

**The 'aperture problem' in complex
moving scenes**

By

David Kane

Acknowledgements

I would like to thank the Wellcome Trust for funding me throughout my PhD thesis. I would also like to thank my supervisors Steven Dakin and Peter Bex for their hard work and patience.

Abstract

The initial encoding of direction by mammals occurs in striate cortex by neurons with small receptive fields that are tuned to narrow bands of the spatiotemporal frequency spectrum. Individual neurons are unable to signal the global direction of 2D motion and are instead sensitive to the 1D component of motion perpendicular to a moving edge. To compute 2D velocity, it is necessary to integrate over a range of 1D velocity sensors. In this work I probe the ability of the visual system to compute 2D velocity from a range of stimulus classes, including naturally contoured scenes, natural scenes and a global-Gabor array. My research shows that the motion stream is highly sensitive to the distribution of local orientations present in a moving image, but is largely insensitive to their spatial second-order statistics. I present a computational model of two-dimensional motion processing that is able to derive precise estimates of 2D motion directly from complex natural scenes. The model produces errors when confronted with stimuli composed of anisotropic orientation configurations and is able to capture many of the biases and errors experienced by human observers. Finally, I argue that observers' misperceptions of 2D motion does not reflect a sub-optimal 2D motion strategy, but reflects a compromise between the competing requirements of defining motions in a spatially discrete manner across space, and the ability to accurately estimate 1D motions, on which the computation of 2D velocity must rely.

PREFACE

The ability of the primate brain to visually infer the temporal dynamics of the world has been the subject of intensive investigation. The study has revealed regions of the primate brain that are highly specialized for motion processing and the identification of a processing hierarchy in which progressively more complex motion signals are inferred from the visual environment. The primary factor through which the hierarchy can be described (and the focus of this thesis) is in terms of the incremental increase in the dimensionality of motion sensitivity in each ascending region of the motion stream; The earliest direction selective cells are found in area V1 of the primary visual cortex and are sensitive to the one-dimensional component of motion orthogonal to a surface orientation. These cells synapse with area MT where a proportion of cells are selective for the two-dimensional component of motion. In turn the cells in area MT synapse with area MSTd that is associated with full field motions such as optic flow.

The focus of this thesis is on the progression from one-dimensional to two-dimensional motion processing. The problem is commonly referred to as the '*aperture problem*' because when a moving straight edge is viewed through an aperture, the two-dimensional motion of the edge is ambiguous. In theory when two or more oriented surfaces are present (either locally, or distributed across space), the '*aperture problem*' may be resolved, however despite this theoretical possibility human observers are often systematically biased when asked to judge the two-dimensional motion of objects. Despite considerable research into brain regions V1 and MT associated with one- and two-dimensional motion processing there is no consensus on how the established

properties of motion sensitive cells may lead to observers' miss-perceptions of direction.

The work in this *thesis* attempts to link the established properties of motion sensitive cells (in area V1 and MT) and the psychophysical literature on the '*aperture problem*'. To do so I rely heavily on the motion energy model of 1D velocity processing to allow the construction of a model of two-dimensional motion processing that can (a) work upon unconstrained natural scenes and (b) incorporates biologically realistic constraints into the model.

The experimental chapters examine the influence of natural contour and natural orientation statistics on motion processing; statistics that are not present in the majority of stimuli used to probe the '*aperture problem*'. The approach stems from the argument that a system can only be truly measured or understood in terms of the natural environment it has evolved to cope with. In this regard I examine the influence of natural contour statistics in motion processing (experimental chapter 1) and I employ a reverse-correlation paradigm to examine the role of naturally occurring textures in motion processing (experimental chapter 2). A model of two-dimensional motion processing is then introduced that is able to work within a few degrees of accuracy upon natural scenes but performs very poorly on the artificial stimuli (e.g. plaids) used to study the aperture problem. A final experimental chapter is designed specifically to test how well the model is able to predict observers' errors in estimating the 2D direction of a global-Gabor array.

(1) INTRODUCTION	12
The motion stream	12
Overview: The 'Aperture Problem'	20
Equations relating 1D motion to 2D motion.....	22
How the distribution of 1D motions varies with 2D motion	23
Solving the 'aperture problem'	25
The curse of dimensionality.....	25
The Geometric Solution	27
Intersection of Constrains	31
Non-veridical solutions.....	33
Biological Vision.....	37
The Early Visual System	37
Linear-systems theory	39
Fourier analysis and wavelets	41
Ganglion cells	43
Lateral geniculate nucleus.....	44
Orientation and direction selectivity	46
Spatiotemporal frequency domain	48
Model of V1 motion energy	50
The 'aperture problem' - Psychophysics	58
Temporal aspect in the computation of 2D velocity	63
EXPERIMENTAL CHAPTERS.....	66
(2) EXPERIMENTAL CHAPTER NO.1	73

The influence of Natural Contours in motion processing	73
Contour Structure	74
Co-incidence of structure across spatial frequencies.....	78
Methods.....	83
Subjects.....	83
Apparatus.....	83
Stimuli	84
Procedure.....	85
Experiment 1: Dependence of direction discrimination on spatial-frequency structure.....	87
Experiment 2: The role of second-order statistics.....	89
Stimuli	90
Results	92
Experiment 3: Low SFs and the effect of scrambling carrier location	94
Methods	94
Results	94
Controls.....	97
Discussion	99
Model	102
Model Results.....	105
Implications for models of global motion processing.....	108
Experiment 4 Number, Density or Area	110
Subjects/Apparatus.....	112

Stimuli	112
Procedure.....	113
Results.....	115
Discussion	115
(3) EXPERIMENTAL CHAPTER NO.2	119
The aperture problem in natural scenes.....	119
Methods.....	121
Subjects.....	121
Apparatus.....	121
Stimuli	122
Procedure.....	123
Conditions.....	124
Observers' error.....	125
Bootstrapping	127
Results	127
Scene Statistics	132
Results	136
Results	137
Results	139
Discussion	144
Appendix.....	147
Scene statistics.....	147

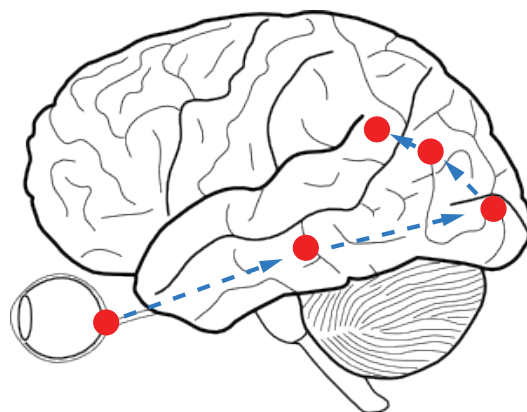
(4) GLOBAL MOTION MODEL	150
Methods.....	153
Local 1D motion sensors (V1).....	153
Global motion sensors (MT)	156
Model Details.....	157
Results	158
Artificial Stimuli.....	158
Natural Scenes	161
Discussion	163
Experimental Chapter No. 3	170
Testing the global motion model.....	170
Methods.....	173
Subjects/Apparatus.....	173
Stimuli (experiment 1).....	173
Procedure (experiment 1).....	173
<u>Data Analysis</u>	174
Results	175
Psychophysics.....	175
Equivalent noise.....	177
Methods.....	182
Subjects/Apparatus /Procedure/Stimuli	182
Ideal observer	182
Results	185

Equivalent Noise	185
Results	188
Testing the 2D model.....	188
Discussion	193
(5)	196
(5) CONCLUSION.....	197
Summary	197
Biological plausibility	204
Model limitations.....	207
Predictable errors	208
Biological realism	215
References.....	224

(1) Introduction

The motion stream

Visual motion has been studied extensively in the field of visual neuroscience. This is in part due to the identification of brain regions highly specialized for the detection of motion and the identification of a hierarchical motion stream in which increasingly complex motion signals are processed at each ascending stage of the motion stream (illustrated in Figure 1; Bartels, Zeki, & Logothetis, 2008; Maunsell & van Essen, 1983a; Movshon, Adelson, Gizzi, & Newsome, 1986a; Zeki, et al., 1991).



Retina → LGN → V1 → MT → MST

Figure 1-1. The primate motion processing stream. The visual signal from the retina is passed via the lateral geniculate nucleus to the primate visual cortex. In area V1 the first direction and orientation selective cells are found. Direction selective cells in area V1 are sensitive to the one-dimensional component of motion orthogonal to an oriented surface. These cells are known to synapse with cells in the medial temporal (MT) area that are sensitive to both the *speed* and *direction* of two-dimensional motion. In the nearby medial superior temporal area cells are found which are sensitive to full field motion such as those generated by an organism passing through a three-dimensional environment.

Primate visual sensitivity begins with the *rod* and *cone* cells found in the retina that are sensitive to the *wavelength* and *intensity* of light. The visual signal is then relayed via the lateral geniculate nucleus (LGN) to the primary visual

cortex located in the posterior region of the brain. Unlike some mammals (e.g. rabbits; H. B. Barlow & Hill, 1963) which have direction selective (DS) cells in the retina, direction selectivity is not noted in the primate visual stream until area V1 of the primary visual cortex. However the direction selective (DS) cells found in area V1 are not sensitive to the two-dimensional velocity of an object across the retina but are instead sensitive to the one-dimensional component of motion perpendicular to a surface orientation (Hubel & Wiesel, 1962). Given that the retina records a spatially two-dimensional representation of the world across time, the rationale behind such a mechanism is unclear, however it can be argued that the approach reflects a rational approach because two-dimensional motion is not always resolvable within a narrow region of time and space. This problem (detailed in depth in the next section) is referred to as the 'aperture problem' and concerns the fact that the two-dimensional motion stemming from a straight edge is locally ambiguous. Theoretically, the motion signal from two or more straight edges is sufficient to disambiguate the two-dimensional motion of an object, however in unconstrained natural environments the region of space and time needed to achieve disambiguation is an unknown variable. Accordingly any system (artificial or biological) that attempts to solve this problem in unconstrained environments will have to dynamically alter the region and time period of integration.

Many authors claim the 'aperture problem' is solved in the medial temporal (MT) region of the primate brain. The homologue of primate area MT in humans is sometimes referred to as area V5. In area MT an unusually high proportion of cells are sensitive to motion (~90%). Cells in area MT are reciprocally connected to area V1 (Maunsell & van Essen, 1983a), have

receptive fields 10 times the area of V1 cells and are known to receive projections from directionally selective cells in area V1 (Movshon & Newsome, 1996). Functionally this structure makes area MT ideally suited to resolving the ambiguity inherited from V1 DS cells. The key evidence in support of this notion comes from the observation that around a third of cells in area MT are responsive to the two-dimensional direction of motion even if the 1D velocity signals are oblique to the two-dimensional vector (Movshon, Adelson, Gizzi, & Newsome, 1985; Rodman & Albright, 1989). Speed tuning is also refined as we move from area V1 to area MT. The temporal frequency tuning of V1 DS cells is either low-pass or high-pass (Foster, Gaska, Nagler, & Pollen, 1985) and the V1 DS cells are not speed tuned because their spatial and temporal frequency tuning functions are separable (Foster, et al., 1985). In contrast, the majority of MT cells have inseparable spatial and temporal frequency tuning (Perrone & Thiele, 2001) and are thus tuned to stimulus speed. MT cells are tuned to all directions and a broad distribution of speeds (DeAngelis & Uka, 2003), but the distribution of speed tuning is heterogeneous with the majority of cells tuned to high angular speeds ($\sim 32^\circ/s$) (Cheng, Hasegawa, Saleem, & Tanaka, 1994). The speed tuning properties of MT cells are heterogeneous and are either low-pass, band-pass, broadband or high-pass (Lagae, Raiguel, & Orban, 1993; Mikami, Newsome, & Wurtz, 1986)

The notion of a hierarchical motion stream receives further support by the identification of reciprocal axons terminals from area V1 to MT and from MT to the medial superior temporal area (MST) (Maunsell & van Essen, 1983a). Within the dorsal regions of MST (MSTd) cells are selective to a range of full field motions such as rotation, radiation and translation, or to combinations these motions (Graziano, Andersen, & Snowden, 1994). With the exception of full

translations, such motion patterns require more than two-dimensions to describe them, because they describe a change in velocity over space. More recently it has been shown that nearly all cells in area MSTd are 'pattern' selective (Khawaja, Tsui, & Pack, 2009).

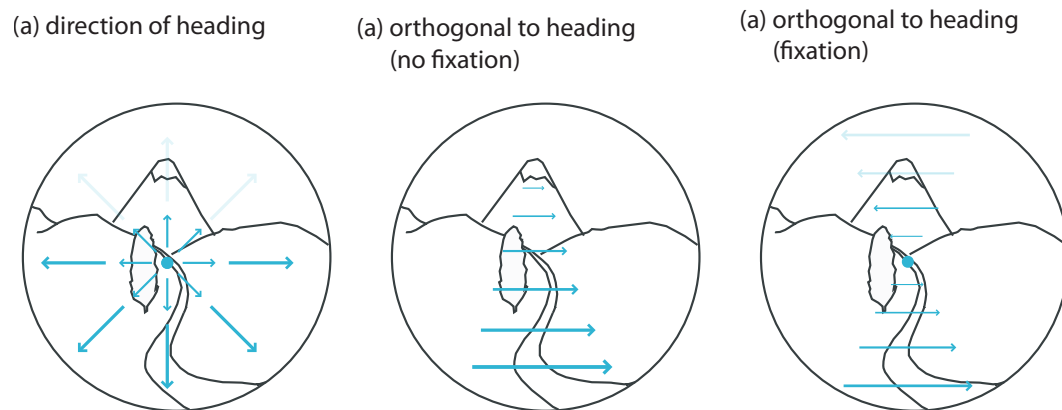


Figure 1-2. Full field motions generated by specific movement and fixation parameters. (a) The pattern of 2D motion (Vector-flow-field) generated by an organisms moving towards the point of fixation. (b) The Vector-flow-field generated by an organism moving orthogonal to the direction of unfocused gaze (c) The Vector-flow-field generated by an organism moving orthogonal to the direction of focused gaze. Note, the sign of motion switches at the point of fixation.

The result of the above *anatomical* and *functional* studies have led to the notion that motion processing occurs in a hierarchical and feed-forward manner with ascending regions becoming selective for more complex representations of visual motion. A number of additional factors contribute to this idea, including the development of phase-insensitivity in area V1 (Hubel & Wiesel, 1968), the development of disparity (Uka & DeAngelis, 2003) and speed (Perrone & Thiele, 2001; Priebe, Lisberger, & Movshon, 2006) selectivity in area MT, and the position invariance of MST cells (Duffy & Wurtz, 1991). However the primary dimension in which motion selectivity develops (and the primary topic of this *thesis*) is in the increasing dimensionality of the motion sensitivity in ascending regions of the primate brain; specifically the *thesis* will

examine how motion signals are combined across space to improve estimates of 2D motion. In this regard I will describe the component of motion orthogonal to a surface orientation as 1D motion.

One may ask why the motion stream utilizes this hierarchical structure; why not infer the necessary motion defined information directly from the retina image of the world? A clue may come from a search of the literature on the computation of optic-flow and the observation that a large number of optic flow models rely on a representation of motion known as the *vector-flow-field* (see Perrone, 2001 for a review). The *vector-flow-field* is a representation of two-dimensional motion at each point in the visual field. In other words a large number of models that infer motion-defined properties about the world assume the 'aperture problem' has been solved. This presents a problem; although area MT is commonly considered to 'solve' the 'aperture problem' and the output from MT can be correlated with behavior (Britten, Shadlen, Newsome, & Movshon, 1992; Newsome, Britten, & Movshon, 1989; Salzman & Newsome, 1994) there is still no consensus on how the known properties of V1 and MT cells may account for human observers' misperceptions of two-dimensional motion. Specifically, human observers are often systematically biased towards the direction of 1D motion, even when there is sufficient information to correctly compute the 2D velocity (K. Amano, M. Edwards, D. R. Badcock, & S. y. Nishida, 2009; Bowns, 1996, 2002; Burke & Wenderoth, 1993; Loffler & Orbach, 2001; Mingolla, Todd, & Norman, 1992; Wilson & Kim, 1994; Yo & Wilson, 1992). Despite a number of proposed theories there is no commonly accepted model that can account for observers' solve the 'aperture problem'. (Adelson & Movshon, 1982; Johnston, McOwen, &

Buxton, 1992; Nowlan & Sejnowski, 1995; Perrone, 2004; Simoncelli & Heeger, 1998; Weiss, Simoncelli, & Adelson, 2002; Wilson, Ferrera, & Yo, 1992).

The work in this *thesis* attempts to reconcile the known properties of area V1 and MT with the psychophysical literature on two-dimensional motion processing. To do so I first set out the 'aperture problem' from a *theoretical* perspective. The point is made that our current theories about how the 'aperture problem' may be solved do not lead to a ready explanation of why observers may sometimes misperceive two-dimensional motion. This has led some authors to propose that two-dimensional motion is computed via a *non-veridical* mechanism (e.g. Yo & Wilson, 1992), whilst others have proposed a number of overlapping motion systems selective for a number of motion defined attributes such as zero-crossings, minima or maxima (Bowns, 1996), contrast defined motion and feature tracking mechanisms (Stoner & Albright, 1996).

In contrast, I argue that the misperceptions of human observers results from a feed forward, two-stage model of two-dimensional motion processing that is optimal in a theoretical sense, but is constrained by the necessity to derive motion estimates from the natural environment. To make this argument (after detailing the theoretical nature of the 'aperture problem'), I introduce the major theories that dominate our thinking about primate vision. Particular emphasis is given to the concept of the wavelet filter (Daugman, 1980;

Gabor, 1946) and the competing constraints of extracting visually defined information that is both localized in time and space and accurately extracts the desired stimulus features (Graham, 1989). This structure is useful because it allows us to incorporate the constraints imposed upon the visual system by the initial detection of motion into our understanding of the psychophysical literature on the 'aperture problem'.

Overview: The ‘Aperture Problem’.

The ‘aperture problem’ refers to the inherent ambiguity of a motion signal arising from a stimulus containing one orientation (i.e. a straight edge). The problem is highlighted in Figure 1-3 which depicts a bar rigidly translating in the rightward direction. The endpoints of the bar are occluded such that only one surface orientation is exposed. The motion stemming from this orientation is inherently ambiguous and is potentially consistent with an infinite range of global velocities as depicted by the graph in Figure 1-3(b). In other words an object moving in any of the velocities denoted in (b) may give rise to the physical stimulus shown in (a).

