reference. The dashed lines indicate veridical performance and so if viewing distance was used correctly to make size and depth settings, the data should lie exactly along these horizontal lines. To assess, and quantify, the reliability and accuracy of the settings, the SD and the mean setting error (see table 5.1) were calculated averaged across the five viewing distances and the two reference sizes (15 cm and 30 cm).

<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Size</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Static Monocular</td>
<td>33</td>
<td>8</td>
</tr>
<tr>
<td>Monocular Motion</td>
<td>28</td>
<td>6</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>25</td>
<td>7</td>
</tr>
<tr>
<td>Binocular Motion</td>
<td>24</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 5.1. Group (n=4) size and depth performance indices (averaged across viewing distance and across the 15 and 30 cm reference figures). The values are expressed as percentages of the veridical setting.
Figure 5.2. Plots the mean size and depth settings, in cm, expressed as a function of viewing distance for the four different viewing conditions: (a) static monocular, (b) monocular motion, (c) static binocular, and (d) binocular motion. The different symbols represent the following experimental conditions: the small triangle base settings (•), the small triangle apex settings (▲), the large triangle base settings (○) and large triangle apex settings (△). The dashed horizontal lines indicate veridical performance. The error bars represent ±1SE. Note the different scale on the ordinate in (a).

From figure 5.2a it is apparent that depth settings in the static monocular viewing condition were erratic and rather inaccurate (note the scale on the ordinate is different from the other graphs in the figure) with even reversals in depth settings (negative values). This poor performance is also reflected in table 5.2 that shows large SD (118%) for the depth settings. Although, the mean setting error was better than one 'depth cue' condition (monocular motion) these figures are averaged over viewing distance, and so disguise the considerable variation over the viewing distances (see
figure 5.2a). Size settings appear much more reliable (SD 8%) although systematically biased (mean 33%) towards underestimates. If judgements of size were just based on maintaining a constant angular size (i.e. no compensation for viewing distance) then settings should have increased by a factor of 2 across the range of viewing distances whereas the ratio of the settings between near and far distances was ~1.5. Taken together, these results suggest that the experimental set-up and controls were sufficient to prevent observers completing the task successfully.

The relatively consistent performance in observers' size settings was surprising and may have arisen from several types of experimental factors that do not depend on the explicit use of viewing distance. One obvious difference between size and depth judgements is that changing the separation of the LEDs to make a size judgement creates a clear difference in the retinal image whereas the movement of the LED along the line of sight (to make a depth setting) does not. In addition, the range of adjustment for the size settings was more constrained (there could not have been a confusion in sign) than for the depth settings. The range of starting positions of the base LEDs could also have contributed to this. The starting position of each LED was each selected at random but the range was constant at each viewing distance (0 to 40 cm) which may, over repeated presentations, have contributed to the apparent constancy (see Glennerster, et al, 1998). Also, the angular movement caused by the LED adjustment varied with viewing distance. Therefore, the level of performance in the size settings may or may not reflect the explicit use of viewing distance. However, if viewing distance did influence the size settings it did not appear to influence the depth settings systematically in the experimental situation.
Figure 5.3. The size to depth ratios of the settings as a function of the viewing distance for each of the three ‘depth cue’ viewing conditions: (a) monocular motion, (b) static binocular, and (c) binocular motion. The symbols denote the 15 cm reference (●) and the 30 cm reference (○). The dashed line indicates a ratio of 1 (veridical performance).

The results from the monocular motion condition are presented in figure 5.2b. Each plot, for both the size and depth settings, is approximately horizontal which suggests that changes in viewing distance were taken into account in making both depth and size settings. Furthermore, depth and size judgements for the references did not differ significantly (F_{1,3}=0.65; p>0.05) which indicates that depth-to-size ratios were near 1 for both sizes of reference (see figure 5.3a). However, the large mean setting error of 24% and 26% and the relative constancy in settings across viewing distance (see figure 5.2b) indicates that both settings were consistently biased. Figure 5.2b shows t
hat this bias a consistent underestimate of the figure size. There are a number of possible explanations for observers' performance. The first account may rest on the fact that it is possible to set the fixed relationship between the size and the depth using only visual information based on Ullman's theorem although absolute size would remain undetermined. If the size were unknown it might be expected that although the size-to-depth ratio remained constant settings would vary considerably both within and across the five viewing distances. Though, of course, it is possible that an estimate of the size of the stimuli could be gained from another source (e.g. angular rotation) and then used to constrain the possible interpretations. An alternative is that 3-d structure from two motion views was scaled by an estimate of the observer's head motion (e.g. angular rotation) appears initially to be inconsistent with the pattern of the results. If the visual system employed this strategy the angular size of the stimuli would be inversely proportional to the absolute distance whereas depth (from motion parallax) would be related to distance squared (see section 1.3.2). As a consequence, the only points at which the estimates of size and depth would be equal are those at which the estimates are veridical (along the dotted lines in figure 5.2). A systematic under estimation of the angular rotation would result in an increased divergence between observers' settings for size and depth with the absolute distance (similar to that in the static binocular condition depicted in figure 5.2c). Therefore, the data in figures 5.2b and 5.3a that indicate a high degree of agreement between systematically mis-estimated size and depth settings are apparently inconsistent with the derivation of these scaling parameters. However, this interpretation of the results assumes that observers are correctly estimating the dimensions of the hand held reference figure. Another possibility is that the internal standard the observers use to make their settings are not veridical but systematically
under estimated. Consequently, the motion information may be perfectly scaled but the resulting setting correspond to an incorrect but constant internal standard. This is supported by the fact that the settings for the two references were both underestimated by the same amount (approximately 67% of veridical) across all viewing distances and so the relative size of the references was preserved. Furthermore, size settings were approximately the same in all three 'depth cue' conditions and, as outlined in section 1.4, it is well documented that observers' perceptual judgements of size on the basis of binocular viewing are generally highly accurate (see for example Gogel, 1960; Baird & Biersdorf, 1967). This interpretation is considered in more detail below.

The settings for the static binocular disparity condition are shown in figure 5.2c. Both size and depth settings were fairly reliable (7% and 12% respectively) although again there were biases in performance. The pattern of the results is clearly different from the monocular motion condition. Although there was no significant main effect when size and depth settings were compared (F_{1,3}=3.1; p>0.05) there was a significant main effect of viewing distance (F_{4,12}=4.91; p<0.05). There was also a significant interaction (F_{4,12}=4.7; p<0.05) which was caused by the progressive divergence of the size and depth plots settings with distance. The depth settings increased monotonically with viewing distance (Pages L=209; p<0.01), with less depth being set at near distances and more at far distances (with veridical settings at about 200 cm). This pattern of results is consistent with the interpretation that vergence angle was used to estimate viewing distance (incorrectly) and is in broad agreement with earlier findings (e.g. Foley, 1980; Johnston, 1991). The increases in the magnitude, in cm, of observers' depth settings with absolute distance in the present experiment are
consistent with an increasing compression of perceived depth in a stimulus with a fixed disparity. There was no significant trend with viewing distance in the size data. These results are reflected in figure 5.3b, which shows that the size-to-depth ratio decreases with increases in viewing distance.

The results from the binocular motion conditions are plotted in figure 5.2d. Here too the size and depth settings were reliable (5% and 11% respectively). This condition is of interest as, theoretically, there should be sufficient visual information available for observers to recover both size and depth veridically. The data show components of the trends apparent in each ‘individual-cue’ condition. In common with the motion-only condition the monotonic increase in the depth settings with viewing distance observed in the static binocular condition was not present (no significant trend, Pages L=194; p>0.05); and in common with the disparity-only condition, the ratio between the size and depth settings was mis-estimated although here it remained invariant with viewing distance (see also figure 5.3c). Depth settings for both sizes of reference did not depart significantly from veridical values (p>0.05). Therefore, although there was some evidence that the addition of motion information affects performance, size settings were underestimated relative to depth settings.

The results from both binocular conditions are apparently inconsistent with those for the shape task (experiment 3.2) because depth is underestimated, and not overestimated, at most viewing distances. However, there are a number of possible ways to reconcile this potential contradiction. Firstly, there are variations in results between previous studies (Foley, 1980; Johnston, 1991, Tittle et al 1995; Glennerster et al, 1996) that suggest although the relationship between perceived and absolute distance is an increasing linear function its slope can vary for a number of reasons, including the nature of the experimental task. Nonetheless, this explanation seems
highly unlikely given the degree of similarity between the two tasks. Secondly, if, as suggested previously, the internal standard were mis-estimated and observers’ size settings are approximately correct with respect to this internal standard. Then the depth settings would indicate an overestimation of perceived depth at each distance.

To quantify the degree to which viewing distance was taken into account in making judgements of size and depth the raw data (in cm) were converted into ‘effective scaling distances’ (Foley, 1980). For depth intervals, the mathematical techniques used were identical to those outlined in chapter 3. For the size intervals, the computation is based on the same assumption that any error in the settings is attributable to a mis-estimate of the viewing distance. The angular size ($\theta$) of a target that is fixated centrally can be derived from its absolute distance from the observer (D) and its physical size in the fronto-parallel plane (p) by the following formula:

$$\tan\left(\frac{\theta}{2}\right) = \frac{p}{D}$$  \hspace{1cm} (11)

The ‘effective scaling distance’ (S) is the distance where the angular size set by the observer would give rise to a veridical (i.e. 15 or 30 cm) setting. By rearranging the above equation and substituting the target veridical setting (v) for the physical size it follows:

$$S = \frac{v}{\tan\left(\frac{\theta}{2}\right)}$$  \hspace{1cm} (12)

Substituting for $\tan(\theta/2)$ from equation 11:

$$S = \frac{vD}{p}$$  \hspace{1cm} (13)