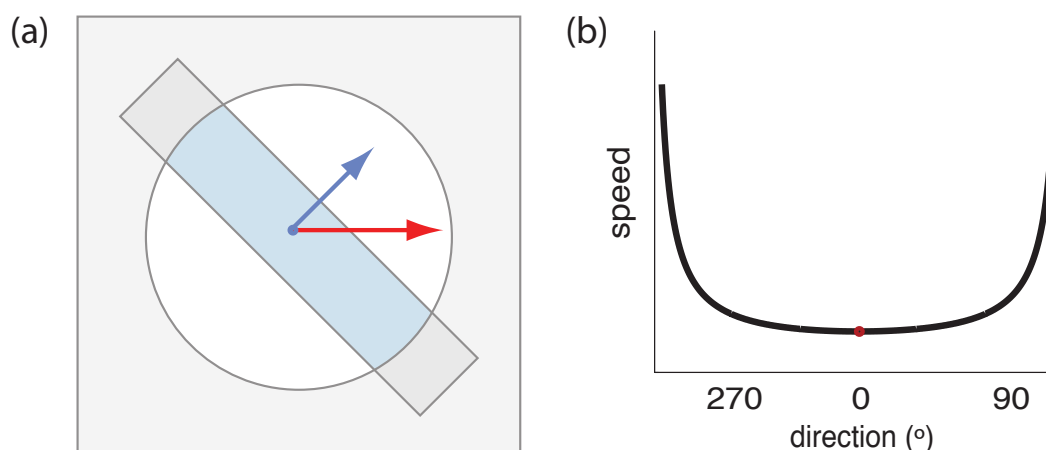


Figure 1-3 (a) a bar rigidly translating rightward is occluded and the only visible portion is a straight edge. The red arrow denotes the 2D direction and the blue denotes the component of motion perpendicular to the surface orientation. (a) The range of possible 2D motions with which the motion signal from the edge is consistent. Observers perceive motion in the direction orthogonal to the edge orientation (also the slowest possible 2D motion consistent with that 1D velocity).

In the absence of any disambiguating cues human observers tend to perceive locally ambiguous motion in the direction orthogonal to the line's orientation (Wallach, 1935), a percept consistent with the established properties of V1 direction selective cells (Hubel & Wiesel, 1968). Although this

2D percept may be incorrect, computing the 1D component of motion, normal to an edge orientation serves a clear and useful purpose as it allows one to lawfully relate the distribution of 1D velocities (speeds and directions) to the 2D motion of an object and the orientation structure of that object.

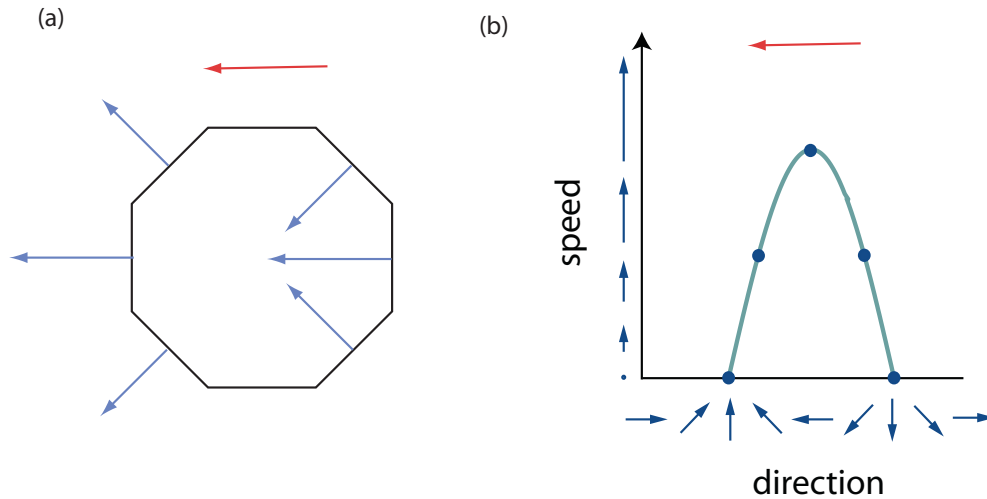


Figure 1-4 (a) the 1D velocity perpendicular to each surface orientation (b) the pattern of 1D velocities as a function of the speed and direction.

To illustrate this point in Figure 1-4 I replace the bar with an octagon rigidly translating in the leftward direction. In (a) the blue arrows denote the component of motion *perpendicular* to each surface orientation. I will term the component of motion orthogonal to a surface orientation as 1D motion in the rest of this *thesis*. In (b) I plot the *speed* and *direction* of each 1D velocity (blue dots). If the dots are joined, it can be seen that 1D speed is related to 1D *direction* in a cosine manner. The equations defining this relationship are shown in the next section.

Equations relating 1D motion to 2D motion

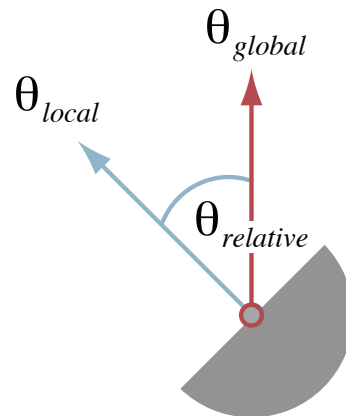


Figure 1-5 If 1D velocity is defined as being in the direction orthogonal to a surface orientation (blue arrow), then the speed (magnitude) of each 1D velocity is lawfully related to the angle between the 1D velocity and the 2D motion (see next Figure).

The 1D component of motion orthogonal to an edge orientation is denoted by the blue arrow (Figure 1-5). The speed of 1D motion, measured normal to an edge orientation, varies in a sinusoidal manner with the angular separation between the edge orientation and the 2D direction. As a result, the range of 1D velocities a 2D velocity may illicit, lie upon a sine wave defined by Equation 1.1, where ϑ denotes the orientation of an edge, ϕ_{1D} the speed of 1D motion and θ_{2D}, ϕ_{2D} the speed and direction of 2D motion.

$$\phi_{1D} = \sin(\theta_{2D} - \vartheta)\phi_{2D}$$

Equation 1.1

According to Equation 1.1 orientations 180° apart will generate speeds of identical magnitude but opposites speed (i.e. the same 1D velocity). To constrain the description of 1D motions in a stimulus to positive speeds, I first calculate the angular separation between each orientation and the 2D

direction, across the half circle (Equation 1.2). This calculation produces a number between $\pm 90^\circ$ which I will term *relative orientation*. This relative-orientation term is used throughout the thesis when referring to the orientations of a moving object. The speed and direction of 1D motion can then be computed from the *relative orientation* term through Equation 1.3.

$$\theta_{relative} = \tan^{-1}\left(\frac{\sin(\theta_{2D} - \vartheta)}{\cos(\theta_{2D} - \vartheta)}\right)$$

Equation 1.2

$$\begin{aligned}\theta_{1D} &= \theta_{2D} + \theta_{relative} \\ \phi_{1D} &= \sin(\theta_{relative})\phi_{2D}\end{aligned}$$

Equation 1.3

How the distribution of 1D motions varies with 2D motion

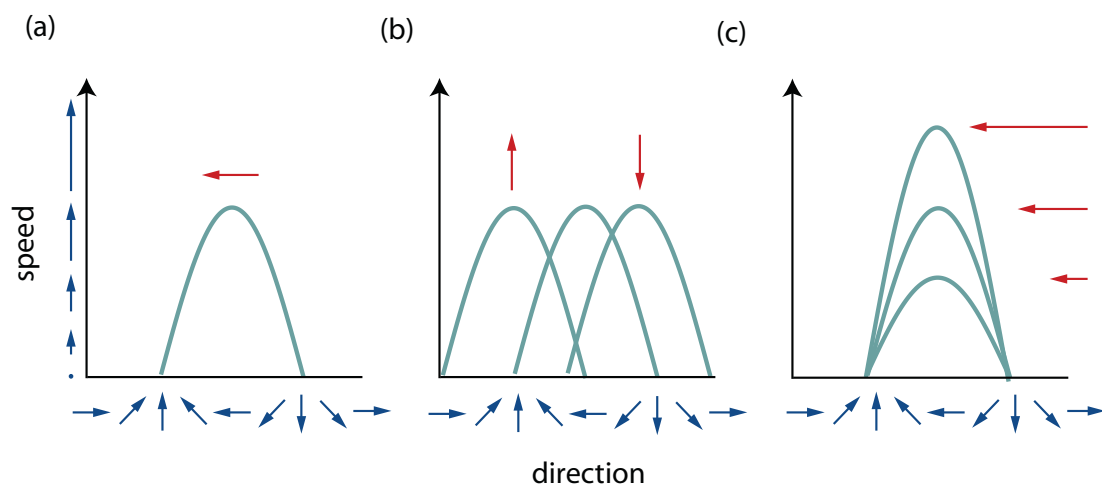


Figure 1-6 Speed vs. direction plot of the relationship between 1D and 2D velocities (a) red arrow denotes the 2D velocity of an object rigidly translating in the leftward direction, the blue line denotes all the possible 1D motions that are consistent with this global motion. (b) The distribution of 1D motions for upward and downward motion (c) The distribution of 1D motions for faster and slower motions.

Figure 1-6 (a) demonstrates how the pattern of 1D velocities varies with an object's 2D velocity. In (b) the 2D direction of motion is shifted clockwise or anticlockwise and respectively shifts the cosine left or right. In (c) the speed of 2D motion is either increased or decreased, stretching or squashing the cosine.

To probe the 'aperture problem' many studies have used just two orientations, as this is the minimum needed to constrain an estimate of 2D velocity. Interestingly observers often miss-perceive the direction of motion across a range of stimulus classes (Bowns, 1996; Bowns & Alais, 2006; Burke & Wenderoth, 1993; Heeley & Buchanan-Smith, 1992; Mingolla, et al., 1992; Wilson, et al., 1992; Wilson & Kim, 1994; Yo & Wilson, 1992), leading some authors to suggest the problem is solved via a sub-optimal processing scheme (Wilson, et al., 1992), accordingly the next section details a number of optimal and non-optimal models of two-dimensional motion processing.

Solving the ‘aperture problem’

In this section I briefly outline potential *veridical* and *non-veridical* solutions to the ‘aperture problem’. For the purposes of this section it is sufficient to say that the psychophysical percept of motion is not always veridical and can be systematically biased towards the direction of the 1D motions in a stimulus (Mingolla, et al., 1992; Rubin & Hochstein, 1993; Weiss, et al., 2002; Wilson, et al., 1992; Wilson & Kim, 1994; Yo & Wilson, 1992) and there is still much debate regarding the mechanisms of two-dimensional motion detection.

Before moving on to potential solutions to the ‘aperture problem’ it is worth noting the difference between the Velocity-space representation of motion and the Speed-Direction representation, that latter of which I have already used in [Figure 1-4](#) and [Figure 1-6](#). The Velocity-space is a Cartesian representation of motion where a 2D motion may be described by its component in the **x** and **y** dimension. The advantage of using a Velocity-space representation is that a straight line describes the range of possible 1D velocities that an individual 1D motion is consistent with, whilst a circle that passes through the origin describes the range of 1D motions that is consistent with a specific 2D velocity.

The curse of dimensionality

The distribution of 1D velocities is determined not only by the *speed* and *direction* of 2D object motion but also by the orientation structure of the stimulus. This presents a problem for a standard template-matching model of

2D motion because an infinite number of 1D velocity distributions are consistent with any 2D velocity.

Instead I introduce two veridical solutions that circumvent this issue, the first is an adaption of the *geometric solution* commonly used by computer scientists searching for circles or ellipses in digital imagery or other unconstrained data (Gander, Golub, & Strebler, 1994). The algorithm is designed to work upon noisy data and thus provides a good source for an ideal-observer-model used in the final section of this thesis. The second, more commonly cited solution is the *Intersection of Constraints* (IOC) solution. This mechanism was first described by Adelson and Movshon (1982) and takes advantage of the fact that each individual 1D velocity is consistent with an infinite range of possible global velocities, but constrained upon a line in *Velocity-space* known as the *constraint line*. Accordingly a veridical solution to the 'aperture problem' can be achieved by looking for the point of intersection between two or more constraint lines.

The IOC solution is commonly referred to in the literature as a system that correctly specifies the conjoint two-dimensional velocity consistent with a range of 1D velocities. As the solution to the 'aperture problem' is still debated the work in this *thesis* will move away from such theory-laden language. The use of IOC is pervasive throughout the literature (perhaps for historical reasons), for instance the review paper by Bradley & Goyal (2008) described the global motion model of Simoncelli & Heeger (1998) as

performing an IOC-like computation, Bradley likely means that the solution is capable of solving the 'aperture problem' rather than with reference to the specific computations made for which there is little comparison. Arguably the geometric solution I present below is closer to the biologically inspired model of Simoncelli & Heeger (1998) than the IOC model.

The Geometric Solution

The *geometric solution* is an iterative procedure that searches for the best fitting circle or ellipse among a number of data points. The algorithm is typically used in *Cartesian* space for problems in which the centre of the circle and the radius of the circle is unknown. As the pattern of 1D velocities generated by a rigidly translating two-dimensional object can be described by a circle in the Cartesian representation of motion (Velocity space), the procedure can be readily adapted to the 'aperture problem'.

The *geometric solution* can be contrasted against the *algebraic solution* that attempts to minimize the distance between all data point and an equation for a circle (i.e. all points on a circle). The *algebraic solution* is inappropriate when only a small arc of the circle is present (Gander, et al., 1994) because the approach attempts to fit all points on a circle to the available data. The algebraic solution tends to place the circle centre just inside of the arc and underestimate the circle *radius*, as highlighted in [Figure 1-7](#).

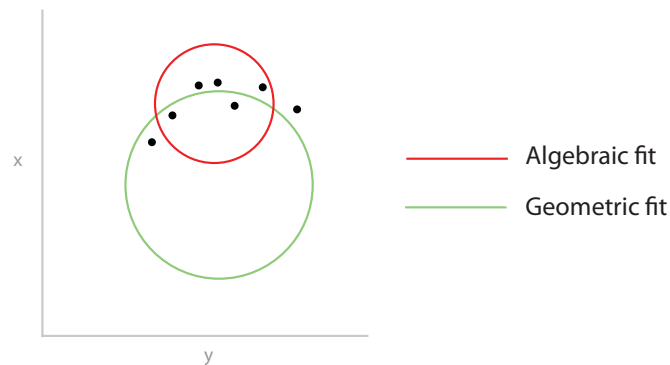


Figure 1-7 Algebraic versus geometric solutions. An algebraic fit (red line) and a geometric fit (green line). Note how the algebraic fit tries to minimize the distance between the data points (black dots) and the full circle; in contrast the geometric fit minimizes the distance between a small region of the circle and the data points.

In contrast, the class of solutions known as the *geometric solution* attempts to estimate which point on a circle the data point may have arisen. In terms of two-dimensional motion processing this is like estimating which orientation led to a 1D velocity. If the noise source is defined in Cartesian space and is normally distributed around the mean, then an optimal solution is to draw a line between each data point and the centre of a circle as illustrated in Figure 1-8. The line bisects the circle at two points; by taking the smaller of the two *Cartesian* distances and taking the *root-mean-square* error I can arrive at an optimal error signal.

Geometric fit

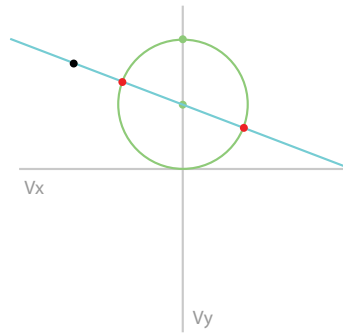


Figure 1-8 The geometric solution. The green circle depicts the range of 1D velocities consistent with a particular 2D velocity and the black dot indicates 1D velocity. The Geometric solution works by estimating the points on a given circle that may have led to a particular 1D velocity and then minimizing the root-mean-square distance between all data points and a circle. In this example the model assumes that the noise is equal in \mathbf{x} and \mathbf{y} . Under such conditions the most likely point on the circle to have led to the 1D velocity can be found by drawing a line between the data point and the centre of the circle. The line will bisect the circle at two points and the distance between the 1D velocity and the closest point of intersection is taken as the error.

This geometric solution is typically designed to fit a circle to a distribution of noisy data points but can be modified to solve the 'aperture problem' in Velocity-space when the 1D velocities are defined in velocity-space.

We know that a global motion is constrained to pass through the *origin* and the centre of each circle (defining the range of 1D velocities consistent with a 2D velocity) $c_x c_y$ in Velocity space can be defined as follows:

Equation 1.4

$$c_x = \sin(\theta_{2D}) \frac{\phi_{2D}}{2}$$

Equation 1.5

$$c_y = \cos(\theta_{2D}) \frac{\phi_{2D}}{2}$$

If we assume that the noise is equal in \mathbf{x} and \mathbf{y} , then our best guess as to what point on the circle led to a given data point can be achieved by drawing a line that bisects the data point and the centre of the circle. As the line will bisect the circle twice, we chose the point of intersection that is closest to the data point as the best estimate. To create a solution, I modify equations for calculating the point of intersection between a line and a circle from (Weisstein, 2009).

For each data point $p_x p_y$ the points of intersection $i_x i_y$ can be defined as

$$\text{Equation 1.6} \quad i_x = \frac{-Dd_x \pm \text{sign}^*(d_y)d_x\sqrt{r^2d_r^2 - D^2}}{d^2}$$

$$\text{Equation 1.7} \quad i_y = \frac{-Dd_x \pm |d_y|d_x\sqrt{r^2d_r^2 - D^2}}{d^2}$$

Where,

$$\text{Equation 1.8} \quad \text{sgn}^*(x) \equiv \begin{cases} -1 & \text{for } x < 0 \\ 1 & \text{otherwise} \end{cases}$$

And $d_x d_y$ is the distance in \mathbf{x} and \mathbf{y} between each data point and the closest point on a circle.

$$\text{Equation 1.9} \quad d_x = c_x - p_x$$

$$\text{Equation 1.10} \quad d_y = c_y - p_y$$

$$d_r = \sqrt{d_x^2 + d_y^2}$$

$$\text{Equation 1.11}$$

$$D = c_x p_y - c_y p_x$$

Equation 1.12

Intersection of Constrains

The *Intersection of Constraints* (IOC; Adelson & Movshon, 1982) solution is a rapid means of estimating 2D velocity from two or more discrete 1D velocities. Unlike the geometric solution it does not require an exhaustive search of potential 2D motions across the *speed* and *direction* dimensions, but instead efficiently achieves an arithmetic estimate of 2D velocity.

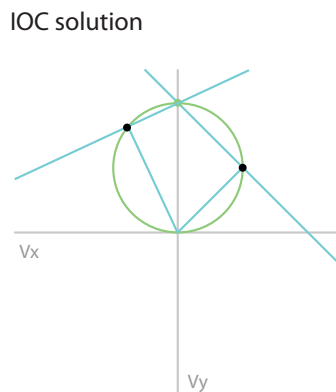


Figure 1-9 The Intersection of Constraints solution. The Intersection of Constraints solution calculates the point of intersection (green dot) between lines denoting the possible global velocities (blue line) consistent with each 1D velocity (black dot)

The IOC solution takes advantage of the fact that the range of 2D velocities that a 1D velocity is consistent with, lies upon a line in *Velocity-space*, known as a *constraint line*. Thus by combining two or more such *constraint lines* and looking for the points of intersection a unique veridical solution can be found (assuming no noise on the estimation of 1D velocities).

The IOC can be computed as follows;

First take the angle orthogonal to the angle between each 1D velocity and the origin.

$$m = \tan(\theta_{local} + 90)$$

Equation 1.13

Then calculate the point the each line bisects the **y**-axis.

$$c = y - mx$$

Equation 1.14

Once you have **c** and **m**, for two or more constraint lines, the point of intersect can be calculated through Equation 1.15 & Equation 1.16.

$$i_x = \frac{c_1 - c_2}{m_2 - m_1}$$

Equation 1.15

$$i_y = m_2 i_x + c$$

Equation 1.16

Both solutions discussed so far are defined in Velocity space. This is a problem if we want to incorporate biologically realistic constraints such as directional or temporal bandwidths (Dakin, Mareschal, & Bex, 2005b; Matthews & Qian, 1999) which are described in the polar dimensions.

Non-veridical solutions

The perception of coherently moving plaids or drifting bar like stimuli is often not in the direction of predicted by an IOC/veridical combination of 1D velocities, but closer to the mean of the individual components (Bowns, 1996; Mingolla, et al., 1992; Rubin & Hochstein, 1993; Yo & Wilson, 1992). This finding has lead authors to suggest that global motion is achieved through a *Vector-Sum* (VS) or a *Vector-Average* (VA) computation. The two solutions are computed by first deconstructing the stimulus into their respective \mathbf{x} and \mathbf{y} components (Equation 1.12 & 1.14), the components are then either *summed* (VS; Equation 1.14) or *averaged* (VA; Equation 1.14) and the resulting vector is the estimate of 2D motion. Both procedures have the advantage that they can be computed efficiently in one step. The *Vector-Sum* and *Vector-Average* represent a plausible strategy because the *orthogonal/fastest* component of motion, closest to the object velocity will contribute most to the final estimate of motion and the fastest 1D velocity is the closest to the true 2D velocity. The *Vector-Sum* solution will predict increasingly fast velocities with an increasing number of 1D signals and does not represent a plausible mechanism for the decoding of speed (see, Figure 1-10) Accordingly the *Vector-Average* solution has been introduced because this mechanism does not produce increasing speed estimates with increasing number of inputs, instead the *Vector-Average* solution tends to underestimate the global speed.

$$x = \sum_{i=1:n} \cos(\theta_{local(i)} \phi_{local(i)}) \quad \bar{x} = \frac{x}{n}$$

Equation 1.17

$$y = \sum_{i=1:n} \sin(\theta_{local(i)} \phi_{local(i)}) \quad \bar{y} = \frac{y}{n}$$

Equation 1.18

$$VS = \arctan\left(\frac{x}{y}\right)$$

Equation 1.19

$$VA = \arctan\left(\frac{\bar{x}}{\bar{y}}\right)$$

Equation 1.20

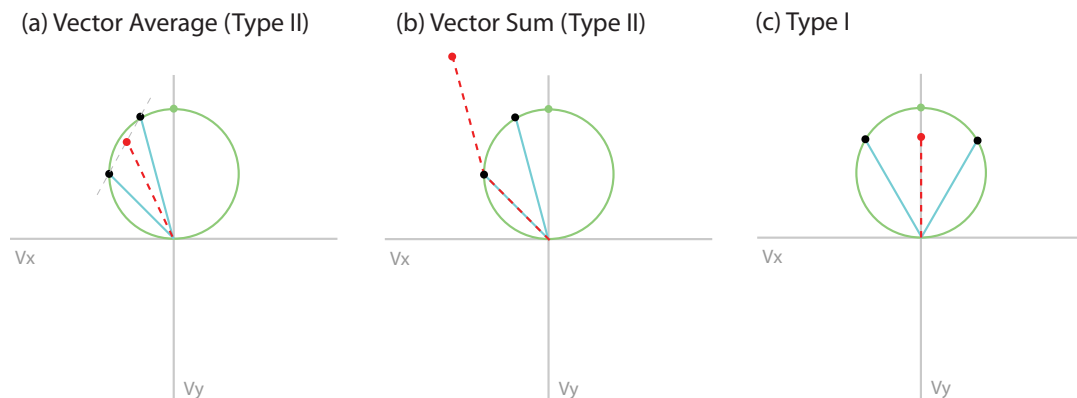


Figure 1-10 Type I, Type II stimulus classes; (a) the Vector Average solution applied to a Type II stimulus, the solution is biased away from the veridical direction denoted by the green dot towards the 1D directions in the stimulus. As the fastest of the two 1D velocities has the greatest magnitude the final estimate is closer to the fastest 1D velocity (b) the Vector Sum solution, note how the addition of the two 1D velocity signals leads to an unrealistically fast motion estimate but the direction estimate is the same as in (a). (c) a Type I stimuli, the red line denotes the Vector Average solution which correctly estimates the speed but underestimates the speed.

In order to distinguish between veridical methods such as the *geometric-solution* and the *IOC solution*, potential stimuli are classified in terms of their

orientation structure. Stimuli whose 1D directions lie to one side of the 2D direction are classed as Type I (Figure 1-10c) because both averaging class of solutions and the geometric and IOC solutions will generate veridical estimates. In contrast when the 1D velocity signals are defined to be one side of the veridical direction (Figure 1-10, a&b) different predictions are generated and stimulus classes that can distinguish between an averaging and a veridical solution to the 'aperture problem' are known as Type II.

Biological Vision

An object moving relative to the eye will generate a changing pattern of light upon the retina, however it is not until we reach the primary visual cortex that cells become selective for the *direction* and *speed* of motion (Hubel & Wiesel, 1968). These cells known as *directionally-selective* (DS) cells are not only sensitive to the *direction* and *speed* of the stimulus but are selective for a number of other stimulus attributes such as spatial frequency, (K. K. De Valois, De Valois, & Yund, 1979) shape and orientation (Basole, White, & Fitzpatrick, 2003; Gizzi, Katz, Schumer, & Movshon, 1990; Mante & Carandini, 2005; Movshon, Adelson, Gizzi, & Newsome, 1986b), not directly related to the 2D motion of an object through space. However, these aspects of motion detection must constrain our thinking of how *two-dimensional* motion is computed. Accordingly, in the following sections I will review the major computations that occur in the early-visual stream. I then re-examine the '*aperture problem*' within the context of the established properties of 1D velocity detectors and in particular the motion energy model (Adelson & Bergen, 1985).

The Early Visual System

The initial registration of the visual world is achieved in the retina by two main classes of photoreceptors: *rods* and *cones*. Rods are the most sensitive, able to capture individual photons of light and have the most rapid temporal response duration. Cones are less sensitive and require more photons to achieve a single spike, but different cone cell morphologies generate

differential selectivity's for the long (red), medium (green) and (short) blue wavelengths and a greater number of cone cells across in the foveal retina result in a finer-scale representation of the world.

Although the spatiotemporal dynamics of retinal sensitivity provides primates with a rich description of the visual world, decoding higher level features directly from a two-dimensional representation of light *intensity* and *wavelength* is fraught with difficulties. Imagine the task of detecting chair in a natural environment - the chair could vary across a number of dimensions (e.g. viewing angle, colour, ambient lighting), which radically change the retinal image. Generating a simple template incorporating all possible patterns of light intensities or wavelengths generated by even an individual instance of a chair rapidly becomes computationally impossible.

The process of going from a low-level representation to a higher-level percept is known as *local-to-global*. Converging evidence from neurophysiology, psychophysics and imaging studies suggest that the primate visual system does not go directly from a representation of *local* intensities and wavelengths to high-level *global* percepts; instead a series of operations occur in the early-visual system that transform the initial retinal input from a temporarily varying representation of image *intensities* and *wavelengths* into representations defined by other stimulus dimensions such as *contrast*, *orientation*, *direction*, *speed* and *spatiotemporal-frequency* (Hubel & Wiesel, 1968). Since the computations of the early-visual system cannot increase the

information content of the initial retinal encoding (except for *stereopsis*, where depth is inferred from disparity between the two eyes perspectives), the computations that occur presumably serve to refine the initial representation of the world in a manner more amenable to the extraction of information pertinent to the organisms' survival. The main transformations appear are often described by *linear-systems-theory* and a wavelet-style decomposition of the world into its localised Fourier components.

Linear-systems theory

A class of model known as linear-systems-theory captures the main functional transformations that occur in the early visual system and is built on the concept of the *linear-receptive-field*. The *linear-receptive-field* is a weighting function that works across the dimensions of *space* and *time* in early areas (e.g. ganglion cells and the LGN) and upon increasing abstract dimensions further up the visual stream. Filtering involves multiplying each filter weight by the underlying input structure and summing the result. The aim of filtering is to transform the original input into another modality defined by the weighting function and is the primary means through which initial registration of the visual world in terms of *intensity* and *wavelength* of light is transformed in to other stimulus modalities such as *orientation* or *spatiotemporal frequency*. Each filter can be thought of as a template for a particular stimulus feature (e.g. *orientation*), the more a stimulus resembles the filter, the greater the magnitude of the filter response. The greatest problem with this approach is that the response of a filter is determined not only by the relative

structure of input but also the total energy of the stimulus. This leads to ambiguity in the output of a sensor known as the *principle of univariance*; for instance a particular sensor may respond equally to a low-contrast stimulus that closely matches the filter as it does to a high contrast stimulus that is a poor match for the stimulus. This is a key source of ambiguity that I will return to later in the *thesis*.

The successes of *linear-systems-theory* in describing the major transformation that occur in the early visual system is surprising considering the number of non-linearities associated with neural architecture not being incorporated into the model. This has led to the suggestion that non-linear processes related to neural coding are undesirable and that the visual system effectively attempts to re-linearise the signal. Evidence for this broad hypothesis comes from the study of orientation tuning in the primate brain. Here, the linear-model is able to capture the contrast invariance of orientation tuned cells noted in area V1 (Ferster & Miller, 2000) but is unable to model the firing rate of neurons which are by definition, only positive. This deficiency can be overcome by the addition of a rectification stage which provides greater neural plausibility but also leads to the *iceberg-effect* (Carandini, 2007), where the orientation bandwidth of a signal increases with (local) stimulus contrast. It is only after the addition of a squaring operation (Heeger, 1992a) and a divisive normalisation mechanisms (Heeger, 1992b) that contrast invariance is recovered (Finn, Priebe, & Ferster, 2007). This is encouraging because it means the complex and non-linear gain mechanism

just described can be well approximated by energy in the Fourier spectra (the square of the amplitude) that does not take into consideration *non-linear threshold* operations or the limited *dynamic-range* of cortical neurons.

Fourier analysis and wavelets

Early auditory and visual psychophysics could either describe the stimulus in terms of a series of discrete units across time and/or space, or in terms of the spectral or Fourier components of the signal (e.g. Campbell & Robson, 1968). A Fourier transform is a technique that can deconstruct a stimulus of arbitrary dimensionality into its spectral components. The notion seemed to gain support because human observers were found to be highly sensitive to spectral components of a signal and a number of low-level response properties such as contrast sensitivity are best described with reference to the spectral aspects of a stimulus (Anderson & Burr, 1989; Campbell & Robson, 1968). However it was noted by Dennis Gabor (1946) that the Fourier transform model could not account for how humans processed sound because the Fourier transform threw away any temporal localisation of the signal. Instead Dennis Gabor introduced the concept of the wavelet that can jointly encode both localisation and spectral components. The *wavelet* filter operates through two components: the *carrier* and the *envelope*. The carrier is the weighting function that determined the features to be extracted, whilst the envelope is a weighting function that limits the operation of the *carrier* to a limited region of space or time. The point is illustrated by the construction of a Gabor filter. The Gabor filter was introduced by Daugman (1980) and

named after Dennis Gabor and is the most commonly accepted model of local orientation processing in area V1 (Parker & Hawken, 1988) and by extension the direction selective cells of V1 explored in depth in the following sections (Adelson & Bergen, 1985). The carrier component of the Gabor is shown in (a), it is a sinusoidal modulation across the horizontal plane at 2 cycles per image. Alone the grating/filter is sensitive to vertically oriented elements across the full visual field. The envelope is shown in (b) and is a Gaussian centred upon the middle of the image and will be most sensitive to structure located at centre of the receptive field. By multiplying (a) and (b) a filter is generated which jointly encodes both spatial position and Fourier information.

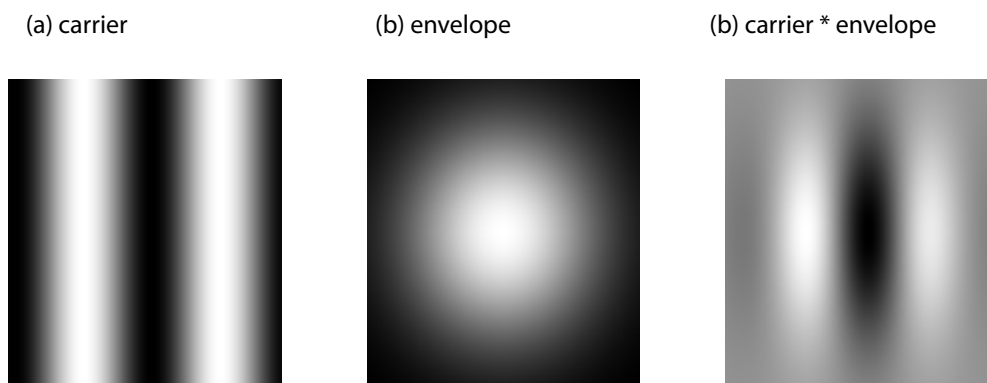


Figure 1-11 (a) a carrier filter tuned vertically oriented structure at 2 cycles-per-image. This filter is works across the entire visual field. (a) a Gaussian envelope, the filter is most sensitive to image structure falling with the center of the 2D Gaussian. (c) is constructed by multiplying (a) with (b), this generates a filter which jointly encodes spectral and spatial components.

The wavelet represents a compromise between the spectral sensitivity and the localisation of a filter. If we take the filter in (a) and assume it works over an infinite spatial region then it is sensitive to, and only to, image structure

vertically oriented at 2 cycles per image. In contrast the filter in (b) is not sensitive to a specific image feature, but is localised in space. (c) is the product a (a) and (b) and is jointly sensitive to both spectral and spatial information. However the flip side is that the sensor is also sensitive to structure near by its peak orientation and spatial frequency tuning. The interplay between spatial localisation and spectral sensitivity is determined by the relative width of the *envelope* and the *carrier* wavelength; increasing the spatial extent increases the spectral specificity of the filter but reduces its localisation, at the other extreme if the envelope was confined to a point in space it would have very precise spatial encoding but no stimulus selectivity. Thus the wavelet represents a fundamental compromise between spatial and spectral localisation, and this compromise must be inherited by later state visual processing. In the context of the 'aperture problem', one can consider the joint requirement of integrating over the finite but unknown region contained by a moving object and the ability to accurately determine the 1D velocities in the stimulus (that are in turn required to compute 2D velocity). For equations defining the play off between *spectral* sensitivity and localisation in *space* or *time* for the Gabor filter the reader is advised to look at Graham (1989),

Ganglion cells

The concept of wavelets is common to many theories of the early visual system and filtering using wavelet filters occurs as early as the Ganglion cells in the primate retina. These cells receive input from rods and cones spread over a region of retinotopic space known as the *receptive-field*. The basic

architecture of the *receptive-field* is that of a *centre* and *surround* and generates selectivity for features of a certain size and polarity (contained within the centre region but not the surround). Alternatively the receptive field configuration can be described as a *carrier* and an *envelope*. Here the *carrier* is a radial-sin wave, which provides spectral tuning, whilst the *envelope* is a Gaussian that constrains the frequency filter to a narrow region of space.

Lateral geniculate nucleus

Nearly 90% of ganglion cell axons relay to the LGN before passing on to the striate cortex (Silveira & Perry, 1991). Accordingly the response properties of the LGN are interesting because the majority of processing in higher visual areas must inherit the signal passed from the LGN and it has been shown that lesions of both pathways abolish almost all input to later visual areas (Shapley & Perry, 1986). The distinction between the magnocellular and parvocellular systems continues in the LGN with parasol ganglion cells and midget cell axons terminating in different layers of the LGN. This distinction continues as the axons of the magnocellular LGN cells synapse in layers 4C β of area V1 and the parvocellular cells of the LGN synapse in layers 4A and 4C α of area V1. The response properties of both layers are strikingly different in a few dimensions, but very similar in a number of others. The most striking difference is in colour with the parvocellular path cells having colour opponency in the red/green or blue/yellow wavelengths meaning they respond to colour change regardless of the relative luminance. In contrast magnocellular cells are insensitive to colour. Magnocellular and parvocellular cells also differ in their temporal frequency response profile. Parvocellular cells are

low pass and known as sustained, whilst the magnocellular cells achieve band-pass temporal-frequency selectivity through a biphasic modulation of the ON and OFF regions of the cells spatial receptive field (Cai, DeAngelis, & Freeman, 1997). Although neither parvo- nor magno-cellular cells are tuned to direction tuning they are important for the perception of motion (and vision in general) because higher visual areas must inherit their signal. For instance the properties of *sustained* and *transient* are also present in the response of V1 Direction Selective cells (Foster, et al., 1985) and are thought to account for behavior of observers' in masking paradigms (Anderson & Burr, 1989; Hess & Snowden, 1992).

Cells in LGN have spatial frequency profiles that smoothly sample to the spatial frequency dimension, however the same is not true for the temporal frequency tuning of LGN cells which are broadly divisible into low-pass and high-pass temporal frequency tuning. This presents a problem for downstream processing to 2D velocities that require accurate estimates of 1D speed and it is the subject of much debate how such broad tuning is converted into precise speed estimates. It has been shown that speed selectivity can be obtained from the ratio of activity in the two channels to produce cells whose activity positive correlates with stimulus speeds (e.g. Johnston, et al., 1992; Thompson, 1982) and through the selective combination of sustained and transient responses to generate speed tuning (Perrone & Thiele, 2002).