The ‘effective scaling distance’ is plotted against the actual viewing distance for size and depth judgements in figures 5.4a-c. The ratios of the ‘effective scaling distance’ to the physical viewing distance for size and depth judgements are plotted against physical viewing distance in figures 5.4d-f.
Figure 5.4. The graphs plot the ‘estimated’ scaling distance against viewing distance, and the ratio of ‘estimated’ scaling distance to viewing distance as a function of the viewing distance for (a-b) monocular motion, (c-d) static disparity, and (e-f) binocular motion conditions. The symbols indicate the experimental conditions: the small triangle base settings (○), the small triangle apex settings (▲), the large triangle base settings (●) and large triangle apex settings (Δ).
Figure 5.4a and b plot the results of the monocular motion condition. From figure 5.4a, it is apparent that observers took account of increases in viewing distance in making their settings (as shown by the linear increase in the estimated ‘scaling distance’ with viewing distance) although they over-estimated the absolute distance in each case. Figure 5.4b, indicates that they preserve constancy (denoted by a slope of 0) reasonably well but there are fluctuations in settings across the viewing distances. Along with figure 5.3a (which indicated near perfect constancy in the size-to-depth ratios) this result is consistent with the use of Ullman’s theorem alongside an unreliable (variable) estimate of the size (based on scaling the angular size by the angular rotation). In figures 5.4c, d, e and f, which plot results from the static and moving binocular conditions respectively, a trend emerges that would be expected on the basis of figures 5.3b and c. Again, figures 5.4c and e indicate viewing distance was taken into account (increasing linear functions) with observers more accurately estimating depth than size. However, figures 5.4d and f reveal a difference in the patterns of the two conditions. In the static binocular condition, there is a linear decrease in the ratio of the ‘scaling distance’ to viewing distance with increases in viewing distance. This is consistent with an increasing compression of depth with viewing distance (Holway & Boring, 1941; Gogel, 1960; Foley, 1980; Wagner, 1985, Loomis et al, 1992). In contrast for the combined disparity and motion condition, the ratio of ‘scaling’ to viewing distance remained relatively constant across viewing distances. Additionally, the results for the depth estimates were closer to veridical (as indicated by the zero line in figure 5.4f) than for the monocular motion condition (figure 5.4b). These differences in performance from the single cue conditions indicate some degree of interaction between the two cues.
Figure 5.5. The graphs plot the 'estimated' scaling distance against viewing distance, and the ratio of the 'estimated' scaling distance to viewing distance against viewing distance for (a-b) monocular motion, (c-d) static disparity, and (e-f) binocular motion conditions. The scaling distance is calculated from an estimate of each observer's internal standard. The symbols indicate the experimental conditions: the small triangle base settings (●), the small triangle apex settings (▲), the large triangle base settings (O) and large triangle apex settings (△).
Another possibility in the interpretation of these results, as introduced above, is that the correct estimate of viewing distance was used to make the size and depth settings but the *internal standard* used by the observers differed from the physical reference used in the experiments. This possibility is seldom addressed in experiments on depth constancy although it is consistent with the patterns of results presented in figure 5.3 where it is evident that observers set a fixed, but incorrect, physical size irrespective of changes in viewing distance.

To assess this possibility further the data was re-plotted using the same technique as for figure 5.4, but this time the physical size of the reference was replaced in the scaling distance computation by an estimate of the observers' internal standard which was based on the overall average of their size settings (across all three 'depth cue' conditions). The size settings were used because previous studies have generally found estimates of size to be close to veridical (see Holway and Boring, 1941; Gogel, 1960; Baird and Biersdorf, 1967). The fact that the depth/size ratio was close to 1 for monocular motion (although the overall size was underestimated) is reflected in figure 5.5a where the data falls along the leading diagonal. This time estimates of size and depth constancy (from figure 5.5b) average ~103% and 99%\(^2\), nearly perfect performance for both indices. This suggests that changes in viewing distance were compensated for correctly — assuming a different internal standard was used in making the settings in this condition. For the disparity only condition (figure 5.5c and d) and the disparity plus motion (figure 5.5e and f) the results form a similar pattern and suggest that an underestimated viewing distance was used to compute depth. However, depth constancy is much better within the combined condition (the slope in

\(^2\) The constancy ratios are converted into percentages (multiplied by 100) when quoted in the text.
5.5f is close to zero and the constancy was \(-88\%\) than the static binocular condition (depth estimates decreased with distance and were on average \(-85\%\) of the viewing distance). Size estimates were highly accurate as a result of the transformation (constancy estimates of 99% and 98% in the static binocular and combined conditions) which is consistent with previous studies.

![Graphs showing scaling distance as a function of viewing distance](image)

**Figure 5.5.** Plots the scaling distance as a function of the physical viewing distance at which the ratio of the size and depth settings were equal to 1 for each of the three 'depth cue' viewing conditions: (a) monocular motion, (b) static binocular, and (c) binocular motion. The different denote the 15 cm reference (●) and the 30 cm reference (○). Symbols are the same as those in figure 5.3. The dashed line indicates perfect shape scaling (i.e. the set of shapes with a size to depth ratio of 1).

Of course it is also possible that different internal standards were adopted both for the size and for the depth settings in these conditions. If so, then the settings for size and
depth in both the binocular conditions would also fall near to the leading diagonal in a similar manner to the monocular motion condition. It seems more parsimonious to assume that observers used the same standard to judge both size and depth. This was their impression. Given this assumption, the data from figure 5.3 can be re-plotted using the technique described in Johnston (1991) to determine the distance where the angular size and disparity set by the observer reflected an object with a size/depth ratio of 1 (regardless of actual size). The results of this analysis are shown in figure 5.6 (which re-plots the data from figure 5.3 in terms of scaling distance).

Figure 5.6a shows the results from the monocular motion condition and indicates that the correct size/depth ratio was set irrespective of absolute size as would be expected on the basis of figure 5.3a. The average slope of the best fitting straight lines is 0.9. In the binocular moving condition (figure 5.6c) an average slope of 0.77 was found compared to 0.5 for the static binocular condition (figure 5.6b). Figure 5.6 clearly shows that the addition of motion information improves performance over the disparity-only condition. Although the disparity also affects performance as settings are less accurate than in the motion-only condition (where several ways of determining size/depth ratios may be available). The results in figure 5.6b and 5.6c are consistent with an interpretation based on mis-estimates of viewing distance or a mis-estimate of internal standard as both types of error cause a change in slope.

If observers did make settings that corresponded to the size of the physical reference, then biases in performance can be attributed to mis-estimates in viewing distance. In this context it appears from figures 5.4 and 5.5 that different estimates of viewing distance were used in making size and depth judgements. This finding is consistent for the two sizes of reference used. Figure 5.7 plots the relationship between the
'effective scaling distance' computed for size and the effective scaling distance computed for depth (as plotted in figure 5.4) in each viewing condition (not including static monocular) to make this relationship explicit. This figure shows how the size and depth settings scale together as opposed to how each scale with viewing distance (as shown in figures 5.4 and 5.5). If the same estimate was used in both judgements then the data should coalesce along the leading diagonal. Clearly there is a strong correlation between the scaling distances used for size and depth.

![Graph showing effective scaling distance for size settings against depth settings for various viewing conditions.](image_url)

**Figure 5.7.** Plots effective scaling distance for size settings (ordinate) against those for depth settings (abscissa) for the monocular motion (○), static binocular (□) and binocular motion (♦) conditions. The full line indicates the best fitting straight line across the three conditions. Point lying on the leading diagonal would indicate perfect correspondence between the distance estimates used for both judgements. The hatched lines on either side of the leading diagonal indicate mis-estimates of the absolute size of 60% (slope greater than 1) and 140% of the veridical size.

Van Damme and Brenner (1997) presented a similar figure. However although they found a close correspondence between the distances used for size and depth it appears the current experiment did not. Rather, the scaling distance used for size was systematically overestimated relative to that used for depth in all 3 viewing conditions. This result was also evident in some of the data plotted in figure 5.4. Before reaching this conclusion, however, the possibility, discussed earlier, of a different internal standard being used, should also be addressed. The hatched lines in
the figure indicate how the results would appear if absolute size of the two references was underestimated (slope greater than 1) or overestimated (slope less than 1). There is a close correspondence between the line of best fit (full line in the figure 5.7) and the hatched line that depicts the outcome of underestimating the reference sizes by 60%. This suggests, in common with the arguments above, that the results are equally consistent with the use of a different internal standard and that similar estimates of viewing distance may have been used in making both size and depth settings. Taken together with the arguments raised earlier, and previous empirical results on this issue, this conclusion seems likely.

5.2.3 Discussion

The present experiment was designed to determine the degree to which size and depth constancy are maintained on the basis of binocular disparity and motion parallax information, presented separately and in combination, using (quasi) natural stimuli and viewing conditions. The results suggest that estimates of size and depth were not maintained accurately when either disparity or parallax are the principle depth cues — presented separately or together (cf. figure 5.2 and table 5.1). In terms of depth constancy, monocular motion produced a reasonably constant overestimation of viewing distance (~115%) whereas on the basis of binocular disparity they declined with the physical viewing distance though on average they were more accurate (97%). When both disparity and parallax were present, which through their combination provides further information for the recovery of both size and depth veridically, they produced a slight overestimation of viewing distance (103%) but depth constancy was
higher. In all viewing conditions size constancy was good but their scaling distances were consistently underestimated. This pattern in the data led to an alternative interpretation of the observers’ performance that was based on the premise that an incorrect standard may have been used to make settings. In this light, the size and depth settings were very good and estimates of size and depth constancy in all three conditions were close to 100%.

In the monocular motion condition the correct size to depth ratio was recovered although both the size and the depth were underestimated at each of the viewing distances. The fact that there is not a proportional relationship between size and depth scaling as a function of distance suggests that a scheme based on estimating ego-rotation was not used here. Rather, this result at first sight appears more consistent with Ullman’s scheme of recovering rotation angle from visual information (higher order motion parameters) which leaves absolute size undetermined. However, if an incorrect standard (67% of veridical) was used in making the settings then perfect size and depth constancy were maintained in this condition (see figure 5.5) and the pattern of results is equally consistent with viewing distance being recovered via ego-motion. Though, it also remains possible even if the internal standard was underestimated that Ullman’s scheme was used along with an estimate of scaling distance. This would explain the very good size to depth ratios (figure 5.3a) for the monocular motion condition compared to the binocular conditions.

In the static binocular condition size to depth ratios were misperceived with progressively more depth being set in the stimulus at farther distances. The size settings, in both conditions, remained reasonably invariant with changes in viewing distance. These results are consistent with the interpretation that observers used a
mis-estimate of viewing distance to scale the disparity information to recover size and depth (Foley, 1980; Johnston, 1991; Rogers & Bradshaw, 1993). This failure is most likely due to the fact that vergence angle cannot be reliably estimated in low-light conditions (see Foley, 1980; Owens & Leibowitz, 1980) and vertical disparities were not available. The bias introduced by this is consistent with the ‘specific distance tendency’ in which the distance of far objects are underestimated and the distance of close objects are overestimated (see Johnston, 1991).

When both disparity and motion cues were available which provides, in principle, sufficient information to recover veridical size and depth information (Richards, 1985) the pattern of results appeared to share important features of both the motion- and disparity-only conditions. In common with the motion-only data, size and depth judgements were consistent across viewing distances; and in common with the disparity-only data, size/depth ratios were less than 1. It is clear that the size/depth ratios of unity obtained in the motion parallax conditions were not used to constrain the settings in the combined cue condition, which Richards’ proposal would predict.

When the raw data were converted into ‘effective scaling distances’ depth constancy in the combined-cue condition was estimated to be 93% compared to 106% and 75% in the motion and disparity conditions respectively. Again depth-from-disparity and depth-from-motion appear to be computed quite separately and averaged together at the end. This pattern of results is actually quite well described by weak fusion (Landy et al, 1995) that is, the perceptual outcome is a simple average of the outcomes in the two isolated conditions. The fact that Johnston et al (1994) found evidence for a strong fusion interaction in their 2-frame binocular motion conditions does not necessarily contradict the results. In their conditions this was the only
strategy open to the visual system to complete the task whereas here, the results from the individual-cue conditions suggest that sufficient information was available to recover size and depth information. This is consistent with the suggestion that the visual system may switch strategies based on the information available and according to the demands of a particular task (Rogers & Bradshaw, 1995a; Glennerster et al., 1996). Therefore the estimate of size and depth constancy were high, particularly with the transformed data which incorporated the assumption that an incorrect standard was used in making settings. The magnitude of these constancy estimates, and those within chapter 3, are similar to those found by Rogers and Bradshaw (1995a), Bradshaw et al (1996) and Glennerster et al, (1996) whereas the magnitude of the mean errors are more typical of those found by Norman et al (1996). However as noted earlier, these studies used a range of different tasks.

A further question addressed by the experiment was whether the same estimate of viewing distance was used to scale information about size and depth. This has been the focus of several recent studies using computer displays. A high positive correlation was found between the effective scaling distances for the size and depth judgements although the slope of the best fitting regression line was considerably greater than 1. At first sight, therefore, the scaling distances used for size and depth seemed different. However, when the possibility that observers may have been using an incorrect standard (67% of actual size) in making their settings is taken into account then the data suggest similar estimates of distance were used in both cases. This is because of the nature of the scaling process: size is inversely proportional to distance whereas depth is inversely proportional to the distance squared. Certainly
this conclusion is consistent with that made by Rogers and Bradshaw (1995b) and van Damme and Brenner (1997).

In conclusion, the experiment found that motion information generated by head movements, is sufficient to recover size/depth ratios correctly although overall size was consistently underestimated. The size/depth ratios were misperceived in the binocular disparity condition with progressively more depth being set in the stimulus at farther distances. When both disparity and parallax were available together, the perceptual outcome was consistent with an averaging of the information provided by either cue when presented separately. However estimates of size and depth constancy were high which indicated that changes in viewing distance were taken into account when settings were made in each of the viewing conditions. The consistent pattern of underestimation in the size settings suggested that observers might have used an incorrect version of the standard size to make their settings. When the data were interpreted in this light, settings were accurate and size and depth constancy was remarkably good. In experiments on depth and size constancy many experimental, stimulus and observer factors can affect performance (Glennerster et al, 1996). In accounting for the markedly different estimates of our ability to recover size and depth on the basis of disparity and motion cues the story is certainly more involved than the issue of natural versus computer generated stimuli.
CHAPTER 6

The Recovery of the 3-d Structure of Physical Stimuli from Binocular Disparity Information within an Illuminated Structured Environment.

6.1 Introduction

The experiments in chapters 3, 4 and 5 demonstrated that binocular disparity and motion parallax information did not alone, or in combination, support a veridical percept of depth magnitude when presented in isolation from other sources of information. These findings were consistent with those of many previous studies using both real and simulated stimuli that have argued that these cues cannot be used to recover metric scene structure (e.g. Foley, 1980; Todd & Bressan, 1990; Collett et al, 1991; Johnston, 1991; Todd & Norman, 1991; Liter et al, 1994; Glennerster et al, 1996, 1998; Rogers & Bradshaw, 1993, 1995a; Tittle et al, 1995; Todd et al, 1995; Norman et al, 1996). However, Durgin et al (1995) and Frisby et al (1996) found considerably better performance and attributed deviations from constancy within many previous studies to the poor range of cues available within the experimental viewing environment. In the current chapter two experiments are presented that examined observers' judgement of depth on the basis of binocular disparity or motion parallax information presented in conjunction with a range of additional depth cues. The first experiment examined judgements using the same stimuli as the earlier experiments but viewed within a fully illuminated structured viewing environment.
The second experiment assessed judgements of the shape of triangular ridges viewed in the dark but with a range of surface texture cues available.

Durgin et al (1995) and Frisby et al (1996) attributed the departures from veridical perception found within previous studies of shape and depth constancy to three main factors (either alone or in combination): (i) impoverished experimental viewing environments which remove depth and distance cues available in normal viewing (e.g. perspective cues), (ii) artefacts resulting from the use of computer generated stimuli (e.g. conflicts from accommodation) and (iii) the use of artificial discontinuous stimuli (i.e. many previous studies have focused upon shapes defined by points, or rods, at their apices rather than continuous surfaces).

Durgin et al (1995) claimed that when viewing simulated or real surfaces within an illuminated environment the available cues facilitated the precise recovery of viewing distance and hence the calibration of binocular disparities (though these cues did not similarly aid the recovery of angular rotation and the calibration of motion information). In particular, they emphasised the importance of perspective cues present on the ground plane which had previously been established as an effective cue to absolute distance (O'Leary & Wallach, 1980). However, the degree to which parallax and disparity information can support veridical perception even in the presence of these additional cues is open to question. Studies conducted outside in fields (Wagner, 1985; Loomis et al, 1992) and in illuminated structured environments (Glennerster et al, 1996; Norman et al, 1996) found systematic distortions of depth persist despite the presence of a wide range of cues. There are, however, at least two differences between these experiments and those of Durgin et al. Firstly, distance
judgement experiments conducted in fields were generally conducted at much greater viewing distances than those within laboratory based studies. The relative effectiveness of different cues is a function of the absolute viewing distance (Cutting & Vishton, 1995). Hence, a combination of cues that supports a veridical percept of depth at close viewing distances might not be as effective at greater viewing distances. Secondly, the use of discontinuous stimuli (Wagner, 1985; Loomis et al, 1992; Norman et al, 1996) rather than a continuous surface.

Durgin et al's claim was assessed within the first experiment which examined depth judgements at five viewing distances between 1.5 and 3 metres in a structured well illuminated environment. These viewing distances were similar to those of Durgin et al but closer than those for the studies conducted in fields (Wagner, 1985; Loomis et al, 1992). The large number of viewing distances employed facilitated the detection of small trends deviating from constancy.

An aim of both of the experiments was to remove any potential artefacts arising from accommodation that may have been present in earlier experiments. Buckley, Frisby and their collaborators have in a series of papers criticised the interpretation of results from studies using computer simulated stimuli for failing to consider the importance of accommodation (Buckley & Frisby 1993; Frisby, Buckley & Horsman, 1995a; Frisby et al, 1996; Brennand, Buckley, Davies & Frisby, 1998). Specifically, these authors claimed that problems arise due to conflicts in either the relative depth (Buckley & Frisby, 1993) or the absolute distance (Frisby et al, 1995a; Brennand et al, 1998) signalled by accommodation and that signalled by binocular cues (disparity or vergence).
Physical stimuli, such as those used within the current thesis, should in principal contain no conflicts between accommodation and other depth cues. However, in practice previous studies (Liebowitz & Owens, 1975; Owens & Liebowitz, 1980) have shown that when isolated point light sources are viewed in the dark accommodation can default to a resting state. Although the exact position of this resting state is subject to considerable individual variation it typically averages around 59 cm (Liebowitz & Owens, 1975). Brennand et al (1999) argued that individual variation in observers' biases within depth estimation tasks conducted in dimly lit environments is caused by conflicts between stereoscopic cues and information signalled by the resting state of accommodation.

The first experiment employed physical stimuli within a well-illuminated environment and the second experiment used physical surfaces defined by a large number of LEDs. As a consequence, the stimuli within both experiments provided effective accommodative information that was consistent with the other distance and depth cues. Therefore, any systematic biases that exist within these studies cannot be attributable to conflicting information signalled by accommodation.

The second experiment addressed Frisby et al's suggestion that their use of continuous stimuli may have been partially responsible for the relatively low magnitude of observers' setting errors within their experiments. There are in principal two main ways in which this factor might effect judgements. Firstly, continuous surfaces are potentially a more effective stimulus for accommodation than point light sources (see above). Secondly, physical continuous surfaces may contain additional depth or slant information from texture cues that can modify disparity based depth
estimates (Buckley & Frisby, 1993). This latter factor was examined explicitly in the second experiment by comparing the judgement of surfaces, viewed either binocularly or monocularly with head movements, which varied in the effectiveness of the texture cues they provided.

The accommodative state of the observers’ eyes was not, however, manipulated within the current experiment due to the practical limitations in doing this whilst viewing dynamic real world stimuli. There are two techniques for manipulating accommodation whilst viewing physical stimuli: (i) the use of a drug that paralyses the ciliary muscles (e.g. homatropine) so not allowing changes in the accommodative state of the eyes or (ii) an artificial pupil that increases the depth of field to bring all points within the scene into focus. The former is impractical due to pragmatic difficulties in gaining ethical approval and access to homatropine. The latter limits the utility of accommodation rather than to introduce the conflicting information, which Frisby et al (1996) claimed caused departures from constancy. Furthermore, the existence of departures from constancy in situations where accommodation information is fully consistent with other depth cues is sufficient to demonstrate that it cannot be their sole cause.

In summary, the two experiments presented within this chapter assess the degree to which failures of constancy found in the earlier experiment can be explained by the impoverished range of available cues within the viewing conditions. They specifically examine the effects of introducing different of monocular pictorial cues upon stereoscopic and parallax based depth judgements.
6.2 Experiment 6.1 - Illuminated Structured Environment Task.