Orientation and direction selectivity

The axon terminals from area LGN predominately synapse in layer 4 of area V1 of the primary visual cortex, it is here that cells are identified which are selective for orientation and direction (Hubel & Wiesel, 1962). By integrating across receptive fields with positive and negative regions aligned in space a filter can achieve selectivity for particular surface orientations. This function (described above, see [Figure 1-11](#)) is known as the Gabor (Daugman, 1980) and is maximally sensitive to structure of a certain spatial-frequency and orientation. Thus while LGN cells are selective for spatial-frequency content at all orientations, V1 orientation cells are maximally selective for a specific spatial-frequency at one particular orientation.

The most commonly accepted model of V1 directionally selective cells is the motion-energy model (Adelson & Bergen, 1985) and is an extension of the Gabor Filter already described; The Gabor filter can develop temporal frequency selectivity by modulating the phase of the Gabor carrier (sinusoid) in time. This process can alternatively be thought of as a rigid translation of the carrier in the direction orthogonal to the orientation of the Gabor. Either way the process generates band-pass temporal frequency selectivity in the sensor in the direction orthogonal to the Gabor orientation. Thus the motion-energy model is not sensitive to 2D motion across the retina, but to the 1D component of motion that is perpendicular to the orientation of the sensor. In the introduction I introduced the 'aperture problem' as a problem of overcoming the motion signal elicited by a moving straight edge, however the selectivity of an individual motion energy sensor is inherently ambiguous

regardless of the stimulus class. In the case of an object with a broad local orientation structure the ambiguity may be resolved locally by integrating across a bank of motion energy sensors tuned to different spatiotemporal frequencies and orientation, but in the case of a straight edge the signal from a local bank of sensors must also be combined with other filters across space. This property has been exploited by Pack & Born (2001) who examining the response properties of MT cells to moving bars whose spatial extent (3°) exceeded the receptive field size of V1 cells, but was smaller than the receptive field of MT cells (Albright & Desimone, 1987). By recording the output of MT cells they were able to show that the initial response was in the direction orthogonal to the bar, but with time the selectivity of a cell moved towards the 2D direction of the bar and it is argued this pattern of firing may underlie the changing perception of bar stimuli (Lorenceanu, Shiffrar, Wells, & Castet, 1993) away from the 1D velocity(s) towards the veridical 2D direction with time.

The extension of the Gabor described thus far is often used as a model of a subset of V1 DS cells known as simple cells because they are sensitive to the phase and polarity of the visual stimuli (Hubel & Wiesel, 1962). In contrast cells known as Complex cells have a degree of position invariance, as their response is invariant with regard to the phase and polarity of a stimulus. Such response properties can be achieved by summing across simple cells tuned to opposite phase but otherwise identical tuning properties. In the case of the motion energy model of V1 direction selective cells operates by combining

the output of two simple cells with identical direction and spectral tuning but opposite phase. The output of each sensor is squared and then summed to produce a phase invariant output (Adelson & Bergen, 1985). The motion energy can be shown to be formally equivalent to a correlation-based method under some conditions (Adelson & Bergen, 1985; Reichardt, 1961; van Santen & Sperling, 1984), but only when the input to the 'correlate mechanism' are derived from orientation tuned cells (van Santen & Sperling, 1984).

Spatiotemporal frequency domain

The spatiotemporal frequency domain is useful because the selectivity of a motion energy filter (Adelson & Bergen, 1985) filter is well described by a Gaussian in the spatiotemporal frequency domain (Figure 1-12a) and a rigidly moving broadband and iso-oriented object is well described by a plane in spatiotemporal space (Figure 1-12b) (Watson & Ahumada, 1983). To uniquely specify a plane, a minimum of three points are needed, however given that the plane describing a global motion is constrained to pass through the origin, only two points are necessary. Pattern selectivity in MT cells are commonly thought to arise via a selective integration of motion signals across a plane in spatiotemporal space (Ahumada & Lovell, 1971; Perrone, 2004; Perrone & Thiele, 2001; Priebe, et al., 2006; Simoncelli & Heeger, 1998), a finding that is supported by psychophysical evidence that contrast detection is best when the motion energy is spread equally across a plane rather than confined to a subset of the plane (Schrater, Knill, & Simoncelli, 2000).

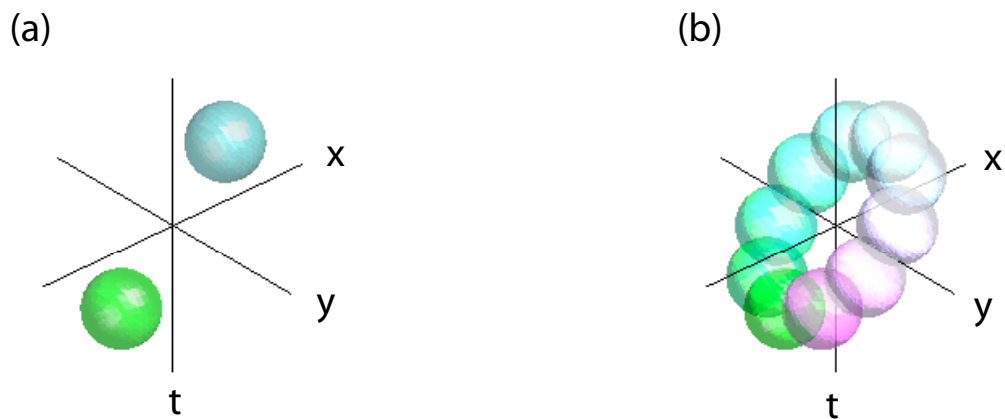


Figure 1-12 (a) The response profile of a motion energy filter in the spatiotemporal frequency domain. (b) A set of motion-energy filters tuned to one 2D velocity at one spatial-frequency.

Recent work has extended the modelling of the motion energy model to a bank of filters (Mante & Carandini, 2005) and demonstrated that the pattern of results in optical imaging studies in response to moving bars of different speeds, lengths direction and orientation (Basole, et al., 2003) is well described by the spatiotemporal properties of the stimulus. However the strength of Mante & Carandini (2005) is in demonstrating how the motion energy of a elongated structure may appear similar to the motion of a dot moving at a different velocity in the spatiotemporal frequency domain - in other words perceptually distinct visual stimuli appear similar in the spatiotemporal frequency domain. In this *thesis* I introduce a new means of representing the output of a bank of motion-energy filters in terms of the *speed* and *direction* of a sensor's tuning. This configuration is useful as it is common for a psychophysicist both to conceptualise the 'aperture problem', and to generate stimuli, in terms of the *speed* and *direction* of the

component of motion perpendicular to a surface orientation (e.g. Lorenceau & Alais, 2001; Lorenceau & Shiffrar, 1992; Mingolla, et al., 1992; Rubin & Hochstein, 1993).

The next section will introduce the methodology behind the modelling of 1D velocity estimation used in this *thesis*. The output of the bank of filters across a range of spatial frequency channels is then illustrated in response to rigidly translating stimuli with either a band-pass or 1/f spatial frequency profile.

Model of V1 motion energy

This section details the configuration of motion energy (Adelson & Bergen, 1985) sensors used to model the one-dimensional motion selectivity used throughout the *thesis* and putatively represents the direction selective cells that are found in area V1 of the primate brain. The aim of the 1D motion stage was to detect the full range of 1D velocities that the stimulus (moving at a known speed but unknown direction) may elicit. To do so, we created a bank of filters tuned to directions between 0-360° and to pseudo-speeds between 0-150% of the 2D speed, where the sensors pseudo-speed tuning is defined by the ratio of the spatial and temporal frequency tuning of the sensor (Equation 1.21)

$$speed = \frac{t_{freq}}{s_{freq}}$$

Equation 1.21

The 1D motion energy (Adelson & Bergen, 1985) filters were constructed in the spatial domain and were the product of a Gaussian envelope G and a Carrier signal S (Equation 1.22)

$$DG = G(x, y, t)S(x, y, t)$$

Equation 1.22

The Gaussian envelope was centred upon the middle frame t_m and upon the coordinate (x_a, y_a) .

$$G(x, y, t) = e^{-\left(\frac{(x-x_a)^2}{2\sigma_x^2}\right)} e^{-\left(\frac{(y-y_a)^2}{2\sigma_y^2}\right)} e^{-\left(\frac{(t-t_m)^2}{2\sigma_t^2}\right)}$$

Equation 1.23

The Carrier signal was a sinusoidal modulation in \mathbf{x} and \mathbf{y} with a wavelength $\lambda_{spatial}$, an orientation θ . The phase of the spatial sinusoid was shifted on each frame by $\Delta\lambda_{temporal}$.

$$S(x, y, t) = \sin\left(\frac{2\pi}{\lambda_{spatial}}(\sin(\theta)x + \cos(\theta)y) + \Delta\lambda_{temporal}t + \lambda_{phase}\right)$$

Equation 1.24

where $\lambda_{phase} = 0$ for even phase, and $\lambda_{phase} = \frac{\pi}{\gamma}$ for odd phase sensors.

The phase shift per frame $\Delta\lambda_{temporal}$ was calculated from the desired pseudo-speed tuning ϕ_{1D} of each local motion sensor, given the spatial frequency of the sensor using Equation 1.25 & Equation 1.26.

$$t_{freq} = \phi_{1D} S_{freq}$$

Equation 1.25

$$\Delta\lambda = t_{freq} 2\pi$$

Equation 1.26

As the desired temporal frequency for a particular speed increases in with spatial-frequency, the range of temporal frequencies is greatest in the high-spatial frequency channels. As the phase shift per frame cannot exceed 90° this sets a limit of the highest spatial frequency used (a phase shift of 90° per frame will lead to ambiguity in the direction tuned of the cell because a phase shift of (90° + x°) is identical to a phase shift of (-x°).

The Gaussian envelope was always a constant ½ of the wavelength of each filter (to maintain a constant direction and temporal bandwidth across filters) and the output of each filter was divided by the sum of the absolute of the respective field. This had the effect of flattening the response of filter to stimuli with a 1/f spatial frequency profile.

Convolution was achieved through multiplication of the signal and the sensor in the Fourier domain. The square root of the sum of the square of the real and imaginary components was taken to represent the motion energy at each point in space for each DS filter, a computation that is formally equivalent to the full rectified square of odd and even phase neurons to generate a phase invariant output (Adelson & Bergen, 1985).

$$E = \sqrt{g_{even}^2 + g_{odd}^2}$$

Equation 1.27

A global motion analysis was achieved by collapsing the spatial domain and summing across all DS filters tuned to the same spatiotemporal frequency and direction. Each spatial frequency channel could then be represented as a 2D pseudo-speed and direction image (Figure 1-13a), in which the intensity of each region represents the global sum of motion energy across DS filters whose velocity tuning is denoted by the regions position in the image. The only filter normalisation employed was to divide the output of each neuron by the sum of the absolute of the receptive field across space and time; this had the effect of evening out the expected $1/f$ spatiotemporal frequency spectrum. No gain control, normalisation or inhibition occurred between neurons.

A full bank of filters could then be defined as follows;

1. Sixteen directions evenly spaced around the clock.
2. Thirteen evenly spaced pseudo-speeds from 0% (static) to 200% of the carrier signal speed (3.95 deg/s).
3. Five spatial frequencies from 1 cycle per image to 64 cycles per image.

In order to allow the reader to move from the spatiotemporal frequency domain a single spatial-frequency bank of filters (tuned across directions and speed) is plotted in Figure 1-13(a). The output of the channel to a translating moving dot is shown in Figure 1-13 (b). A dot is an iso-oriented and

broadband stimulus; encouragingly the configuration of motion energy filters just described is able to capture the *cosine* relationship between the *speed* and *direction* of 1D velocities. This is a key observation of the *thesis* because it inspired the creation of a model of two-dimensional motion processing described later in the *thesis*.

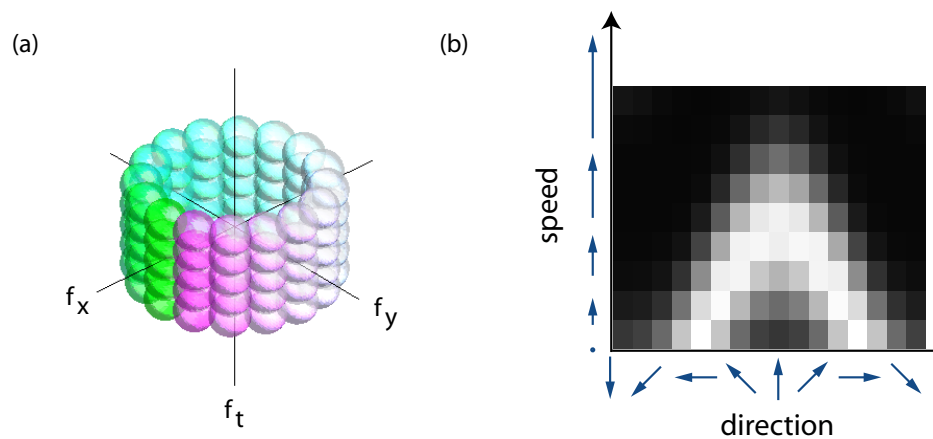


Figure 1-13 (a) bank of filters tuning to one spatial frequency channel and a range of speed and directions (b) the output of the bank of filters in (a) to a rigidly translating dot stimuli.

As the motion-energy filter is not truly speed tuned it is important to examine the response properties of the model for stimuli of various spatial frequency profile. As such Figure 1-14 plots the output of the filter bank to random noise patterns of either band-pass and $1/f$ spatial-frequency profiles. Like the dot stimuli, random noise patterns are iso-oriented and the cosine pattern of energy across speed and direction is present in all spatial-frequency channels. However in response to the band-pass stimuli the pattern of activity varies across the spatial-frequency channels; appearing squashed at low spatial frequencies and stretched in the high spatial frequency channels. The

cosine is only where we would intuitively expect when the peak spatial frequency of the stimulus matches the peak of sensitivity of the filter bank. This pattern results from the fact that motion energy filters exhibit an independence of spatial and temporal frequency tuning, consistent with the response properties of V1 DS cells (Foster, et al., 1985). The result is that when the spatial frequency of the sensor and stimuli do not match, the speed tuning of sensor is no longer reliable. Fortunately the filters responds best when both the spatial-frequency and temporal-frequency of the filter and stimulus are matched; accordingly summing across the spatial-frequency channels will allow the recovery of speed, assuming the spatial-frequency profile is symmetric (i.e. not skewed) or the profile is flat across the spatial-frequency channels.

Although natural scenes may vary across a huge number of dimensions, the spatiotemporal properties of natural scenes are relatively stable and exhibit an approximately $1/f$ drop of in amplitude (Dong & Atick, 1995; van Hateren, 1997). There is evidence in the both neurophysiology (Atick & Redlich, 1992; Carandini, et al., 2005; Dong & Atick, 1995) and psychophysics that the visual system accounts for this profile by *whitening* the distribution of natural scenes such that the response to natural scenes across all channels is even. Psychophysically evidence in support of the idea is found as observers over-estimate the high spatial and temporal frequency structure of a stimulus (Brady & Field, 2000; Cass, Stuit, Bex, & Alais, 2009), further subjects contrast detection functions are flat as a function of spatial and temporal frequency

when the stimulus is embedded within (masked by) a natural scene (Bex, Dakin, & Mareschal, 2005; Bex, Solomon, & Dakin, 2009).

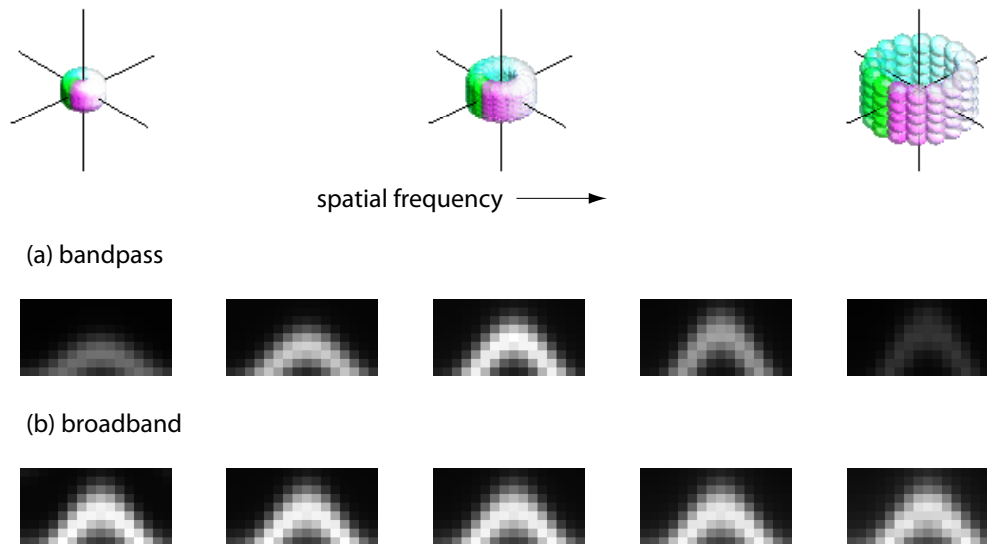


Figure 1-14 Output of the bank of motion energy filters tuned to a range of speed and directions and to five spatial-frequencies. The first row denotes the bank of motion energy filters in the spatiotemporal frequency domain (see [Figure 1-13](#)) (a) the second row shows the output of the filters in response to a band-pass moving dot; the spatial-frequency of the dot and filters are matched in column three and the amplitude of the filters produces a cosine as predicted from [Equation 1.1](#) in the introduction, however when the spatial-frequency of the filter is lower than that of the stimulus the cosine is squashed, when the spatial-frequency of the filter is greater than the stimulus the cosine is stretched. There is also less energy in the channels with non-matching spatial-frequencies. (b) In row three the output of the bank of filters to a broadband dot is shown, the cosine is centered upon the correct speed and direction in each spatial-frequency channel.

The 'aperture problem' - Psychophysics

The 1D distribution of motions is determined not only by an objects *speed* and *direction* but by the *orientation* content of the stimulus. Accordingly by manipulating the range of orientations present in the stimulus, one can alter the computational requirements for the successful estimation of global motion. Experiments exploiting this paradigm have often used plaid stimuli composed of two superimposed gratings whose 1D velocities specify a unique two-dimensional direction. By exploiting configurations in which the two-dimensional motion is not in either of the 1D directions of motion one can separate between one and two-dimensional selectivity. This paradigm was first exploited by Adelson and Movshon (1982) who showed that the percept of a symmetric Type I plaid was consistent with the conjoint two-dimensional motion. Soon after Movshon et al. (1985) were able to show that a proportion of cells in area MT were consistent with the property, leading to the notion that MT is the area where the 'aperture problem' is solved. However a number of subsequent psychophysical studies have shown that the perception of plaid motion is often not is always in the conjoint two-dimensional direction but towards the direction of 1D motion (Bowns, 1996; Burke & Wenderoth, 1993; Wilson & Kim, 1994; Yo & Wilson, 1992) and no firmly established model has been able to account for these differences.

Without an accepted model of two-dimensional motion processing it cannot be ascertained whether the pattern of results stems from a sub-optimal model of two-dimensional pooling (e.g. Wilson, et al., 1992), whether the pattern of

results is due to a number of stimulus attributes such as second-order structure (Yo and Wilson, 1992) or features such as zero-crossings, maxima or minima (Bowns, 1996; Stoner & Albright, 1996) or whether the pattern of results is optimal under some unknown behavioural or biological constraint.

To avoid the feature complications (e.g. zero crossings) associated with plaid stimuli, a number of other have studied the 'aperture problem' using stimuli composed of spatially disparate locally one-dimensional motions. As a discrete object cannot be described by just two orientations experimenters use an 'aperture paradigm' in which the orientation content of a stimulus is selectively occluded/exposed by viewing the stimuli passing under apertures. Using Type II stimulus classes both Mingolla et al. (1992) and Rubin & Hochstein (1993) demonstrated the perception of line elements passing under apertures was not in the conjoint 2D direction, but was also (like work using plaids stimuli) biased towards the direction of 1D motion. More recently the same problem has been studied using a global-Gabor array (K. Amano, M. Edwards, D. R. Badcock, & S. Nishida, 2009). The global-Gabor array is composed of number of local Gabor elements whose local one-dimensional velocities were configured to be consistent with a two-dimensional motion vector. Using a Type II orientation configuration paradigm Amano et al. also showed that observers' perception of motion was biased in the direction of 1D motion, consistent with previous research.

Contrary to studies employing Type II stimuli a number of lines of evidence point indicate that two-dimensional motion is often correctly estimated. By using a stimulus that rotated through space, rather than following a single trajectory, Lorenceau et al (1998) were able to simultaneously probe the perception of *speed* and *direction* (unlike the majority of studies on the 'aperture problem' which just probe direction judgements). If the computation of velocity were non-veridical either in *direction* or *speed* then perception of a structure in circular motion would be perceived as moving elliptically, yet observers correctly estimate the trajectory of motion. It was also shown by Amano et al (2009) that both the *speed* and *direction* estimates of global Gabor arrays was consistent with a veridical estimate of motion, whilst the combination rule used to infer the global speed and direction of locally unambiguous elements (plaids) employed a different rule, closer to the vector average. The studies employing the 'aperture paradigm' also note that the percept of direction is more accurate than would be predicted from a simple averaging solution (K. Amano, et al., 2009; Bowns, 1996; Mingolla, et al., 1992; Rubin & Hochstein, 1993), suggesting that some level of disambiguation is being achieved.

In an effort to consolidate or rationalise the mixed psychophysical results work by Weiss, Simoncelli and Adelson (Weiss & Adelson, 1998; Weiss, et al., 2002) used a Bayesian probabilistic framework and invoked the assumption that global IOC-like motion computation relies on a prior expectation of slow motion. This prior not only helps predict why slower speeds are generally

perceived at low contrasts (Johnston, Benton, & Morgan, 1999; Thompson, 1982) but also biases the perception of Type II stimuli towards the 1D motion signals. The sign of bias is always in the direction of 1D motion because the speed of a 1D motion must be either equal to, or slower than the global velocity, thus the influence of the prior is to draw the percept of motion towards the slower component motions and closer to the predictions of a vector-averaging scheme, consistent with psychophysical observations (Mingolla, Todd et al. 1992; Yo and Wilson 1992; Rubin and Hochstein 1993; Bowns 1996; Burke and Wenderoth 1993). The rationale behind the model is that the majority of motion occurs at low-temporal frequencies (Dong & Atick, 1995); because all perceptual judgements are noisy, the prior maximises the likelihood of correctly estimating motion. This approach is welcome because it attempts to provide a framework from which observers' misperceptions may be understood. However the use of a speed prior has not been constructed with reference to the behavioural or environmental context needed to determine optimality (Geisler & Ringach, 2009; Simoncelli & Olshausen, 2001), instead the prior was constrained by the capacity of the prior to determine psychophysical performance. Although this approach (also see; Stocker & Simoncelli, 2006) makes predictions about the 'shape' and function of a prior, the statistical advantage of the model has yet to be demonstrated. At this stage the use of a prior is open to the critique that the prior was simply a factor through which the model could be made to match the psychophysical data.

The psychophysical literature on motion transparency provides the inspiration for an alternative explanation; It is known that the ability to perceive multiple motions is contingent upon the difference in *speed* (Greenwood and Edwards 2006) and *direction* (Braddick, Wishart et al. 2002) difference between overlapping motion fields (i.e. the greater the speed or directional difference, the higher the probability that transparency will be perceived). This phenomenon is believed to result from the bandwidths of motion sensitive neurons; rather than each motion being described by discrete vector, motions are represented by distributions of activity; when the distributions of activity elicited by separate motions is spatiotemporally close the distributions overlap and are hard to distinguish. For instance it is claimed that transparent motions are harder to detect around the oblique directions (Greenwood & Edwards, 2007) due to the increased directional bandwidth of motion sensors in these directions (Dakin, et al., 2005b; Gros, Blake, & Hiris, 1998; Li, Peterson, & Freeman, 2003).

Evidence that this same line of argument can be applied to the 'aperture problem' comes from studies in which the angular separation of the 1D velocities has been systematically manipulated (Burke and Wenderoth 1993; Bowns 1996). In both studies misperception only occurred when the stimulus were in Type II configuration *and* when the 1D velocities were close in Velocity space. The reader is advised to look at (Weiss & Adelson, 1998) for a re-plot of the (Bowns, 1996) data in which observer bias is plotted as a function of the angular separation between the component gratings of a

plaid. If the motion stream can correctly identify the 1D velocities in a moving stimuli then the motion stream can theoretically solve the 'aperture problem' in a unbiased manner, however it is known from studies of transparency that motion signals may interact and cause a phenomena known as motion repulsion in which the angle between two motions is overrepresented (Marshak & Sekuler, 1979; Rauber & Treue, 1999). Regardless of the underlying mechanism, evidence from motion transparency and motion repulsion demonstrate that the veridical discrimination of motions is harder when the signals are close in velocity space and it may be suggested that the inability to correctly determine the 1D velocities of a stimulus, is the primary cause of observers' errors in the estimation of 2D direction. Note, this theory is in sharp contrast to those authors who propose that 2D velocity is computed by a sub-optimal model such as a Vector-Average scheme.

Temporal aspect in the computation of 2D velocity

Studies employing perceptual (Lorenceanu, et al., 1993; Yo & Wilson, 1992), oculomotor (Masson, Rybarczyk et al. 2000) and neurophysiological (Pack and Born 2001) paradigms have revealed that the response of the motion stream is initially biased in the direction orthogonal to the locally dominant orientations, but then switches (partially or fully) to the direction of 2D motion. The temporal duration of the switch appears to depend on the exact nature of the stimuli, for instance suprathreshold stimuli with distinct but locally overlapping components appear to refine relatively quickly (~160 ms; Yo and Wilson 1992), whilst studies employing translating lines oriented obliquely to the direction of motion (Lorenceanu, et al., 1993; Masson, Rybarczyk, Castet, &

Mestre, 2000) are resolved more slowly (~400ms). Thus although the visual system appears capable of responding to the unambiguous signals that arise from local elements with broad orientation structure, the detection of motion signals stemming from ambiguous line elements appears more immediate.

EXPERIMENTAL CHAPTERS

Overview of methodology and rationale

Psychophysics is the process of relating a stimulus to behaviour. To do this, the psychophysicist makes inferences about the internal mechanisms that bring about an observers' behaviour. Broadly speaking two main classes of experiment can be conducted: the *hypothesis-driven* test and the *explorative* test. In the former, experiments are designed to *falsify* specific hypothesis whilst in the latter the aim is to identify which aspect of a stimulus drive observers' behaviour and to speculate upon the mechanism. The choice of stimulus will often reflect the aim of the study; *hypothesis-driven* experiments tend to employ highly constrained stimulus classes so that behaviour can be easily related to the stimulus and theory to hand, whilst more *explorative* experiments may incorporate more complex stimuli with the aim of indentifying which of a number of stimulus dimensions drive behaviour. Each approach is open to criticism; if the experimenter employs a very complex stimulus he or she runs the risk of not being able to constrain any conclusions; conversely incorporating too little detail runs the risk of throwing the proverbial '*baby, out with the bathwater*'.

Interestingly, this axiomatic debate (common to all scientific practice) has recently lifted its head in vision science (Felsen & Dan, 2005; B. A. Olshausen & Field, 2005; Rust & Movshon, 2005). Here the argument concerns the stimulus class used with a contemporary drive to move away from highly constrained stimuli such as bars or gratings towards stimuli that incorporate richer or more

naturalistic statistical properties (Bex, Mareschal, & Dakin, 2007; Carandini, et al., 2005; Dakin, et al., 2005b; Dumoulin, Dakin, & Hess, 2008; Felsen & Dan, 2005; Felsen, Touryan, Han, & Dan, 2005; Geisler, Perry, Super, & Gallogly, 2001; Geisler & Ringach, 2009; van Hateren, Ruttiger, Sun, & Lee, 2002). While the success of highly constrained stimuli to reveal the primary properties of the transformation that occur in the early visual system has been rigorously defended (Rust & Movshon, 2005), it is argued that many of the properties of the visual system are not revealed without exposure to the stimulus classes they have evolved and adapted to process (Felsen & Dan, 2005), or in the case of *natural-system-theory*, the behavioural context in which a visual task is carried out (Geisler & Ringach, 2009). Such claims are largely based on principles of efficiency and the observation that maximal efficiency can only be defined as function of both the sensor system and the range of stimuli (i.e. statistics) a system is designed to process (Simoncelli & Olshausen, 2001). The approaches may be viewed as complimentary (Carandini, et al., 2005); naturalistic stimuli can both help identify which important dimensions are absent from constrained, artificial stimuli (Felsen & Dan, 2005) and allow us to build a picture of how well existing models predict response to natural scenes (e.g. David & Gallant, 2005; Hsu, Borst, & Theunissen, 2004; van Hateren, et al., 2002).

The approach taken in the first two experimental chapter of this thesis is consistent with the contemporary drive to incorporate more 'naturalistic' statistics into the stimuli used to probe the '*aperture problem*'. In the first

experimental chapter I examine whether the contour statistics of natural images influence our ability to determine two-dimensional motion. Specifically, it is known that human observers are better at detecting contours when the orientation statistics conform to those in naturally occurring images (Geisler, et al., 2001) but it is less clear whether contour statistics affect the ability of observers to group local motion across space (i.e. solve the 'aperture problem'). Experiment one probes this issue by manipulating the second-order contour statistics of a scene across space, whilst maintaining the identical local motion signals and examining observers' direction thresholds in the two-alternative forced choice paradigm.

In the second experimental chapter a *reverse-correlation* paradigm is employed to examine which stimulus dimensions in natural images influence observers ability to determine two-dimensional motion. This approach is notably different from the majority of studies that probe the 'aperture problem' using highly constrained stimulus classes. Such studies probe the 'aperture problem' using stimuli composed of just two oriented elements (the minimum number needed to uniquely specify a two-dimensional velocity), whilst natural scenes tend to contain much broader orientation content. The results from such experiments using just two orientations have yielded very little consensus as to how two-dimensional motion processing occurs and the behaviour of observers has not easy to express in terms of a rational solution (but see; Weiss & Adelson, 1998; Weiss, et al., 2002). Given the observation that the optimality of a biological system may only be described in the

context of the environment the visual system has evolved to process (Simoncelli & Olshausen, 2001) I ask whether the psychophysical data may be better understood in terms the response to naturally occurring stimulus classes.

Experimental chapter II uses natural scenes whose exact stimulus properties are unknown. To relate the underlying statistics of the natural scene to observer behavior it is necessary to estimate the stimulus properties. The chosen method was to use biologically inspired filtering with Gabor (Daugman, 1980) and motion-energy (Adelson & Bergen, 1985) filters. This has the disadvantage that the results of the analysis are dependant on the choice of filter parameters, but it has the advantage that it forces the experimenter to incorporate practical and biological constraints into the analysis. The reverse engineering approach reveled that the observers' behavior could be reliably related to the output of Gabor filters in response to natural scene. Moreover the application of a bank of motion energy filters revealed that the theoretical distribution of 1D velocities (i.e. a cosine) was apparent in the response of the filters to natural scenes. This observation inspired the construction of a template-matching model of two-dimensional motion processing in the proceeding chapter that was shown to produce qualitatively similar behavior to observers.

The approach taken in the first two chapter of the thesis is to use more complex and 'naturalistic' stimuli when probing the 'aperture problem', however the work does not 'go all the way' and probe vision using type of visual diet that humans are exposed to on a day-to-day basis. The visual diet that humans or primates are exposed to is typically referred to as natural

vision and it is the end goal of vision science to build models that are able to capture human or primate visual function in response to natural scenes. To a limited extent this had been achieved in neurophysiology for models of early visual function such as retinal and LGN processing (for a review, see;Carandini, et al., 2005). The advantage of such an approach is that it allows for a robust test of our models of visual processing.

In this work I confine the experimental stimuli to full-field translations in the fronto-parallel plane. This clearly limits the extent of the research because the range of naturally occurring motions may occur in any plane and I do not consider the problems associated with extracting the motion of an object from a static background, a question directly confronted in (Johnston, et al., 1992), the computation of 3D motion (e.g. Harris & Drga, 2005), natural motion (e.g. Neri, Morrone, & Burr, 1998) or any other of a multitude of additional problems associated with natural vision. Instead the work is best viewed as a small step towards the end goal of natural vision. I ask whether the pattern of observers' responses to highly constrained stimulus (e.g. plaids) translating the fronto-parallel plane is consistent with observers' response to my complex visual stimuli translating in the fronto-parallel plane.

The final experimental chapter differs from the previous two chapters in that it was designed specifically to probe the ability of the model described in the computational chapter to predict observers' errors in a two-dimensional motion task. In this regard the stimulus was a global-Gabor array (K. Amano, et al., 2009) in which the experimenter had close control over the orientation content of the stimulus (unlike the preceding chapters). A double pass technique was used to divide the observers' variance into predictable and

unpredictable components. This procedure allows one to compare the total variance the model is able to capture against the total variance captured by the observer on stimulus retrials.

(2) Experimental Chapter No.1

The influence of Natural Contours in motion processing

When global motions are composed of locally ambiguous elements (e.g. straight edges or lines) the local motion they produce must be integrated across space to achieve an unambiguous motion estimate. The '*aperture problem*' then becomes both the problem of correctly integrating across *speed* and *direction* and the problem of correctly grouping elements belonging to individual objects and segregating those motions belonging to other objects. This experimental chapter will examine the role of natural contour structures in determining our ability to solve the '*aperture problem*' by combining motion signals across space.

Relative to noise with a spatial frequency matched to that of natural scenes, contours in natural scenes exhibit two statistical regularities that I will probe in this chapter. Firstly, contours are regions in which the phase structure across spatial-frequency channels is aligned (Attneave, 1954; Barlow, 1961), this property is not always present in other studies designed to probe contour integration because it means the signal from each contour element is not constrained to a limited region of space. Secondly, contours tend to vary smoothly across a scenes and it is know that observers' are more likely to detect contours that conform to the second-order orientation statistics of natural scenes (Geisler, et al., 2001). Accordingly this chapter will probe 2D

motion processing using a two-alternative-forced-choice paradigm (2-AFC) for stimuli containing natural and disrupted contour structure and for stimuli with a broadband or high-pass spatial-frequency structure.

Contour Structure

Natural scenes contain a preponderance of edges (Attneave, 1954; Barlow, 1961) whose properties tend to vary smoothly across a scene, a characteristic termed 'good continuity' by the Gestalt psychologists (Wertheimer, 1958.). More formally the relationship has been defined in terms of the probability that one edge point predicts the occurrence of another edge point at a given distance (d), orientation difference (ϕ) and contour angle (θ) (Geisler, Perry, Super, & Gallogly, 2001). Broadly speaking the smaller ϕ , θ & d , the more likely one is to encounter another edge point. Psychophysicists have examined if and how the visual system exploits such regularities using paradigms in which small oriented elements (typically Gabor) are used to build contours with particular second-order relations (e.g. co-circularity) which are then embedded in a field of randomly-oriented distracter elements (e.g. Field et al, 1997). In this paradigm, contour detection must involve global integration since it operates over spatial distances and across spatial phase in a manner that could not be achieved by conventional V1 neurons (Hess & Dakin, 1997). Sensitivity to contours has been shown to increase with lower curvature (smaller ϕ & θ) and contour length (Field, Hayes, & Hess, 1993), consistent with the statistics of natural scenes.

While it is clear that the second-order distribution of orientations across the visual field is critical for determining our ability to see static extended contours, the role of such statistics in motion processing is less clear. Second-order orientation statistics can certainly influence motion processing when the underlying elements are locally ambiguous. This point is illustrated by Lorenceau & Shiffrar (1992) who demonstrate that the perceived directions of four moving bars (Figure 2-1) can be dramatically altered by changing the appearance of occluding elements. Although the bars in Figure 2-1(a&b) move in an identical fashion (sinusoidally translating in the direction perpendicular to their orientation) the perceived directions of motion are different. In Figure 2-1(a) the bars are perceived to move as independent pairs, but when the occluders are present in Figure 2-1(b), the individual components 'cohere' and appear to move as a rotating diamond whose vertices are occluded. This dramatic change in percept is thought to arise from a change in the classification of the end points from 'intrinsic' (i.e. part of the object) to 'extrinsic' (arising from occlusion by another object). This argument is intuitive. When the endpoints are considered part of the object eliciting motion there is only one physically realistic interpretation: independent motion. However, if the endpoints are due to an occluding object, the motion signal generated at the intercept bears no relation to object motion. In isolation, this leaves ambiguous the speed and direction (velocity) of each bar (Figure 2-1d) and motions must be combined across space to achieve a veridical 2D motion estimate.