In experiments 3.2 and 5 it was demonstrated that observers could not derive a metric representation of scene structure on the basis of binocular disparity information presented in isolation from other information sources. The results showed that under those viewing conditions the observers' perception of shape was systematically biased and this was attributed to the mis-estimation of the appropriate scaling factors. One question left open by these experiments is the degree to which observers can recover a veridical estimate of shape when a greater range of cues is available.

The current experiment examined the degree to which the addition of a range of monocular pictorial information, including perspective and familiar size cues, affected observers' percept of shape in a task matched to that used within experiment 3.2. It assessed the degree to which the task could be performed on the basis of pictorial cues alone (monocularly) and the degree to which they enhanced performance in conjunction with binocular disparity information. A greater range of distances was used than employed in the previous shape task (experiment 3.2) to help identify small departures from a veridical percept. Any departures from constancy would, if they existed, almost certainly be of a lower magnitude than those found within the dark viewing environment due to the increase in the range of available information sources.

Observers viewed stimuli on a table top that provided three sources of ground plane perspective cues: (i) the natural texture (grain) of the wooden table which stretched away from the observer along the line of sight approximately parallel to the centre
line, (ii) the edges of the table and the tracks along which the adjustable LED moved (which were also parallel to the centre line) and (iii) a regular pattern of horizontal lines orthogonal to the centre line that were added to the surface of the table. The familiar size cue arose from observers viewing the stimuli and its supporting framework at different distances and so comparing the change in its angular size across the viewing distances (Predebon, 1992). Additionally, as the experiments were performed in a fully illuminated environment any information signalled by accommodation would be fully consistent with other depth cues. Hence, there was a rich array of monocular pictorial information sources available within the viewing environment. The use of a monocular condition allowed an assessment of the degree to which the task could be performed on the basis of the pictorial cues alone and consequently whether the presence of these cues actually enhanced the accuracy of disparity based depth estimates.

6.2.1 Method

6.2.1.1 Observers

Four observers participated as unpaid volunteers within the experiment. Three had participated within the experiments in chapters 3 and 4. The fourth was an experienced psychophysical observer who was naïve to the experimental procedure.

6.2.1.2 Apparatus

The stimuli consisted of configurations of 3 bright yellow LEDs, which were
presented in a well-illuminated room (see chapter 3). The LEDs were carefully
aligned in the horizontal meridian of the observer’s eyes (28 cm above the surface of
the table) and were centred in a 4 cm square piece of black cardboard. The adjustable
LED moved along tracks lying down the centre line. The two fixed LEDs were each
attached to separate horizontal steel bars (1 cm diameter) which in turn were joined to
similar vertical bars clamped to the table either side of the centre line. For each trial
the LEDs were positioned either 7.5 (small triangle) or 15 (large triangle) cm either
side of, and orthogonal to, the centre line. The base lights were 150, 178, 212, 252 or
300 cm from the origin (observer). The natural texture cues of the table’s surface
were supplemented by strips of blue tape (1.5 cm wide) placed at a regular series of
intervals orthogonal to the centre line. The initial strip was positioned 35 cm from
the observer with subsequent strips positioned at 25 cm intervals along the table.

6.2.1.3 Design and procedure

As within the shape task (experiment 3.2), the observer was required to move an
adjustable LED in front of two fixed LEDs, on either side of the centre line, such that
the interval \((h)\) between it and their centre point was the same as their separation \((b)\)
(see figure 3.3). The experiments were performed under two viewing conditions: (i)
static monocular (control) and (ii) static binocular. Each observer made 6 settings for
both viewing conditions and base sizes (15 and 30 cm) at all 5 viewing distances. The
observer’s eyes were closed between trials. The experiments were performed in 4
blocks of 30 trials run over several days. The 2 viewing cue conditions, 5 viewing
distances and 2 stimulus sizes were presented in a random order across all 4
6.2.2 Results

The settings which observers made for the large (30 cm) base size were halved in order to convert them to the same scale as those for the small (15 cm) base size (see experiment 3.2). There were no theoretical expectations for the settings for the two triangles to differ and no significant difference ($F_{1,3}=3.39$, $p>0.05$) was found. Therefore, data for the two base sizes was combined for the purposes of further analysis. Figure 6.1 presents the signed means, in cms, for the four observers' settings with the error bars representing the standard error of the means. Table 6.1 shows the accuracy (mean setting error) and reliability (SD) of the settings expressed as percentages of the veridical value averaged across the five viewing distances.

![Figure 6.1](image)

*Figure 6.1.* Plots the mean (n=4) depth settings, in cm, as a function of the viewing distance with error bars representing ±1SE. The different symbols represent the individual viewing conditions: static monocular viewing (●) and static binocular viewing (▲). The dashed horizontal line indicates ideal (veridical) performance at both viewing distances. Note the scale of this graph is much smaller (50%) than that used for the shape task (experiment 3.2).
<table>
<thead>
<tr>
<th>Viewing Condition</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Monocular</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>Static Binocular</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 6.1. Group (n=4) depth judgement performance indices (expressed as a percentage of the veridical setting).

The performance indices in table 6.1 demonstrate that observers’ settings were least accurate and reliable within the *monocular condition*. There was no significant difference in observers’ mean settings between the two viewing conditions ($F_{1,3}=0.11$, $p>0.05$). However, as the observers’ sensitivity within the monocular condition was considerably less than for the static binocular condition (table 6.1 shows that the observers’ average SD differed by a factor of almost 3) it was concluded that the task could not be completed optimally on the basis of monocular cues.

A comparison of the current task with the shape task (section 3.3) shows that the additional cues from a fully illuminated viewing environment resulted in a considerable improvement in accuracy and sensitivity within both conditions. Unlike the shape task, one-way t-tests revealed that neither condition exhibited a significant bias ($p>0.05$) from zero at any of the viewing distances. However, there was a highly significant difference between mean settings at different viewing distances ($F_{4,12}=7.59$, $p<0.005$). Post-hoc tests found that there was a significant trend within both monocular (Pages L=212; $p<0.001$) and binocular (Pages L=209; $p<0.01$) conditions for settings to increase as a function of the viewing distance. This clearly demonstrates that perceived depth intervals are systematically compressed with distance despite the availability of a large range of monocular and binocular depth cues.

The results were converted into ‘estimated scaling distances’ (as outlined in section
3.3.2) and are plotted as (i) ‘viewing distance’ against ‘estimated scaling distance’ (figures 6.2a), and (ii) the ratios of the ‘estimated scaling distance’ to the ‘viewing distance’ (figures 6.2b). Although, the monocular condition does not contain disparities it was converted using the same formula to quantify the degree to which observers could take distance into account on the basis of the monocular cues.

The results in figures 6.2a and 6.2b confirm that observers’ settings exhibit a relatively small, but systematic, departure from a veridical percept and they appear to take the different viewing distances into account. However, observers do not exhibit perfect constancy in either condition (on average 96% and 95% in monocular and binocular conditions respectively) as their biases tended to increase with the viewing distance. They are, however, considerably more accurate than within the dark viewing environment (the average constancy within the binocular condition was
6.2.3 Discussion

The results from the current experiment show that monocular pictorial cues can improve the estimation of distance and relative depth when viewing physical stimuli. However, the observers' settings also demonstrate that systematic distortions still exist in their perception of physical space. This provides further support for the claim that visual space is non-Euclidean and that the perceived depth of a fixed interval declines as the absolute viewing distance increases (e.g. Foley, 1980; Todd & Bressan, 1990; Collett et al, 1991; Johnston, 1991; Todd & Norman, 1991; Liter et al, 1994; Glennerster et al, 1996, 1998; Rogers & Bradshaw, 1993, 1995a; Tittle et al, 1995; Todd et al, 1995; Norman et al, 1996).

Observers' performance within the binocular viewing condition was more accurate than within the monocular condition. Furthermore, as their stereoscopic performance was considerably better than observed earlier on the basis of disparity information alone (see the shape task section 3.3) it is possible to conclude that disparity and pictorial cues interacted. The considerably improved constancy suggested that the increased accuracy was due to an improvement in observers' estimates of viewing distance. Additionally, the deviations from constancy cannot be attributable to conflicting information from accommodation as no such conflicts existed within the viewing conditions.
The results within the binocular condition (Mean and SD of 10 and 10 respectively) were in general of a similar magnitude to those found by Frisby et al (1996) (Mean and SD of 11 and 8 respectively; see their experiment 6) who concluded that binocular viewing could support accurate representations of the metric scene structure. The difference between their conclusions and those of the current experiment partially reflected the difficulty of determining generally accepted criteria of 'good' performance. An examination of the Frisby et al's data (see figure 10 pp. 151) revealed that they found a small trend consistent with the mis-estimation of scaling distance. All of the depth intervals at the far viewing distance (355 cm) were overestimated and those at the closer distances (78 and 158 cm) were both underestimated. Furthermore, those at the closest distance generally produced the largest underestimates. Consequently, it is possible to conclude that though their data indicated more accurate performance than many other studies it is still entirely compatible with a systematic compression of visual space.

Although observers' performance within these experimental conditions was better than for the dark viewing environment it still fell far short of the extremely high levels of constancy found by Durgin et al (1995). The greater accuracy of observers within their study might be attributable to the stimuli being continuous surfaces rather than a shape defined by LEDs at its apices, this issue is addressed within experiment 6.2.

In summary, the current experiment provides further evidence to support the suggestion that the accuracy and reliability of observers' estimates of depth magnitude are dependent on the range of available visual cues (Cutting & Vishton, 1995). Although the results demonstrate that monocular pictorial cues can improve the
accuracy and reliability of observers' depth settings because several cues were introduced simultaneously it is not possible, purely on the basis of these results, to determine the relative benefits afforded by any particular information source, e.g. the relative benefits of perspective and accommodation. Although Mather (1996, 1997) has reported findings that casts doubt on the utility of accommodation within the ranges of distances and depths used within these experiments (see chapter 7 for a more detailed discussion). However, the results show that even in a multi-cue viewing environment observers can still demonstrate systematic deviations from veridical perception.

6.3 The Derivation of Surface Slant from Texture Information

Gibson (1950) noted the significance of surface texture as a potential source of information about surface slant or inclination. However, he did not attempt to specify formally the different sources of information that are potentially available within a textured surface. For randomly patterned surfaces, Cutting & Millard (1984) have proposed that there are three main sources of available texture information: (i) density which is the tendency of texture elements on a surface to become more closely packed together with distance, (ii) compression which is the tendency of the aspect ratio (width-to-length ratio) of texture elements to change on a curved surface with distance or to change with increasing orientation in depth for a planar surface and (iii) scaling which is the tendency of the retinal size of a texture element to decrease as the absolute distance increases. These can all be considered instances of perspective
information which Howard & Rogers (1995) define as the reduction in the size of the retinal image of an object in the frontal plane with increases in its absolute distance. They distinguished three further sources of information available in surfaces containing a regular structured pattern (e.g. one comprised of horizontal or vertical line elements): (i) **image elevation** which is the tendency of parts an image above and below the fixation point to draw closer to the fixation point with increases in absolute distance, (ii) **foreshortening** which is the tendency of vertical line elements on a surface inclined about a horizontal axis (or horizontal line elements on a surface slanted about a horizontal axis) to become shorter as absolute distance increases and (iii) **linear perspective** which is the tendency of parallel vertical line elements on a surface inclined about a horizontal axis (or horizontal line elements on a surface slanted about a vertical axis) to converge as absolute distance increases. In addition to the internal texture of a surface, the outline of a regularly shaped surface could also provide perspective information. For example, the aspect ratio of the outline of a circular surface will become increasingly compressed as it is increasingly orientated in depth, and the outline of a rectangular surface can contain image elevation, foreshortening and linear perspective cues.

Although, many sources of potential information exist within textured surfaces they might not necessarily be exploited by the visual system in practice. Several studies have attempted to assess the relative importance of different texture cues both alone and in combination with disparity information (e.g. Gillam, 1968; Cutting & Millard, 1984; Todd & Akerstrom, 1987; Stevens & Brookes, 1988; Buckley & Frisby, 1993; Cumming et al, 1993; Blake et al, 1993; Johnston et al, 1993 Frisby et al, 1995b).
Gillam (1968) found a high degree of accuracy for observers' monocular judgements of the slant of a planar surface about a vertical axis when it was covered with a grid or horizontal lines, but that their estimates were poor for surfaces covered with a random texture or vertical line elements. Therefore, Gillam concluded that the major determinant of perceived slant within those surfaces was linear perspective and that observers could make little use of foreshortening information. In a further experiment Gillam used a cue conflict paradigm to assess the interaction of texture cues and binocular disparities. Observers viewed surfaces in which the texture information was consistent with a flat fronto-parallel surface whilst disparity was varied using an anisekonic lens to simulate slants between 0° and 30°. The experiment was performed for two texture conditions, horizontal and vertical lines. For the surfaces with vertical line elements depth was almost entirely determined by the disparity cue. However, observers consistently underestimated the slant of surfaces containing only horizontal line elements. This demonstrated that binocular disparity and linear perspective information interact in the determination of surface depth. However, Stevens and Brookes (1988) challenged the generality of this conclusion by arguing that texture is only effective when a surface exhibits either a depth discontinuity or curvature. Cumming et al (1993) examined the interaction between binocular disparity and three texture cues (density, compression and scaling). They found compression accounted for 97% of the variance when the three texture cues were manipulated independently on a surface covered with circular texture elements. Blake et al (1993) also found that texture density was relatively ineffective for a stereoscopically generated parabolic cylinder in which all of the depth cues were consistent.
Johnston et al (1993) examined the strategies employed by the visual system to integrate information from binocular disparity and texture information using three stimulus shapes: (i) half-cylinders (orientated around either vertical and horizontal axes), (ii) two intersecting planes and (iii) an ellipsoid. In their first experiment they examined the effect of adding a congruent and incongruent (consistent with a flat fronto-parallel surface) random texture. They found that adding texture increased the amount of perceived depth and may at close viewing distances even cause the observers' percept to diverge further from veridical. In a second experiment they varied the texture cue in order to assess the way in which texture and binocular disparities were combined. They found that their data was well described by a weighted linear combination rule similar to that outlined by Landy et al (1995) (see section 1.1). The weighting of texture cues increased with the absolute distance, which is consistent with Cutting and Vishton's (1995) suggestion that depth cues are differentially weighted at different distances.

In summary, texture information has been shown to interact with disparity in the judgements of slant for vertically orientated surfaces that exhibit a discontinuity or curvature. The most effective texture cues in previous studies were linear perspective and compression. Conversely, observers appeared to receive little benefit from either foreshortening or changes in texture density.
6.4 Experiment 6.2 - Textured Surface Task

This experiment examined observers' judgements of surface shape on the basis of texture cues in combination with disparity or parallax information within a dark viewing environment. Observers viewed two planar surfaces of LEDs hinged around a vertical axis and then set the angle of two adjustable lines on a computer display to match the intersection angle of the surfaces. The primary aim of the experiment was to assess the degree to which observers were able to recover a veridical percept of surface slant on the basis of each of the different cue combinations.

The texture information was manipulated by comparing performance for a regularly patterned surface (25 LEDs equally spaced within a 25 cm square grid) with performance for a randomly patterned surface (25 randomly spaced LEDs within a 25 cm square). For the regularly patterned surface, there was a strong outline cue and a number of texture cues available: (i) image elevation as the surface projects above and below the meridian of the eyes, (ii) foreshortening as the LEDs were grouped in vertical lines and (iii) linear perspective as the LEDs were also grouped in horizontal lines. The results from previous studies (see section 6.3) would predict that the linear perspective should provide an especially effective slant cue within these surfaces. For the randomly patterned surface the only one of these cues available was image elevation. In principal compression, density and scaling cues were available on both surfaces. However, the effectiveness of the compression and scaling cues were minimised due to the small size of the texture elements (5 mm at distances of 150, 212 and 300 cm) and the difficulty of precisely estimating the size of a bright light source.
In addition, previous studies found little evidence that the visual system makes use of variations in texture density on a randomly patterned surface. There were two further cues potentially available on both surfaces which were not present in the experiments within chapters 3, 4 and 5: (i) vertical disparities (within the binocular condition) and (ii) accommodation (the array of lights provided a more effective stimulus for this cue). However, both cues were available within each surface and so any differences between the surfaces could only be attributed to variations in the available texture information.

6.4.1 Method

6.4.1.1 Observers

Nine observers participated in the experiment with normal, or corrected to normal, visual acuity and were able to distinguish depth in the standard Frisby stereo test. All were naïve to the purpose of the experiment and seven had never previously participated in a psychophysical experiment. Five observers were undergraduate students who received research credits for participating and the other four participated as unpaid volunteers.

6.4.1.2 Apparatus and Stimuli

The layout of the experimental apparatus is illustrated in figure 5.3. The test stimuli consisted of two planar reference surfaces of LEDs hinged about a vertical axis viewed in the dark. When viewed from above the intersecting surfaces formed a forward
pointing triangle. Each surface comprised 25 LEDs either forming (i) a *regular pattern*, with the LEDs equally spaced within a 25 cm square grid, or (ii) an *irregular pattern* with the LEDs randomly arranged within the same area. The irregular pattern was derived from the regular surface by randomly shifting each of the LEDs between 3 and -3 cm horizontally and then between 3 and -3 cms vertically. The LEDs on each surface were wired into one of four parallel circuits (chosen at random) and the luminance of these circuits was varied at random throughout the experiment (by potentiometers wired into each circuit). The surfaces were positioned such that the central row of LEDs were aligned in the horizontal meridian at the observer’s eye height and the axis around which they were hinged lay down the centre line.

The observer’s perception of the intersection angle of the surfaces was measured using a computer generated probe, which comprised the angle formed at the apex of two lines. These lines could be adjusted by the observer to form an angle between 0° (vertical) and 180° (horizontal), and they always moved symmetrically about their apex. The initial apex angle of the line was set randomly between 30° and 150°. They were generated by Macintosh LC III and displayed on a 14” Apple monitor that was positioned behind the reference surface, 350 cm from the observer, and 60 cm above the horizontal meridian of the observer’s eyes.

A cardboard screen attached to a step motor screened the experimental set-up from the observer between trials. This was lowered after the experimenter had set-up each trial, when the room lights were switched off, and raised once the observer had completed the trial, when the lights were switched on.
The observer viewed the LEDs from a headrest which was either fixed for static viewing or free to move from side-to-side (see chapter 4). When viewing stimuli statically, the headrest was fixed such that the centre line of the stimuli was aligned with (i) the dominant eye (monocular viewing) or (ii) midway between the eyes (binocular viewing). In motion conditions the observer's head moved 6.5 cm either side of the centre line (twice the interocular distance) at a rate of 1 Hz, which was synchronised by a metronome.

**Task: match probe on monitor to angle of surface**

![Figure 6.3](image.png)

*Figure 6.3.* Illustrates the procedure for the textured surface task. The task is being performed at 150 cm with the regularly patterned surface (the pattern is denoted by open circles that represent illuminated LEDs). The observer's task was to use a keypad to adjust the angle between two 2-d lines on a computer monitor to match the angle of intersection of the two planes of the hinged 3-d surface. The black vertical lines represent the axis around which the surfaces are hinged at each viewing distance.

6.4.1.3 Procedure

The observer's task was to adjust the apex angle of two computer generated 2-d lines to match the intersection angle of two 3-d reference surfaces, defined by a series of LEDs, hinged around a vertical axis and slanted in depth away from them (see figure 6.3). These 3-d surfaces intersected at one of 5 angles selected at random (60°, 80°,
90°, 100° or 120°) and contained either a regular or an irregular configuration of LEDs. The apex angle of the lines could be adjusted (inwards and/or outwards) using the numeric keypad of a standard Macintosh keyboard. After completing each trial a screen was raised and a desk light switched on. This prevented the observer adapting to the dark significantly during the course of the experiment and allowed the experimenter to set up each trial.

Each observer completed 2 settings for the 5 surface angles and 2 texture patterns at each of the 3 viewing distances (150, 212 and 300 cms). These were all performed under 2 viewing conditions: (i) static binocular, and (ii) monocular with head motion.