The stimulus of Lorenceau & Shiffrar (1992) is ideal for study because without information that can correctly constrain the percept in one direction or another, it is possible to probe the *priors* and *assumptions* the visual system uses to bind elements or individuate elements. A number of factors increase the probability that elements will be integrated; observers' are more likely to individuate elements presented in the fovea and more likely to integrate eccentrically viewed elements. If the occluding elements are sharp squares then segregation is more likely than for blurred edges (Lorenceau & Shiffrar, 1992), reducing the contrast of the stimulus promotes integration (K. Amano, et al., 2009; Lorenceau & Shiffrar, 1992; Lorenceau & Zago, 1999) and changing the percept of the intersection from intrinsic (to the moving object) extrinsic (occluding the object) promotes integration (Shimojo, Silverman, & Nakayama, 1989).

Recent research has indicated that one-dimensional (1D) and two-dimensional (2D) signals are treated differently by the motion stream; by measuring the perceived direction of multiple Gabor stimuli Amano et al. (2009) have shown that integration of 1D plaids occurs in a veridical manner, whilst integration of 2D plaids produces answers in line with predictions from a Vector average (VA) rule. Furthermore Bowns and Alais (2006) have shown that adaptation to stimuli yielding a VA solution generates a large shift in the perceived direction towards the IOC interpretation and vice-versa. Such adaptation suggests that the two solutions operate independently and compete to determine the overall percept of motion.

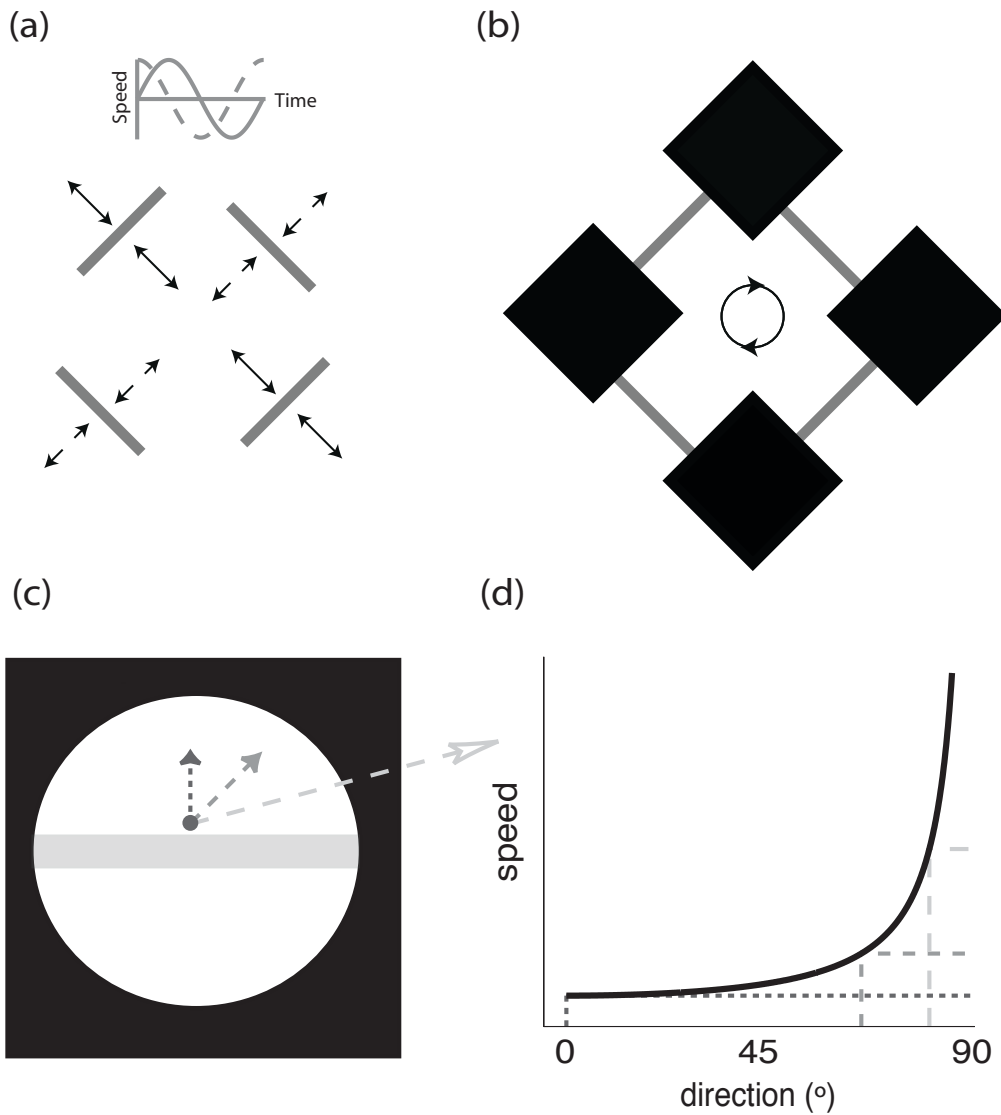


Figure 2-1 Occluded diamond. The influence of form on motion integration (Lorenceanu & Shiffrar, 1992). The movement of the bars is identical in (a) and (b) (sinusoidal translation in the direction perpendicular to their orientation). In (a) the bars appear to move independently of each other, but in (b), when the apertures are made explicit, the individual components 'cohere' and appear to move in directions consistent with a rotating diamond. (c) The ambiguity associated with a moving bar. The exact speed and direction (velocity) of the bar is unknown, however it is known that the veridical velocity must fall on a 'constraint' line that can be inferred from the speed perpendicular to the bars' orientation, as shown in (d) – By solving for two or more such lines, a unique vector can be found, in the case of a rigid object moving in 2D space this vector reflects the veridical velocity.

Co-occurrence of structure across spatial frequencies

As well as the second-order spatial regularities discussed so far, natural scenes have the tendency for content across spatial frequencies to be spatially aligned (Attneave, 1954; HB Barlow, 1961). The early decomposition of retinal signals cannot fully encapsulate this property, as their spatial frequency tuning is too narrow (Anderson & Burr, 1989; Blakemore & Campbell, 1969), accordingly signals must be recombined to achieve the broad SF tuning observed in the integration of both static (Dakin & Hess, 1998) and moving contours (Bex & Dakin, 2003; Ledgeway & Hess, 2006; Ledgeway, Hess, & Geisler, 2005). Such broadband integration is not without danger - an inflexible integration mechanism increases the risk that inappropriate or noisy signals may be integrated. Variation in the extent of integration across scale has been shown in static contour tasks, with contour integration being spatially broadband along straight elements but narrowband at areas of high curvature (Dakin & Hess, 1998). Functionally, this arrangement should reduce the impact of noise by integrating where the signals are likely to be the same across scale (straight edges) but selectively integrating when the signal will vary across frequency (curved edges).

In the motion domain global integration has been shown to be broadband in detection tasks (Bex & Dakin, 2002) and in motion after effects (MAE) when isotropic flickering test stimuli are employed (Ashida & Osaka, 1994; von Grunau & Dube, 1992). For instance, while participants are unable to detect the motion of locally band-pass dots whose spatial frequency content do not

overlap (Bex & Dakin, 2003; Ledgeway, 1996) the perception of global motion (rotation, translation & expansion) can be masked by noise elements at spatial frequencies that are remote from signal elements (Bex & Dakin, 2002). This suggests that global motions detectors are not only SF broadband but are unable to selectively tune their input with respect to the stimulus type at hand. Such 'rigid' integration has also been observed in the orientation domain where Schrater, Knill & Simoncelli (2000) found that thresholds for a signal embedded white noise are near optimal when the energy is uniformly spread around one speed plane, but sub-optimal when the energy is confined to isolated sub-sets of the space.

The study of apparent motion has established that d_{\max} (the greatest distance that motion may be detected over two successive frames), scales inversely with SF under most conditions (Baker, Baydala, & Zeitouni, 1989; Cleary & Braddick, 1990; Eagle & Rogers, 1996; Morgan, 1992) but much less work has studied the influence of SF in global motion tasks. There is some evidence that low SFs play a special role. In Bex & Dakin (2002), masking was strongest for low SF noise elements, even when matched for visibility, suggesting that coarse information is preferentially integrated. Further evidence that low SFs are used to 'bind' high SF comes from the phenomenon of 'motion capture', where high SF structure is perceived as moving in the direction of the low SFs, even when the directional signals are centred on opposing directions (Ramachandran & Cavanagh, 1987).

This chapter introduces uses a novel stimulus to explore the influence of second-order statistics and spatial-frequency in a 2AFC direction discrimination task. Stimuli are generated by band-pass filtering white noise and then performing a thresholding operation. The results are binary blob images (Figure 2-2) containing smooth and relatively sparse contours, which I term “naturalistic” simply because this form of contour structure is more commonly observed in natural scenes than in e.g. two-dimensional noise. The SF profile may be described as low-cut: The initial filtering is band-pass (0.75 c/deg), but the threshold operation introduces high spatial-frequencies to the signal. Thus, the spatial-frequency profile has a low-cut off, but not a high-cut off.

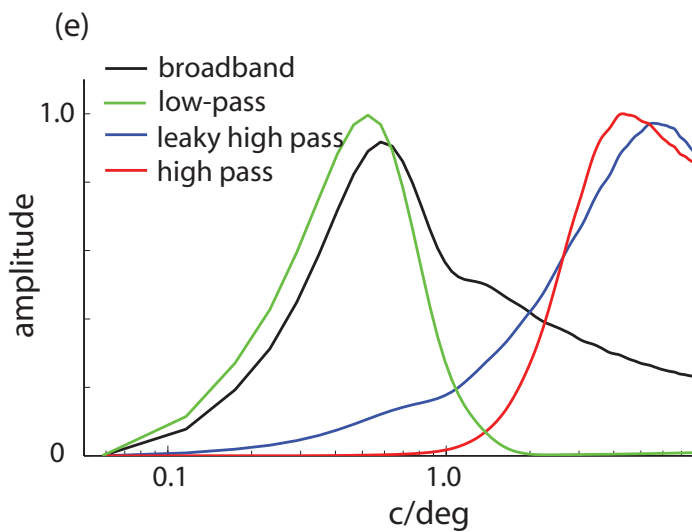
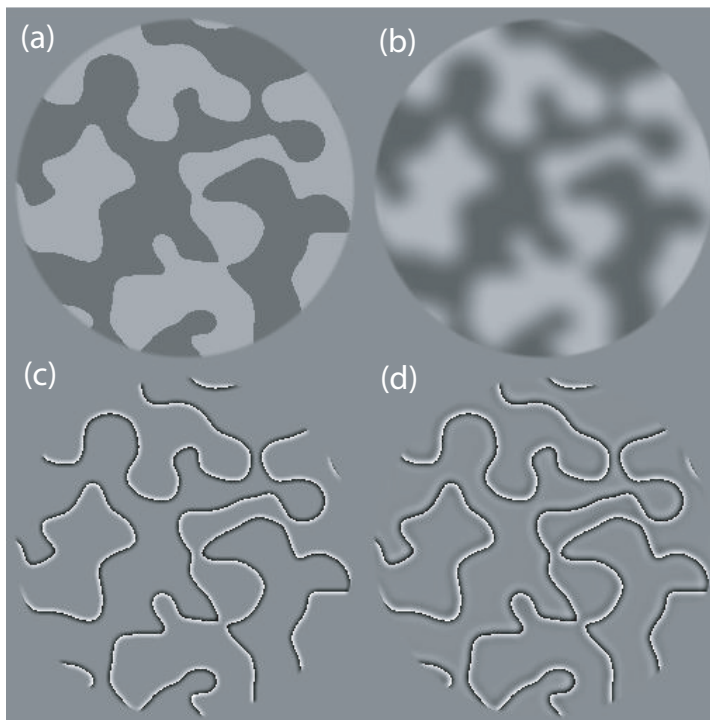


Figure 2-2 Examples of the stimuli employed. (a) Broadband (b) Low-pass, Gaussian filtered version of (a) (c) Leaky high-pass – generated by subtracting a Gaussian blurred version of (a) from (a) (d) (Strictly) High-pass stimulus – generated by further subtracting a Gaussian blurred versions of (c) (see methods). (e) Amplitude spectra of (a, b, c & d), note how the low frequency component of (c) is leaky but the no-illusion stimuli reaches an amplitude of zero at a low SF.

The stimuli used are a significant deviation from the type of stimuli used in most motion studies. For instance, while dot-stimuli may be resolved locally, the use of drifting-Gabors or straight edges forces the visual system to combine signals over space to disambiguate 1D velocities (Kaoru Amano, et al., 2009; Lorenceau & Alais, 2001) In contrast the ability of the visual system to locally resolve motion signals stemming from curved elements is unclear. In this regard there is an interesting inter-play between the ability of the motion stream to accurately identify component motion (presumably easier for straight edges) against the ability to resolve signals locally (presumably easier for areas of high-curvature). Furthermore areas of low-curvature may aid the binding of spatially disparate elements (as shown in contour detection paradigms e.g. Geisler, et al., 2001). For the purpose of the current study it is worth noting that increasing the area of the carrier signal exposed to the observer leads to large improvements in discrimination thresholds. Thus when the aperture size is small, if any disambiguation is occurring on a local level, the precision of such estimates is poor.

Like many studies designed to probe the aperture problem, I restrict the analysis to motion within two-dimensions, I concede that this excludes many of the spatiotemporal relationships present in natural environments, but note that 2D motion is consistent with the sub-set of naturally occurring motions that occur within the fronto-parallel plane.

Given the non-fractal nature of our stimuli, Experiment 1 probes the role of the low and high SF components of our stimulus. This Experiment provides an essential control for Experiments 2&3, which explores the effect of disrupting the second order statistics in a direction discrimination paradigm.

Methods

Subjects

Three psychophysically experienced observers (DK, SD, JG) with normal or corrected-to-normal vision took part in all experiments. In Experiment 2, subject JG was replaced with JC.

Apparatus

Stimuli were generated on an Apple iMac, running MATLAB (MathWorks) using elements of the Psychtool-box (Brainard, 1997; Pelli, 1997). Stimuli were displayed on a Dell, Trinitron CRT with spatial and temporal resolution set to 1024 * 768 pixels and 85 Hz respectively. The screen was viewed from a distance of 1.5m so that one pixel subtended 0.35 arcmin. of visual angle. The monitor signal was passed through an attenuator (Pelli & Zhang, 1991), following which the signal was amplified and copied (using a line-splitter) to the three guns of the monitor resulting in a pseudo 12 bit monochrome image. Monitor linearization was achieved by recording the relationship between the signal and the monitor intensity (Minolta LS 110 photometer), to create a linearization look up table that was passed to the Psychtoolbox internal colour look up table.

Stimuli

The mean luminance of the stimuli was 30.5 cd/m² with a root-mean-square contrast of 0.20. Stimuli were viewed through a large 2D raised cosine aperture (tapered annulus radius; 1.38 arc min) presented in the centre of the display. The radius of the aperture was either 2.95° or 1.17° (two viewing areas were employed to control against ceiling effects; see below). The smaller aperture size was equal to the total signal area in the locally apertured condition in Experiment 2. Due to the tapered annulus used, the visible area was taken to be the area above contrast detection threshold in keeping with the detectable area of Gabor stimuli (Fredericksen, Bex, & Verstraten, 1997).

Stimuli were generated by spatially band-pass filtering random noise using a 2D Laplacian-of-Gaussian filter - $\sigma = 22.8$ arc min - and then thresholding the result at mean luminance to generate binary "blob" images. An example stimulus is illustrated in Figure 2a. This procedure allowed us to rapidly generate complex shapes with a broad SF profile. 200 such images were generated. On each trial a random image was selected, with replacement. Low-pass images were generated by convolving the broadband images with a Gaussian filter ($\sigma = 5.4$ arc min., Figure 2b). One set of high-pass images (Figure 2c) was generated by subtracting a Gaussian ($\sigma = 2.1$ arc min.) filtered version of the broadband images from the source image. This process is "leaky" – allowing through some low-frequency information and leading to the Craik–Cornsweet–O'Brien (CCOB) illusion (Cornsweet, 1970; Craik, 1966; O'Brien, 1958) to be present in our stimuli (observe how the areas within the contours appear to be light or dark even though the luminance of each

patch is equal). To control the potential influence of this illusory coarse-scale structure the low-pass image was subtracted twice more (Figure 2d). This procedure has previously (Dakin & Bex, 2003) been shown to completely abolish the CCOB effect by both further attenuating the low-frequencies and nulling the effect by reversing the polarity of the inter-blob areas.

The carrier component of stimuli translated at a speed of 3.93 deg/s for 0.3 seconds (refresh rate 85 Hz = 26 frames) in near-upwards directions. Motion was generated using operations built in to the computer's graphics card, accessed using the OpenGL programming language. During each trial, the stimulus was passed to the graphics card buffer. Stimuli (11.5 X 11.5 deg.) were greater in size than the viewing aperture (radius 2.95 deg./1.17 deg.), during each frame a segment of the original image was displayed. By smoothly varying the region of the original image presented to the monitor/subject a percept of rigid translation of the image through the aperture was generated. To avoid the potential effects of an orientation bias, the underlying image was randomly flipped from left to right between trials. Between trials a phase-scrambled version of the original broadband stimulus was placed within the viewing area and the following trial was initiated immediately following the observer's response.

Procedure

A method of constant stimuli (MCS) was used to assess fine direction-discrimination with such patterns. A small offset clockwise (CW) or

anticlockwise (ACW) was added relative to vertical upwards motion. The observer's 2AFC task was to fixate on a continuously present cross at the centre of the monitor and to indicate the direction of motion (CW or ACW of vertical upwards motion), guessing if necessary. Audio feedback was provided following incorrect answers. The offset was between $\pm 7^\circ$ (large radius) and $\pm 10^\circ$ (small radius) at 17 equally spaced intervals. Each point was measured 17 times per run and all participants completed at least 2 runs (i.e. 578 trials per condition), extra trials were added if the psychometric function was under or over constrained. All conditions were randomly interleaved.

The procedure for deriving thresholds was identical to (Dakin, Mareschal, & Bex, 2005a); the psychometric function was fit with a wrapped Gaussian and the standard-deviation parameter of the best fitting function was taken as the estimated threshold. A bootstrapping technique was employed to estimate 95% confidence intervals on these estimates; data were re-sampled with replacement across each point (assuming binomial error) in the psychometric function a total of 1024 times and the function refit. In all plots, error bars indicate 95% confidence intervals on the threshold estimates.