6.4.2 Results

Figure 6.4a-e presents the signed means, in degrees, of the nine observers’ settings for each surface angle (60°, 80°, 90°, 100° & 120°) with the error bars representing the standard error of the mean. The dashed horizontal line indicates perfect, unbiased, performance. Table 6.2 shows the accuracy (mean setting error) and reliability (SD) of observers’ settings averaged across the five viewing distances expressed as a percentage of the veridical value. The data was analysed using a four way ANOVA with viewing distance, surface type (regular/irregular), depth cue condition (parallax/disparity) and surface angle as factors.
Figure 6.4. Plots observers' mean settings, in degrees, expressed as a function of viewing distance. Each graph depicts settings for a different surface angle: (a) 120°, (b) 100°, (c) 90°, (d) 80° and (e) 60°. The symbols represent the following conditions: binocular viewing of the regular pattern (●), binocular viewing of the irregular pattern (○), monocular motion viewing of the regular pattern (■) and monocular motion viewing of the irregular pattern (□). The dashed horizontal lines indicate veridical performance and the error bars represent SE.
Table 6.2. Group (n=9) performance indices are summarised separately for the regular and irregular reference textured surfaces, viewing condition and surface angle. The values are expressed as percentages of the veridical setting.

The results in figure 6.4 show that in general observers’ settings were highly accurate within both the static binocular and monocular motion conditions. There was no significant difference between the observers’ mean settings within these two depth cue conditions ($F_{4.32}=3.565, p>0.05$). Both figure 6.4 and the results in table 6.2 indicate there was a tendency for observers to become less accurate and sensitive as the angle of the surface increased though there was no significant main effect of surface angle ($F_{4.32}=1.567, p>0.05$). Figure 6.4 shows that in general observers exhibited little bias and were as accurate as their sensitivity limits permitted (denoted by their SD scores being considerably larger than their mean setting error). There was a highly significant interaction ($F_{4,32}=10.494, p<0.0001$) between the depth cue condition and surface angle. Table 6.2 indicates that was due to the much greater increase in the mean setting error within the motion conditions at oblique angles. Although the bias in the settings increased as the surface approached fronto-parallel ($180^\circ$) the depth of the surface (a minimum of 13.5 arc min' of equivalent disparity) at $120^\circ$ was well above the depth detection thresholds found earlier (see the nulling task - experiment 3.1). Hence, the results are unlikely to be caused by a difficulty in merely detecting depth from the motion signal.
Figure 6.5. The ‘estimated’ scaling distance as a function of the viewing distance (a, c, e, g & i), and the ratio of the ‘estimated’ scaling distance to the viewing distance (b, d, f, h & j). The symbols represent: binocularly viewed regular pattern (○), binocularly viewed irregular pattern (○), moving monocularly viewed regular pattern (■) and moving monocularly viewed irregular pattern (□). The dashed lines indicate perfect correspondence between estimated and actual viewing distance.
The difference between observers' settings for the two surface patterns was highly significant ($F_{1,8}=27.814, p<0.001$). This can be seen in table 6.2 which indicates that in most (9 out of 10 instances) instances the accuracy of observers' settings were greater for the regular surface pattern. This indicated that the texture cues only available within the regular surface pattern (e.g. perspective) aided the recovery and calibration of both binocular disparity and motion parallax information.

Although, there was no significant main effect of viewing distance ($F_{2,16}=0.674, p>0.05$) the results were once again converted into 'estimated scaling distances' (see chapter 3). This allowed a more thorough quantification of the degree of depth constancy exhibited by observers within the task. The converted results are plotted as (i) 'viewing distance' against 'estimated scaling distance' (figures 6.5a, c, e, g and i), and (ii) the ratios of the 'estimated scaling distance' to the 'viewing distance' (figures 6.5b, d, f, h and j).

The results in figures 6.5a and 6.5b demonstrate that observers' settings take changes in absolute distance into account. They exhibit almost perfect constancy (denoted by the flat lines in graphs 6.5b, d, f, h & j) for this task within the both of the tested viewing conditions. They further confirmed that observers' accuracy decreased as the size of the surface intersection angle increased. The settings for the smallest surface angle are almost perfectly veridical whilst those for larger angles were increasingly underestimated particularly on the basis of motion parallax (on average estimates of the 120° angled surfaces are ~90% of the veridical value).
The results from the current experiment indicate that observers can demonstrate highly accurate estimates of the angle of intersection of two planar surfaces on the basis of binocular disparity and motion parallax information in combination with additional depth cues. The differences found in performance between the regular and irregular surface patterns show that observers benefit from specific texture cues (e.g. perspective and outline information). However, observers' performance for the irregular surface was still highly accurate indicating an ability to benefit from other cues (e.g. texture cues, vertical disparities and/or accommodation) that are present within both surface types. Taken together the results further support the suggestion that the range of available cues determines the degree of constancy which observers can demonstrate in a shape judgement task (Cutting & Vishton, 1995; Durgin et al, 1995; Frisby et al, 1996). The identification of the additional sources of information that aided the observers recover the surface shape and the extent to which their accurate performance reflects the derivation of metric scene structure is considered in more detail within the general discussion below.

Observers' settings within the parallax condition were, in general, less accurate than for those in the disparity condition. However, there was no systematic difference in biases across the viewing distances in either condition, which would be expected on the basis of mis-estimated scaling parameters. This suggests that observers' less accurate performance within the parallax condition may be attributable to a greater difficulty in precisely registering the magnitude of retinal motion compared to
binocular disparities. Hogervorst and Eagle (1999) have previously argued that errors in the estimation of surface angle on the basis of parallax information can be well described by a Bayesian model that assumed the existence of noise in the measurement of early motion. Their model predicts that observers would demonstrate greatest accuracy for surfaces around a 90° angle. Although, this was not evident within the current experiment the accuracy of settings within all conditions was extremely high and it is likely, therefore, that performance was at a ceiling.

### 6.5 General Discussion

Both of the experiments within this chapter showed that the range of available cues can have a major effect upon the degree of accuracy observers demonstrated in depth assessment tasks (Cutting & Vishton, 1995; Durgin et al, 1995; Frisby et al, 1996). This may be due to the additional information allowing observers to more accurately recover scaling parameters to calibrate disparities and/or motion information thus deriving an unbiased Euclidean representation, or a more accurate approximation to an Euclidean representation. However, the finding of the earlier experiments (see chapters 3-4) which showed that observers switched strategies on the basis of the available information raises another possibility. In an environment with a large numbers of available cues observers may be able to access a range of strategies that enable them to derive a number of different representations that could allow them to perform tasks effectively without resorting to a full metric representation. A consideration of the nature of the information available within the two tasks suggests
that this latter possibility is the most parsimonious interpretation for the extremely accurate performance in the textured surface task.

Observers' settings within the textured surface task indicated they experienced no systematic bias in their perception of the shape in either static binocular or monocular motion viewing conditions. There are two possible explanations for this finding within both viewing conditions: (i) observers might have accurately recovered the appropriate scaling parameters using the additional cues to absolute distance or (ii) they may have combined slant information gained from monocular texture cues with information from disparities, or relative motion (either by averaging estimates from separate depth modules or by using multiple sources of information to constrain the possible interpretations). Additionally in the motion condition Ullman's (1979) suggestion for estimating shape from the rotational component of motion was also possible.

Theoretically there are only three available sources of distance information available within this experiment: (i) accommodation, (ii) vergence and (iii) vertical disparities. However, all of these, and other distance, cues were available in experiment 6.1 (the illuminated structured environment task) but observers could not complete that task without bias. Additionally, the utility of these cues would have remained constant across all surface angles and for both surface patterns. Therefore, a reliance on these information sources would predict no difference between the two types of surface. The difference between the two patterns was, however, highly significant. This would suggest that these absolute distance cues were not the information sources that allowed observers to perform so accurately for the regular surface pattern.
Another possibility is that slant estimates were made separately on the basis of the perspective information and the disparity/motion information and then averaged together, which is the type of interaction predicted by Landy et al.'s (1995) modified weak fusion model. This seems less likely as unbiased performance would require the independent unbiased recovery of slant for each cue. Experiment 6.1 demonstrated that even with a greater range of distance information available disparity based depth estimates were biased.

The difference in accuracy for the two surface patterns also suggested that observers completed this task by using an estimate of surface slant gained from monocular information, or via Ullman's suggestion in the motion condition. As noted in section 6.3, the most effective cues to slant available on a textured surface are linear perspective and compression (Gillam, 1968; Cumming et al., 1993). The compression cue remained the same for the irregular surface pattern (as the elements themselves were constant) but linear perspective was diminished as groups of surface elements no longer formed parallel lines. Therefore, it seems likely that observers used linear perspective information to estimate surface slant and constrain the interpretation of disparities and motion information. This argument is analogous to Richards (1985) suggestion for gaining surface slant from the acceleration component of the motion flow field. It also suggests that observers did not need to derive a metric representation to complete the task. Furthermore, the differences between the regular and irregular surfaces in the motion condition indicate that texture information was also employed in that condition rather than Ullman's proposal. The rotational component of the object motion should have been equally effective for the two surface types.

Frisby and Buckley (1993) have reported results that confirm surface perspective information effects the interpretation of disparity information. They examined the perception of textured stereoscopically viewed parabolic ridges when stereo and texture information conflicted and reported findings consistent with those for the structured
surface task. For physical surfaces, their experiments showed that although stereo
cues dominated in the interpretation of depth observers' judgements were still affected
by the slant specified by the texture cues.

In contrast to the textured surface task the results of the illuminated structured
environment task were consistent with the use of absolute distance information to
calibrate disparity information. Observers' depth settings demonstrated a bias that
varied systematically across the five viewing distances, unlike the textured surface
task where no effect of viewing distance was found. These results provide further
evidence for Foley's (1980) suggestion that failures in stereoscopic constancy are due
to a mis-estimation of viewing distance. Furthermore, they demonstrate that such
failures of constancy are not a merely product of isolated stereoscopic viewing
conditions (Durgin et al, 1995) nor can they be entirely explained by conflicting
distance or depth information specified by accommodation cues (Frisby et al, 1996).