Experiment 1: Dependence of direction discrimination on spatial-frequency structure

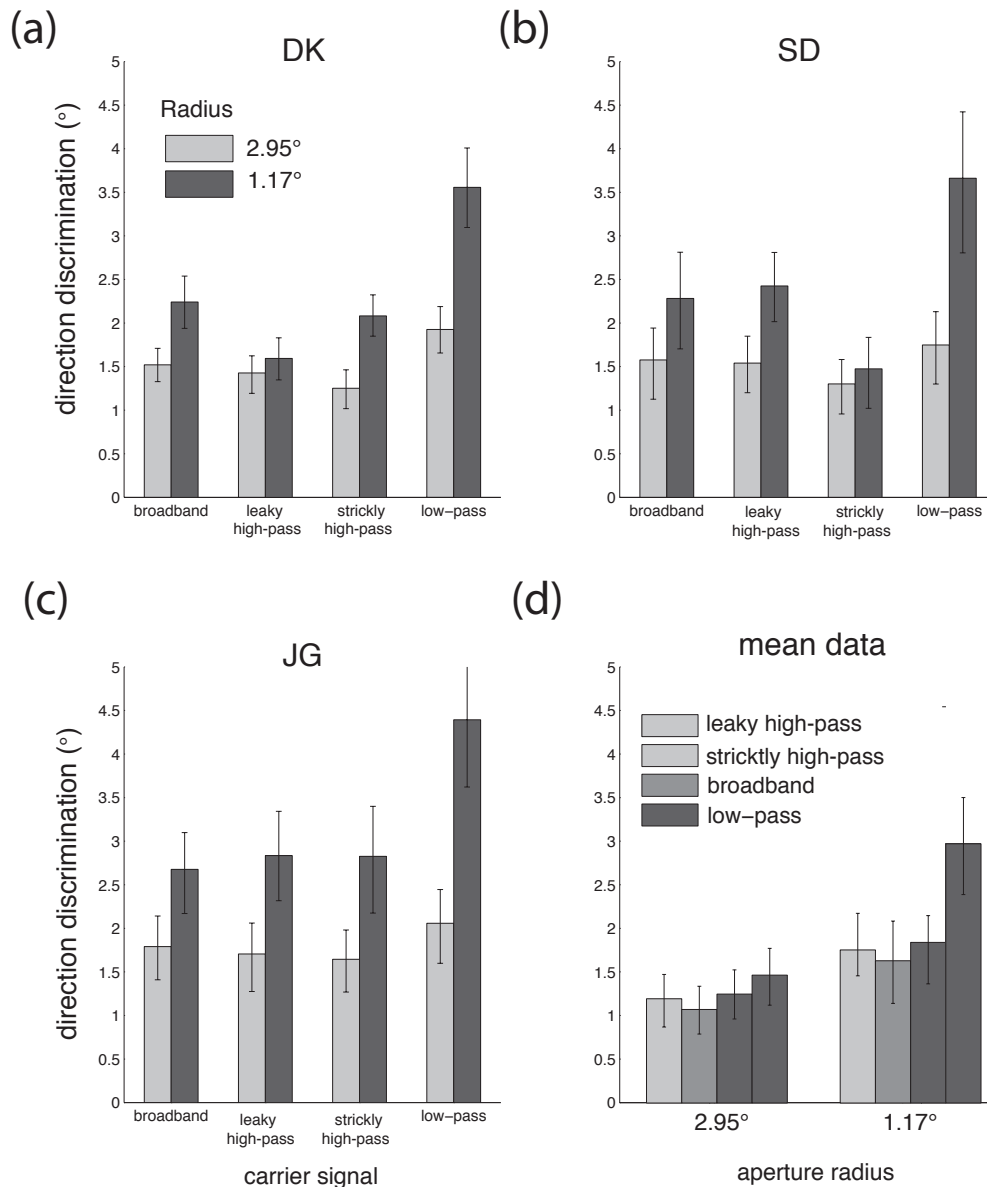


Figure 2-3 Results, experiment 1 (a, b & c) Direction discrimination thresholds for three observers (DK, JG & SD), measured with four underlying carrier signals (broadband, leaky high-pass, strictly high-pass and low-pass – see text for description). Error bars indicate 95% confidence intervals. Note that performance was worse over the smaller aperture (dark grey) condition indicating that performance was not at ceiling. (d) Mean thresholds for the three observers after normalization (ie centering psychometric functions on zero) to correct biases, then pooled across participants. Thresholds were lower for high-pass than broadband conditions,

but not significantly so. Thresholds were significantly higher for low-pass stimuli in the smaller aperture condition.

In the first experiment, I sought to determine the relative influence of information across SF channels in our 'naturally' contoured stimuli. [Figure 2-3](#) plots direction discrimination thresholds for DK, JG & SD, measured with four underlying carrier signals (broadband, leaky high-pass, strictly high-pass and low-pass). Error bars indicate 95% confidence intervals. Note that performance is worse with the smaller aperture (dark grey), indicating that performance in the smaller aperture conditions is not at ceiling. In [Figure 2-3d](#), thresholds for DK, JG & SD were first mean adjusted to zero to correct for biases (i.e. the mid point of the psychometric function were centred on 0 deg), then pooled across participants. Thresholds were broadly similar for the high-pass and broadband conditions, but thresholds were significantly higher for low-pass stimuli in the smaller aperture condition. Thus, direction sensitivity increases either by increasing spatial frequency or increasing aperture size (at least for the conditions tested). This indicates that the signal is less reliable at the low SF's, despite there being an identical number of cycles in the contour structure of each SF channel. Finally, these results reveal no special role for low SFs, unlike that observed motion capture (Ramachandran & Cavanagh, 1987). Given that I tested only four spatial frequency and two aperture-size conditions, I cannot make more general assertions about the relationship between these parameters and direction discrimination performance. For example, there may be subtle inter-actions between parameters, effects that saturate with increasing SF, etc.

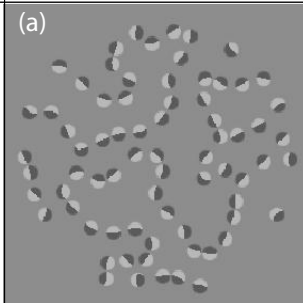
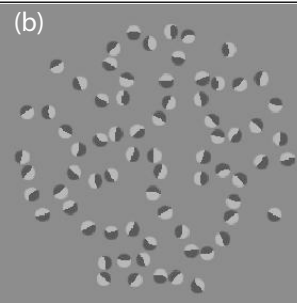
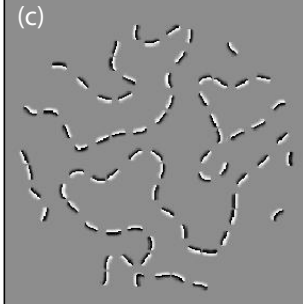
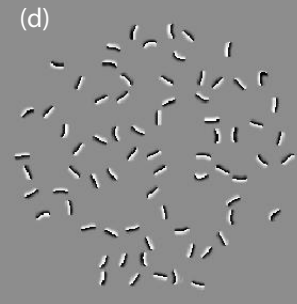
Experiment 2: The role of second-order statistics

Given that removing the low SF information from broadband images did not substantially impair direction discrimination thresholds, I next asked if the “naturalistic” contour structure within our stimuli was promoting the integration of high SF motion signals. It is certainly the case that motion signals can inform observers about the form of objects as shown in slit-motion studies (Nishida, 2004) and studies of spatiotemporal boundary formation (e.g. Shipley & Kellman, 1993) but much less work has demonstrated the influence of form on motion processing (Lorenceanu & Alais, 2001; McDermott, Weiss, & Adelson, 2001). To test this hypothesis, I assessed the impact of disrupting the second-order motion/orientation statistics of our stimuli by placing apertures over the stimuli (Figure 4a). Global structure could then be disrupted by randomly switching the signals passing under each aperture with another randomly chosen aperture (Figure 4b,d). Scrambling in this manner across all apertures preserved local signals, but disrupted global structure. Note that breaking global structure in this way disrupts both the second-order statistics and the low SF components of the signal. Therefore the effect of scrambling can only be identified by comparing performance across both the high-pass and broadband stimuli. Thus if motion processing exploits the statistical regularities of second-order structure in naturalistic images, then performance should deteriorate in both the high-pass and broadband conditions as this structure is abolished. Alternatively, if a detriment to performance is observed only in

broadband stimuli, then disruption to the low SFs is driving any observed reduction in performance.

Stimuli

Stimuli were identical to the broadband (Figure 2a) and high-pass (Figure 2c) stimuli of Experiment 1 but were viewed through a mask consisting of a series of circular raised-cosine apertures (radius 16.2 arc min.; tapered region radius 1.38 arc min.). All apertures were positioned within a circular region (radius of 2.95 deg) centred upon the fixation point. The underlying noise carrier translated upwards and each contour passed through the middle of each aperture during the middle frame of the trial (Figure 4f). This arrangement of the apertures and contours rendered the global structure of the stimuli easily apparent to the observer. Further, centring the apertures over the contours reduced between-trial variability that would have resulted from a random placement of the apertures. Due to the random nature of the stimuli the number of apertures varied, with a mean of 86.4 and a standard deviation of 6.8. Scrambling was achieved by swapping the signal under one aperture with that of another randomly chosen aperture. Scrambling in this manner preserved local signals but disrupted global structure.

		Factor 2 - Structure	
		Un-scrambled	Scrambled
Factor 1 - Spatial Frequency	Broadband	(a) 	(b) 
	High-pass	(c) 	(d) 

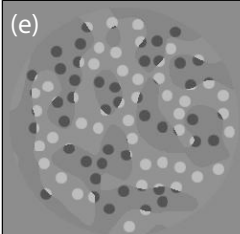
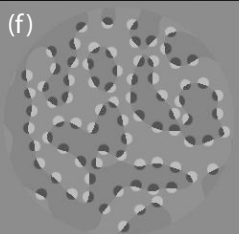
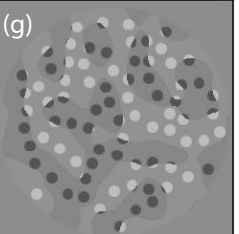
		Example frames		
		First frame	Middle frame	Last frame
(e) 	(f) 	(g) 		

Figure 2-4 Stimuli (a-d) Middle frames of the four conditions used in Experiment 2 (contrast has been maximised to improve visibility). (a) Underlying stimuli were similar to Experiment 1, but were viewed through a series of small stationary apertures that were centred on the contours in the middle frame of the sequence. (b) Global structure was disrupted by randomly swapping the signals viewed behind each aperture. (c,d) Shows a high-pass filtered version of the same image. (e-g) depict the first, middle and last frames of an example broad-band unscrambled trial. For illustration purposes the underlying image is superimposed upon the occluding surface of the apertures. Note that apertures were densely placed over the whole contour structure of the image and that the contour passes through the middle of each aperture during the middle frame (f).

Results

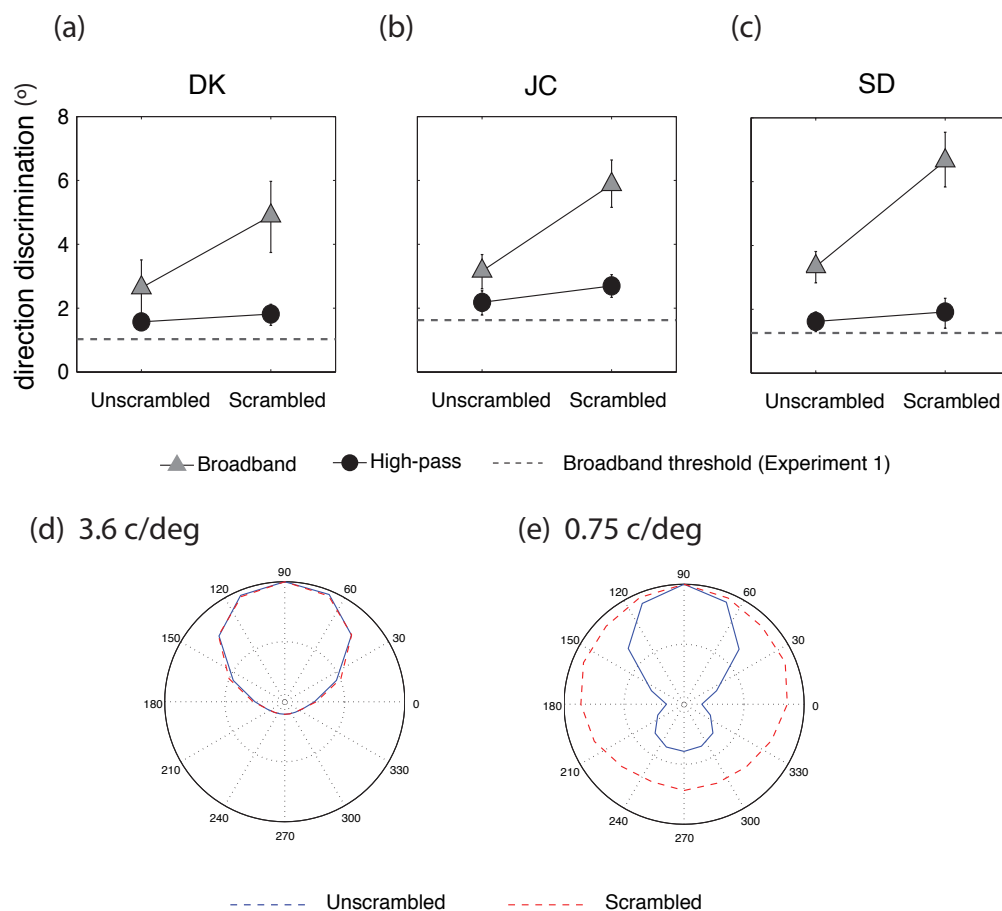


Figure 2-5 Results, experiment II Direction discrimination thresholds measured with locally apertured stimuli for three observers (DK, JC & SD). Dashed lines indicate the mean direction discrimination threshold for each subject for the non-apertured, broadband stimuli from Experiment 1. Thresholds for the broadband stimuli (grey triangles) are always higher than the high-pass (black circles) stimuli. The effect of scrambling is highly significant in the broadband stimuli whilst only a small effect is observed in the high-pass stimuli. This suggests that 'coherent' global structure is not necessary to achieve low discrimination thresholds but that disrupting global structure is detrimental to performance when the low frequencies are present. (d, e) depict the motion energy at 3.6 c/deg and 0.75/cdeg respectively across a channel of V1 neurons tuned to the object speed, note how the distribution of motion energy is identical in (d) but not in (e) highlighting how scrambling dramatically increases the direction bandwidth of the signal at low SFs (see appendix for model details).

Results from Experiment 2 are plotted in [Figure 2-5](#) and show two main effects. First, thresholds for the broadband stimuli (grey triangles) are higher than thresholds obtained without the obscuring apertures (dashed lines). Second, scrambling increased thresholds two-fold across the broadband condition but only had a weak effect on the high-pass stimuli. Interestingly participants reported a percept of rigid translation under all conditions except the broadband scrambled condition where a small amount of spatial incoherence was observed. This pattern of results suggests that the second-order statistics do not significantly influence motion processing in our experiment because scrambling would have predicted an equivalent effect in both the high-pass and broadband signals. Instead, our results are consistent with a global motion mechanism that pools directional information across space and SFs but is insensitive to the relative motion information in nearby locations. In this model, scrambling increases the directional bandwidth at low SF's ([Figure 2-5; d,e](#)) leading to a loss of sensitivity. The weaker effect observed in the high-pass conditions reflects the weak signal in the low SFs (see [Figure 2-2e](#)). Later sections attempt to justify this position further by isolating the low SF component of the signal (Experiment 3) and assessing the variability in the signal through a model of V1 neurons (see Model).

Experiment 3: Low SFs and the effect of scrambling carrier location

Experiment 3 was designed to probe the role of low SFs in the scrambling effect observed in Experiment 2. This was achieved by progressively attenuating the high SF component of the broadband signal to isolate the low SF component by convolution of the carrier signal with Gaussian filters of progressively larger spatial extent.

Methods

Subjects, procedure and apparatus were identical to Experiment 2. Stimuli were low-pass versions of the broadband stimuli in Experiment 1 from which five low-pass conditions were created by convolving the broadband images with a Gaussian filters set to $\sigma=5.4$ 7.8 11.4 16.2 or 22.2 arc min. After convolution, the contrast for all conditions was set to a root-mean-square contrast of 0.20 (6.0 cd/m²). The five new stimuli were then tested across both the scrambled and unscrambled conditions of Experiment 2 to generate 10 new conditions.

Results

Figure 6 shows the results of Experiment 3, which are in good agreement with the results of Experiment 2. Scrambling induced a twofold increase in thresholds at low levels of stimulus blur ($\sigma=0.09$). To examine the effects of increasing blur, a straight line was fit to the log of thresholds across the scrambled and unscrambled conditions. The exponent of the fit was

recorded and error bars were generated using a bootstrapping procedure with 1024 iterations. The results of the fitting procedure (Figure 6 d-f) show that the exponent is higher in the scrambled condition (significantly so for DK and SD). This means that motion discrimination thresholds increase more quickly with blurring for scrambled than unscrambled conditions. Since increasing the level of blur in the images does not alter the second order statistics I conclude that it is the disruption of low SF components of the signal that is driving the effect of scrambling. An alternative interpretation of the data is that lateral interactions occur over increasing distance with de-creasing SF (e.g. Polat & Sagi, 1993) - given the fixed radius of the display this may lead to an increased impact of lateral interactions with increasing blur.