In summary, the results of the two experiments are consistent with the increasing
evidence that the visual system cannot compute a full Euclidean representation of
spatial relationships (Gilinsky, 1951; Gogel, 1960; Wagner, 1985; Toye, 1986;
Norman et al, 1996). They show that the range of available information sources
affects the accuracy of observers' judgements of depth. However, they also show
that systematic mis-perceptions of depth do occur in environments in which a large
range of depth cues are present (the illuminated environment task). Furthermore,
they confirm the results of chapters 3 and 4 that the visual system has a range of
potential strategies available for completing tasks accurately, which is consistent with
the existence of a hierarchy of geometrical representations of visual space (Tittle et al, 1995).
CHAPTER 7

General Summary and Conclusions

The experiments in the thesis examined the derivation of the depth structure of a visual scene on the basis of binocular disparity and motion parallax information. They addressed a number of theoretical issues that are all broadly linked by the theme of the underlying representations that the visual system can derive on the basis of these information sources (alone and in combination). A primary question was the extent to which the visual system is capable of deriving a metric (Euclidean) representation of the depth structure of a visual scene. There was no evidence within any of the studies that disparity and parallax cues allowed the visual system to derive such a representation either in isolation from other cues (experiments 3.2, 4.2 and 5) or when supplemented with a range of pictorial information (experiment 6.1). For all of the tested tasks that required a metric representation observers' estimates of depth were systematically biased. A finding that supports previous studies which argued that departures from constancy for disparity and motion defined depth intervals are caused by a systematic mis-estimation of the requisite scaling parameters (e.g. Foley, 1980; Johnston, 1991; Norman et al, 1996).

The experiments still leave open the possibility that under certain circumstances the visual system may be capable of deriving a metric representation. It is, of course, logically impossible to prove that the visual system never derives a metric representation as any number of negative findings do not preclude the existence of a situation in which such a representation is utilised. However, they contribute to the mounting body of empirical evidence that the visual system does not compute such a

The experiments examined a number of factors that affected observers' performance in depth assessment tasks and concluded that it was determined, in part, by the demands of the perceptual task and the nature of the available depth information. In viewing environments where observers could not perform a metric task they were found to be able to perform depth nulling tasks (experiments 3.1 and 4.1) and depth matching tasks (experiments 3.3 and 4.3) with no systematic bias on the basis of both disparity and parallax. These results demonstrated that the visual system is capable of using a variety of strategies to complete depth assessment tasks depending on the demands of the task and the available information (Glennerster et al, 1996).

The experiments also indicated the visual system assessed the reliability of individual information sources when determining the optimal strategy for completing a task. A comparison of the direct viewing (chapter 3) and telepresence experiments (chapter 4) suggested that when the relationship between different sources of information changed the visual system could switch strategies. There was little interaction between disparity and parallax information in the direct viewing experiments but when observers performed a similar group of tasks viewed using a HMD the combined cue condition was consistently most accurate and reliable, which was attributed to the decrease in reliability of the disparity cue relative to the motion cue.

These results leave open to further investigation the range of different strategies that the visual system can utilise, and the circumstances under which they become available. The experiments within the thesis have shown that that one way of assessing which of these strategies are used is by examining the limits of the different
geometric representations that they produce. Tittle et al (1996) pointed out that any geometric transformation of an image leaves certain properties relating points in space unchanged and that geometries can be categorised into a hierarchy based upon the properties that they specify. Therefore, it is possible to determine if a particular representation is used by establishing that observers can accurately perform tasks that assess the different properties it specifies and by showing they cannot perform tasks that require properties that it does not specify.

The thesis specifically examined two potential strategies that the visual system might use to derive depth structure when scaling parameters cannot be reliably estimated: (i) Ullman's (1979) theorem to recover the 3-d shape up to an isotropic size scaling from the rotation of the stimuli configuration relative to the observer and (ii) Richards' (1985) suggestion that motion and disparity information can be combined to constrain the possible interpretations of depth structure in a visual scene. The size and depth task (experiment 5) was the only task in the thesis where the results were potentially consistent with Ullman's theorem, i.e. observers correctly estimated the size-depth ratio in the motion-only condition but not in the disparity-only condition. However, observers exhibited extremely good size constancy across the five viewing distances, which is not predicted by the theorem, and they did not appear to use Ullman's theorem in the shape task (experiment 3.2), which was very similar to the size and depth task. These results suggested that within the experimental viewing conditions observers could only recover the size-depth ratio when information in addition to the rotation of the stimulus was available\(^1\). One possibility is that the additional information allowed observers to more reliably assess the rotation of the stimulus.

\(^1\) A potential source of additional information available in the size and depth task, but not the shape task, was the constant rate and extent of movement of the base lights, which could provide an indication of the viewing distance.
This suggestion is supported by the lower SDs in all of the 'depth cue' conditions of the size and depth task compared to both the shape and matching tasks (experiment 3.2 and 3.3).

Although there was evidence in a number of the experiments for interactions between disparity and motion information (discussed in more detail below) none of the findings were consistent with Richards' suggestion. Therefore, it was concluded that the visual system could not use his proposal within these experimental viewing conditions to accurately recover shape. Interestingly, there was evidence within the patterned surface task (experiment 6.2) that the visual system did combine estimates of slant gained from linear perspective and/or surface outline information with both disparity and motion information. This implies that merely because a certain combination of cues can be used to derive a specific geometric representation it is not necessarily available from another combination of cues even when they, theoretically, specify qualitatively similar information about depth structure.

These findings open the possibility that accurate performance within an environment where a range of information sources are available is due to the visual system having access to a greater range of strategies rather than being able to compute a metric representation, or a more accurate approximation to a metric representation. In limited cue viewing environments (like those employed in chapters 3-5) the range of potential cue interactions are greatly constrained and performance can be attributed with little ambiguity to specific information sources, providing that sufficient care is taken to control extraneous variables. For this reason considerable emphasis in the current thesis was placed upon observers' inability to perform the tasks in static monocular conditions as theoretically few, if any, cues should have remained available. Therefore, this condition provides a way of establishing that residual
information sources have been minimised. For example, the results of pilot studies showed that observers could make considerable use of auditory cues in monocular conditions and that their accuracy decreased dramatically once they wore ear defenders. In addition, the extremely accurate performance of observers in all of the static monocular conditions in Frisby et al's (1996) experiments questioned the degree to which residual cues had been successfully controlled and hence their interpretation of the findings.

In experimental environments with multiple cues available it is far more difficult to identify the strategies that the visual system uses as the increased range of information sources introduces a far greater number of potential interactions. However, the results of the experiments in the thesis demonstrated that it is possible to draw conclusions about the nature of the representations employed by the visual system in a multiple-cue environment. This was achieved by examining the limits of the information sources across different viewing conditions and tasks. For example, it was argued that the extremely accurate performance within the textured surface task (experiment 6.2) was not due to a more accurate estimation of distance being used to derive a metric representation of scene structure. This was because the cues to distance (e.g. accommodation and vertical disparities) available within this task were insufficient to support the unbiased recovery of distance within the illuminated structured environment task (experiment 6.1) even when supplemented with additional information. However, although the increased accuracy within the latter task shows that the range of additional cues effected observers' percept of depth it was not possible to specifically quantify the impact of specific information sources.

This difficulty was largely due to the problem of independently controlling and manipulating a range of information sources in real multiple cue environments. This
problem can be partially obviated by the use of computer simulated stimuli, which allow a greater control over the magnitude of cues used to specify the stimuli. However, the traditional approach of viewing the stimuli on the monitor in the laboratory still leaves a large number of cues in the physical environment that are difficult to vary incrementally. Landy et al (1995) have argued that if the visual system assesses the reliability of cues (as the experiments in the thesis suggest) then it will tend to discount hugely discrepant cues, especially in multiple-cue environments where there are a large number of information sources signalling different depth. They argued that an appropriate way to assess the effect of cues is to quantify the impact of very small changes, which they termed perturbation analysis. One way to allow the experimenter such a fine degree of control over more of the information sources within the environment is the use of telepresence, virtual reality or augmented reality (which combine real and virtual information) systems viewed through a head mounted display. These systems potentially allow experimenters to more easily assess the effects of co-varying large numbers of cues rather than isolating a limited number of cues (as within the traditional approach psychophysics). A possible criticism of the use of these systems, shown by the experiments in chapter 4, is that the specific relationships between sources of information may differ from real world situations. However, those experiments demonstrated that it is still possible to derive insights into the mechanisms that the visual system uses for processing depth information. Furthermore, traditional limited cue experimental viewing conditions are also artificial and can produce results different from real world viewing.

The experiments in the thesis also examined the nature of the interactions that occur between motion parallax and binocular disparity information. The results of the combined cue condition in the size and depth task (experiment 5) were consistent with
the weighted averaging of depth information that is predicted by Landy et al's (1995) modified weak fusion model. Furthermore, Landy et al's model also predicts that the visual system would assess the reliability of different information sources to determine how to combine them, i.e. by weighted averaging (see earlier). Nonetheless, neither of these interactions are precluded by strong fusion models and a number of the interactions found cannot be accounted for by Landy et al's model in its current form. The combined cue condition of the telepresence depth nulling task (experiment 4.1) showed an improvement in sensitivity that was greater than could be expected on the basis of linear probability summation of independent depth estimates from the two sources of information (Bradshaw and Rogers, 1996). The results were, therefore, consistent with an early non-linear combination of the two information sources prior to estimation of depth, suggested by Bradshaw and Rogers (1996), rather than the modularization predicted by Landy et al's model. Additionally, the extremely accurate performance within the textured surface task (experiment 6.2) was consistent with the use of perspective information to estimate surface slant, and so constrain the interpretation of the parallax and disparity information.

Additionally, some suggestions for the potential reasons for the early combination of disparity and parallax information were generated by the results of the experiments in chapter 2. These examined observers' perception of depth for sinusoidal surfaces at a variety of orientation and corrugation frequencies. The results suggested the visual system has a differential sensitivity to specific components of the geometric transformations generated between retinal images in the disparity and motion domains, e.g. there appeared to be a greater sensitivity to compression information within the motion domain. This suggested that the non-linear interactions of motion and disparity within the early stages of visual processing proposed by Rogers and
Bradshaw (1996) could be due to the visual system combining the most sensitive primitive information source from each domain to derive depth structure, or even interactions and threshold summation between these primitive information sources. These suggestions remain open to further empirical investigation (see chapter 2 - for a further discussion).

A further aim of the current experiments was to assess whether departures from depth constancy on the basis of disparity and parallax information reported within many previous experiments (e.g. Johnson, 1991; Todd & Bressan, 1990; Tittle et al, 1995; Glennerster et al, 1996; Norman et al, 1996) could be explained by conflicting depth information signalled by accommodation in computer generated displays (Buckley & Frisby, 1993; Frisby et al, 1995a, 1996; Brennand et al, 1998). The experiments demonstrated that observers experienced systematic distortions of perceived depth when viewing physical stimuli (in which the depth information signalled by accommodation should be consistent with that specified by the other available cues). Deviations from constancy were shown to occur when observers viewed stimuli in a well-illuminated environment with a range of consistent distance and depth cues available (experiment 6.1). Therefore, conflicting depth information specified by accommodation does not provide a sufficient explanation within itself for failures of depth constancy in previous studies (e.g. Todd & Bressan, 1990; Johnson, 1991; Tittle et al, 1995; Glennerster et al, 1996; Norman et al, 1996).

Furthermore, it is possible to discount Buckley and Frisby's (1993) suggestion that deviations from constancy are due to the absence of accommodative blur within computer simulated stimuli. The combinations of stimulus depth and viewing distances used within the current experiments were outside the range which the human eye could distinguish such blur information (Mather, 1996, 1997). As a
consequence, the experiments demonstrated that both large departures from constancy (experiments 3.2, 4.2 and 5) and extremely accurate performance (experiment 6.2) occurred in situations in where the visual system could not rely upon relative blur.

However, the results of some of the experiments were consistent with variations in the absolute state of accommodation contributing to the magnitude of observers' biases within depth perception experiments (Brennand et al, 1998). The largest distortions in perceived depth occurred in the dark viewing environment (experiment 3.2) and telepresence experiments (experiment 4.2), where it is likely that the observers' accommodative state resulted in very large conflicts with other depth cues. In the real world dark viewing experiments isolated light sources would be a poor accommodative stimuli so it is likely that accommodation would have lay in its resting state (Liebowitz & Owens, 1975; Owens & Leibowitz, 1980). Brennand et al (1998) have demonstrated that there is a correlation between the distance specified by the resting state of accommodation and observers' estimates of depth in a dark viewing environment. However, it is not possible to quantify the degree to which this effected the observers' depth settings in these experiments for two reasons: (i) the observers' accommodative resting state could not be measured for technical reasons (i.e. equipment availability) and (ii) there is such a large inter-observer variation in the position of accommodative resting state that an average figure cannot reasonably be used for a sample size of four.

The quantification effects of accommodation upon estimates of depth on the basis of disparity and parallax information remains a major question for further research. However, a precise answer will almost certainly require the manipulation and measurement of the observers' accommodative state along with these other sources of information. In order to manipulate dynamically accommodative information in a
simulated display it would be necessary to alter the degree of blur across an image depending upon the viewing distance to the observers' point of fixation, the size of their pupil and their current accommodative state. The technical difficulties within such an undertaking at least partially explain why this question has not been addressed previously.

In summary, the experiments have shown that under a range of viewing conditions the visual system cannot derive a metric (Euclidean) representation of space on the basis of disparity and/or motion information. However, the visual system can derive and make use of a range of representations depending on the information available within the viewing environment and the demands of the task. These may enable it to complete tasks accurately without recourse to a metric representation. Furthermore, the visual system seems to be able to combine different information sources in a variety of ways and to switch strategies depending on its assessment of the reliability of specific information sources. When combining information from disparity and motion it appears to do so in the early stages of processing, which is most consistent with a strong fusion model of cue integration. This may be a result of differences in sensitivity to the various components of the geometric transformations generated between retinal images within the disparity and motion domains.
References

Angell R B, 1974 “The geometry of visibles” Nous 8 87-117


Blake A, Bulthoff H H, Sheinberg D, 1993 “Shape from texture ideal observers and human psychophysics” Vision Research 33 1723-37


Blank A A, 1960 “Curvature of binocular visual space. An experiment” Journal of the Optical Society of America 51 335-9


Bradshaw M F, Rogers B J, 1996 “The interaction of binocular disparity and motion parallax in the computation of depth” Vision Research 36 3457-68


Bradshaw M F, Parton A D, Eagle R, 1998 “The interaction of binocular disparity and motion parallax in determining perceived depth and perceived size” Perception 27 1317-33

Bradshaw M F, Rogers B J, 1999 “Sensitivity to horizontal and vertical corrugations defined by binocular disparity” Vision Research 39 3049-56

Bradshaw M F, Hibbard P B, Gillam B, 1999 “Perceptual latencies to discriminate stereoscopic slant and inclination” Manuscript submitted to Perception and Psychophysics


Bruno N, Cutting J E, 1988 “Minimodularity and the perception of layout” *Journal Of Experimental Psychology: General* 117 161-70

Buckley D, Frisby J P, 1993 “Integration of stereo, texture and outline cues in the shape perception of three-dimensional ridges” *Vision Research* 33 919-33


Cagenello R, Rogers B J, 1993 “Anisotropies in the perception of stereoscopic surfaces: the role of orientation disparity” *Vision Research* 33 2189-201


Cobo-Lewis A B, Yeh Y Y, 1994 “Selectivity of cyclopean masking for the spatial frequency of binocular disparity modulation” *Vision Research* 34 607-20


van Damme W, Brenner E, 1997 “The distance used for scaling disparities is the same as the one used for scaling retinal size” *Vision Research* **37** 757-64


Foley J M, 1980 “Binocular Distance Perception” Psychological Review 87 411-34


Gilinsky A S, 1951 “Perceived size and distance in visual space” Psychological Review 58 460-82

Gillam B, Finlay T, Flagg D, 1984 “Evidence for disparity change as the primary stimulus for stereoscopic processing” *Perception and Psychophysics* **36** 477-83


Gillam B, Ryan C, 1992 “Perspective, orientation disparity and stereoscopic slant perception” *Perception* **20** 441-8

Glennerster A, Rogers B J, Bradshaw M F, 1996 “Stereoscopic depth constancy depends on subject's task” *Vision Research* **36** 3441-56

Glennerster A, Rogers B J, Bradshaw M F, 1998 “Cues to viewing distance for stereoscopic depth constancy” *Perception* **26** 1331

Gogel W C, 1960 “The perception of a depth interval with binocular disparity cues” *Journal of Psychology* **50** 257-69


Graham M E, Rogers B J, 1982 "Simultaneous and successive contrast effects in the perception of depth from motion-parallax and stereoscopic information" *Perception* 11 247-62


Hadani I, 1991 “Corneal lens goggles and visual space-perception” *Applied Optics* 30 4136-47

Harris M, Freeman T, Hughes J, 1992 “Retinal speed gradients and the perception of surface slant” *Vision Research* 32 587-90


Hegarty M R, Hibbard P B, Bradshaw M F, De Bruyn B, Parton A D, 1997 “Anisotropic temporal integration in the perception of stereoscopic processing” *Perception* 26S 27

Hibbard P B, Langley K, 1998 “Plaid slant and inclination thresholds can be predicted from components” *Vision Research* 38 1073-84


Hogervorst M A, Bradshaw MF, Eagle R A, 1999 "Evidence for independent channels in the perception of motion-parallax defined surfaces" *Vision Research* in press.


Howard H J, 1919 “A test for the judgement of distance” *American Journal of Psychology* 2 656-75


Indow T, 1991 “A critical review of Luneburg’s model with regard to the global structure of visual space” Psychological Review 98 430-53

Johnston E B, 1991 “Systematic distortions of shape from stereopsis” Vision Research 31 1351-60


Johnston E B, Cumming B G, Landy M S, 1994 “Integration of stereo and motion cues” Vision Research 34 2259-75


Kappers A M L, Koenderink J J, Oudernaarden G, 1997 “Large scale differences between haptic and visual judgements of curvature” Perception 26 313-20

Kappers A M L, 1999 “Large systematic deviations in the perception of parallelity” Perception 28 1001-12


Lanca M, Bryant DJ, 1995 “Effect of orientation in haptic reproduction of line length” Perceptual and Motor Skills 80 1291-98

Liebowitz H W, Owens D A, 1975 “Anomalous myopias and the intermediate dark focus of accommodation” *Science* 189 646-8


Massaro D W, 1988 “Ambiguity in perception and experimentation” *Journal of Experimental Psychology: General* 117 417-21


Mather G, 1997 “The use of image blur as a depth cue” *Perception* 26 1147-58


Mon-Williams M, Wann J P, Rushton S, 1993 “Binocular vision in a virtual world - visual deficits following the wearing of a head-mounted display” Ophthalmic and Physiological Optics 13 387-91


Nakayama K, Shimojo S, 1992 “Experiencing and perceiving visual surfaces” Science 257 1357-63


Ogle K N, 1932 “Analytical treatment of the longitudinal horopter; Its measurement and application to related phenomena, especially the relative size and shape of ocular images” Journal of the Optical Society of America 22 665-728


Pepper R L, Smith D E, Cole R E, 1981 “Stereo TV improves operator performance under degraded visibility conditions” Optical Engineering 20 579-85


Predebon J, 1992 “The role of instructions and familiar size in absolute judgements of size and distance” Perception & Psychophysics 51 344-54
Predebon J, 1993 “The familiar size cue to distance and stereoscopic depth perception” *Perception* 22 985-95


Rogers B J, Graham M E, 1979 “Motion parallax as an independent cue for depth perception” *Perception* 8 125-34

Rogers B J, Graham M E, 1982 “Similarities between motion parallax and stereopsis in human depth perception” *Vision Research* 22 216-70

Rogers B J, Graham M E, 1983 “Anisotropies in the perception of three-dimensional surfaces” *Science* 221 1409-11


Rogers B J, Bradshaw M F, 1995a “Disparity scaling and the perception of fronto-parallel surfaces” *Perception* 24 155-80


Schumer R A, Ganz L, 1979 “Independent spatial channels for different extents of spatial pooling” *Vision Research* 19 1303-14


Sobell E C, Collett T S, 1991 “Does vertical disparity scale the perception of stereoscopic depth?” *Proceedings of the Royal Society of London B* 244 87-90


Teghtsoonian M, Teghtsoonian R, 1965 “Seen and felt length” *Psychonomic Science* 3 465-6


Todd J T, 1985 “Perception of structure from motion is projective correspondence of moving elements a necessary condition” *Journal Of Experimental Psychology: Human Perception and Performance* 11 689-710


Wagner M, 1985 “The metric of visual space” *Perception and Psychophysics* **38** 483-95

Wallach H, O’Connell, 1953 “The kinetic depth effect” *Journal of Experimental Psychology* **45** 205-17

Wallach H, Bacon J, 1976 “Two forms of retinal disparity” *Perception and Psychophysics* **19** 375-84