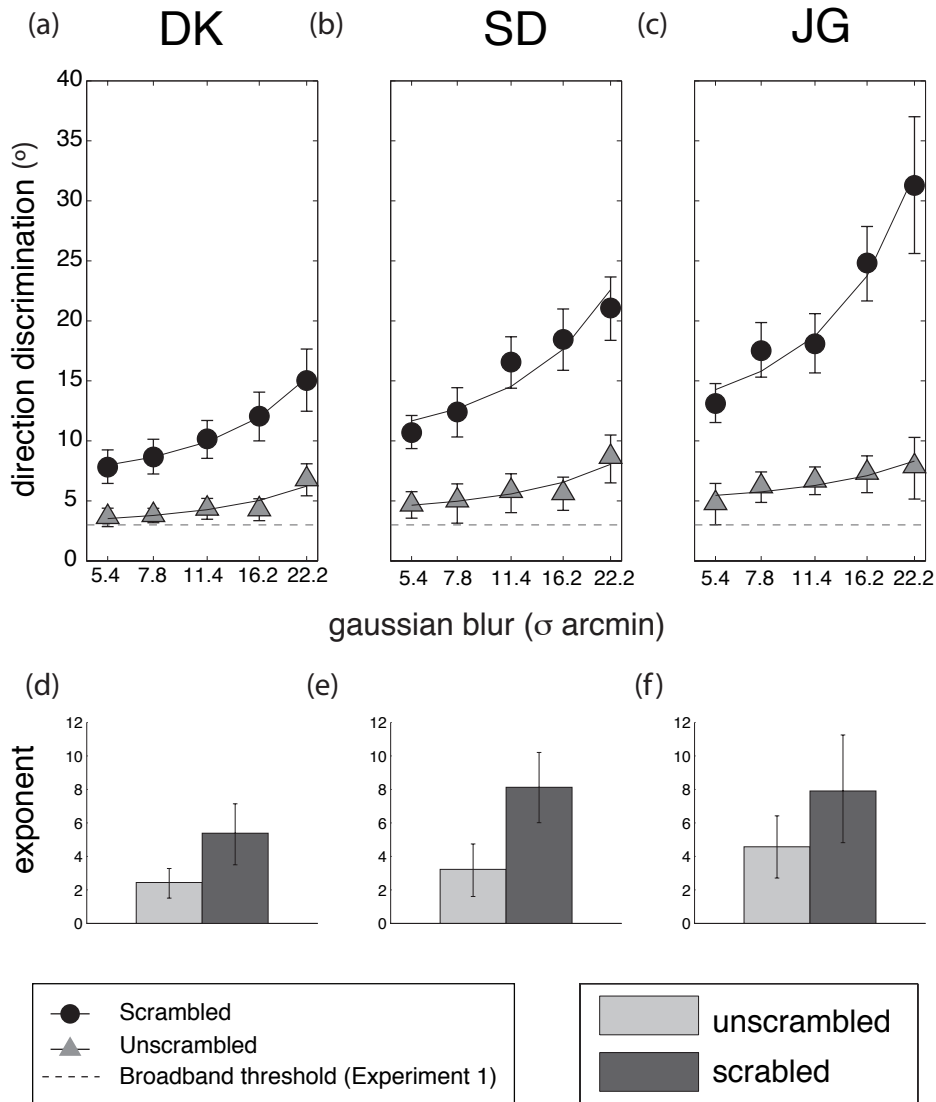


Figure 2-6 Results of Experiment 3 for three observers (DK, SD & JG). Direction discrimination thresholds for scrambled (black circles) and unscrambled (grey triangles) apertured stimuli are shown as a function of the standard deviation of Gaussian blur applied to the underlying contour image. The curves show the line of best fit generated by fitting a straight line to the log of the data, the slope of which is shown in (d - f) for unscrambled (grey bars) and scrambled (black bars) conditions. Error bars show 95% confidence intervals on all graphs. The exponent is always greater in the scrambled condition, (significantly for DK and SD). This suggests that increasing reliance upon the low frequency component is of greater detriment to the scrambled stimuli, further indicating that it is the low-frequency component of the signal rather than the second order statistics that is driving the effect of scrambling.

Controls

The above analysis has implicitly assumed that the psychophysical data is the result of local signals being combined across space to yield a global estimate of direction. However, the stimulus used is theoretically resolvable at the local level. To ascertain what level of disambiguation is being achieved at the local level, I perform a control in which I vary the number of apertures. The control experiment was identical in all regards to the broadband unscrambled condition of Experiment 2, except I vary the area of the image presented to the observer by varying the number of apertures presented from 1, 4, 16 or 32. In all conditions the spatial positioning of the apertures was random but constrained to fall within a radius of 2.95° from fixation. Results are shown in [Figure 2-7](#). Discrimination thresholds improve with increasing aperture number, strongly suggesting that the degree of precision achieved in Experiment 2 could not have resulted from a local analysis alone and that information must have been combined across space. Note that performance in the single aperture condition is better than if the information were truly ambiguous (i.e. straight edges) in which case a simple model which detects the direction orthogonal to an elements orientation will produce discrimination thresholds of around 65° . Thus some level of local disambiguation is being achieved.

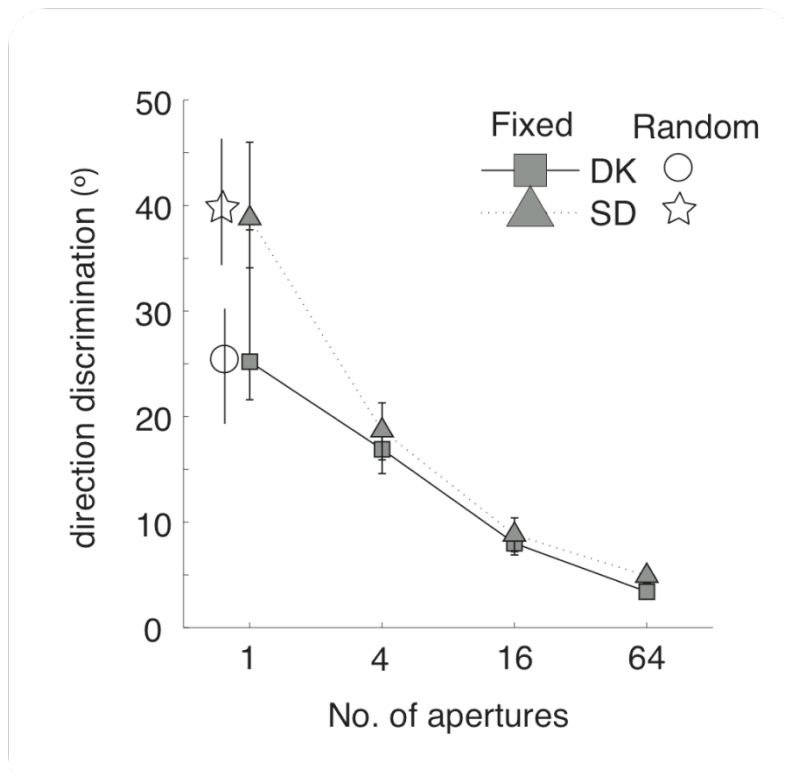


Figure 2-7 Control experiment 1 examined the ability of the observers to locally resolve the information presented in each aperture of Experiment 2. Results demonstrate that performance improves rapidly with increasing aperture number and strongly suggests a global analysis is needed to achieve the level of precision observers achieved in Experiment 2. Closed symbols represents thresholds when random aperture positions. Open symbols denote performance when the aperture position was held constant.

A second criticism is that the second-order statistics of Experiment 2 are only present during the middle frames of our experiment as the apertures largely obscure the contour structure during the beginning and end frames. The criticism is valid because the strength of the second-order relations falls with increasing distance between elements (Geisler, et al., 2001). Since the full contour structure of the stimulus is only exposed during the middle frames of the trial, the mean distance between elements will be larger during the beginning and ends frames thus reducing the strength of the second-order statistics. To address this criticism I repeated Experiment 2 in full, but slowed

down the translation of the underlying carrier to 1 deg/s so that the contour structure was exposed for the full duration of the trial. The results are shown in Figure 8 for subjects DK, JG and SD. Results are consistent with Experiment 2 and reveal no significant difference between the high-pass un-scrambled and scrambled conditions but again reveal that the scrambling significantly lowers the precision of observers in the broadband conditions. It should be noted that performance is worse at the slower carrier speeds of the control experiment, a finding that is expected because a slower carrier speed and identical trial duration will reveal much less of the carrier to the observer.

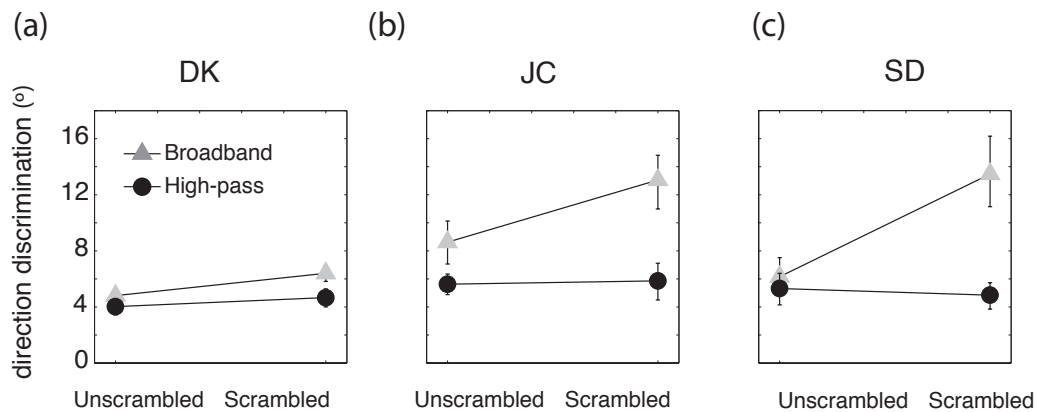


Figure 2-8 Control Experiment 2 repeats Experiment 2 using a slower carrier speed so that the full contour structure is present on each frame. Results follow the same pattern as Experiment 2 with scrambling always causing a significant increase in observers' threshold in the broadband (grey triangles) but not the high-pass condition (black circles).

Discussion

The accurate estimation of motion-direction is trivial for isolated objects containing isotropic orientation structure. Under such conditions the distribution of motion energy is predictable and veridical estimates of the direction of motion can be obtained by simply calculating the centre of

motion energy. However in natural, unconstrained environments this is rarely, if ever the case and biases in motion energy render such a strategy unreliable. The paradigm I have described is able to probe the influence of imbalances in motion energy simply because the stimuli used exhibited anisotropies in the orientation structure that varied randomly from trial-to-trial. In Experiment 2&3 scrambling will induce 'spurious' correlations in the low SF component of the signal (see model), increasing anisotropies in the motion energy and in turn raising psychophysical thresholds.

The lack of an effect of disrupting the second order statistics is surprising considering the importance of second-order statistics in the detection of static (Field, Hayes, & Hess, 1993) and moving contours (Bex, Simmers, & Dakin, 2001; Ledgeway & Hess, 2002, 2006). More directly, this work appears to contradict the findings of Lorenceau and Alais (2001) who show performance on a motion discrimination task is better for 'closed' forms than 'open' forms. Although both studies used very similar paradigms, the stimuli employed differed in terms of their perceptual ambiguity: The class of stimuli employed by Lorenceau and Alais (2001) has been well studied and the percept of global motion is ambiguous and bi-stable (McDermott, et al., 2001) reflecting the potential of such displays to be consistent with more than one physical interpretation (see [Figure 2-1](#)). In contrast, the signal presented in the current paradigm was consistent with only one interpretation. This suggests that global second-order statistics may only influence performance in motion discrimination tasks when there are very high levels of uncertainty in

the binding of spatially disparate elements. The finding also implies that studies of global motion with random second-order orientation statistics (e.g. Kaoru Amano, et al., 2009) are designed to an appropriate level of abstraction.

Although our results suggest no role for the second-order statistics (within one SF channel) like that shown in contour detections paradigms (Field, et al., 1993) the effect of scrambling highlights the importance of the low SF component of motion and how manipulations of spatially disparate elements can dramatically influence the directional signal at this frequency. This observation has implications for a number of other studies using apertured but broadband stimuli (e.g. Lorenceau & Alais, 2001; Mingolla, et al., 1992) where the directional signal of the low-pass component may play an important role.

The rigid integration of the disrupted low SF component observed in our study indicates the motion stream is unable to filter out or 'ignore' SF channels on the basis of a high directional bandwidth in the distribution of motion energy. Although in the present stimuli 'ignoring' the low frequency component of motion would likely improve psychometric thresholds, the relationship between signal bandwidth and reliability is not straightforward. For instance, a broad directional bandwidth is often the hallmark of an unambiguous directional signal (e.g. small dot stimuli) - an observation has been incorporated into the model of Weiss and Adelson (1998) where signals with a broad directional bandwidth are able to constrain estimates of global

judgements to a greater extent than signals with narrow directional bandwidths.

The present work does not distinguish between the predictions of IOC or VA theories, as the stimuli used are essentially Type I. Using the aperture positions of Experiment 2 to restrict the range of orientations presented to the observer may provide a promising route through which this issue may be investigated.

Model

In this section, I explore the interaction between the motion energy model of V1 directionally selective (DS) neurons (Adelson & Bergen, 1985) and the stimuli used in Experiment 2. The theory behind the applications of the motion-energy model is discussed in the introduction and the full battery of DS filters across direction, pseudo-speeds and spatial-frequency is defined as follow;

1. Thirty-two directions evenly spaced around the clock.
2. Thirteen evenly spaced pseudo-speeds from 0% (static) to 150% of the carrier signal speed (3.95 deg/s).
3. Eight SFs from 50% to 700% of the peak SF of the broadband carrier signal (0.75 c/deg) in eight half-octave steps.

The spatial frequency and directional bandwidth of all the model neurons was held constant at 1.5 octaves and 45° (half width and full height)

respectively in keeping with the observed bandwidths of primate area V1 (R. L. De Valois, Yund, & Hepler, 1982; Snowden, Treue, & Andersen, 1992)

The stimuli were accurate reconstructions of trials used in Experiment 2 in terms of the aperture positions and the spatial (256*256) and temporal resolution (26 frames). However, to avoid the artefacts introduced by the horizontal/vertical pixel raster, the direction of motion on each trial was randomised.

Convolution of the signal and sensor took place in the Fourier domain and was inverse-transformed back into the spatial domain. The square root of the sum of the square of the real and imaginary components was taken to represent the motion energy at each point in space for each DS filter, a computation that is formally equivalent to the full rectified square of odd and even phase neurons to generate a phase invariant output (Adelson & Bergen, 1985). A global motion analysis was achieved by collapsing the spatial domain and summing across all DS filters tuned to the same spatiotemporal frequency and direction. Each spatial frequency channel could then be represented as a 2D *Speed-Direction* image, in which the intensity of each region represents the global sum of motion energy across DS filters whose velocity tuning is denoted by the regions position in the image. The only filter normalisation employed was to divide the output of each neuron by the sum of the absolute of the receptive field across space and time; this had the effect of evening out the expected 1/f spatiotemporal

frequency spectrum. No local gain control, normalisation or inhibition occurred between motion filters.

Noise and the sampling rate of neurons were not considered essential to the model output because discrimination thresholds were not derived from the output of the motion filters. Additional factor such as the addition of Poisson noise (e.g. Dakin, et al., 2005a) would have been necessary if direction discrimination thresholds were to be predicted. Further, additional complexity could have been added by varying the bandwidths of the motion filters that simulated V1 neurons as a function of spatial or temporal frequency as both the physiology (e.g. Bair & Movshon, 2004) or psychophysics (e.g. Burr, 1981) would deem necessary, however this would make the resulting motion energy more complex to analyse. For instance, it would be more difficult to ascertain whether the directional bandwidth of the signal was the result of the stimulus or the sensor: By keeping the bandwidth of the sensor fractal across the SF domain and constant across the speed tuning of the sensor, the changes in signal bandwidth across these dimensions could be attributed to the stimulus, not the sensor.

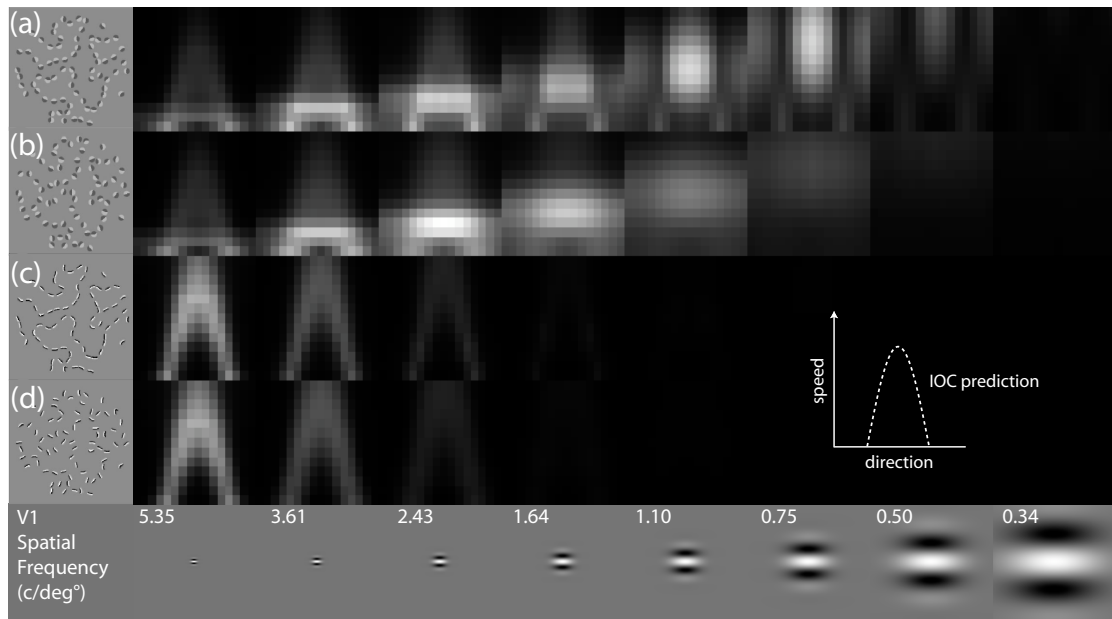


Figure 2-9 Motion-energy (experiment II) 'raw' motion energy plots for the 4 conditions used in Experiment 2. Motion energy is plotted as a function of the pseudo-speed and direction tuning of the model DS filters as illustrated in the inset and previously in the introduction (see [Figure 1-13b](#)). (a); each SF is plotted separately in each column from 'fine' to 'coarse' scale. (a) Motion energy for the band-pass unscrambled condition of Experiment 2, note how the peak of motion energy follows the temporal frequency tuning of the DS filters, not the pseudo-speed tuning (owing to normalisation within, but not across SF) and that the motion energy is centred on the veridical direction only when the SF of the carrier signal and DS filters are matched (0.75 c/deg). (b) Motion energy for the band-pass scrambled condition, note how the directional bandwidth is higher the low SFs relative to the unscrambled condition. (c&d) Motion energy for the high-pass conditions; the motion energy is concentrated in the high-SF channel, and the directional bandwidth is least in the high-pass conditions, reflecting decreased superposition of signals from the lower SF channels.

Model Results

[Figure 2-9](#) reveals the interaction between the stimulus used in Experiment 2 and the motion energy model of V1 directionally sensitive neurons (Adelson & Bergen, 1985). Each row illustrates the averaged motion energy across 256

example trials for one of the four conditions of Experiment 2 (depicted in the leftmost row). The illustrations in each column show the motion energy as a function of the speed and direction of the DS filter within each SF channel. For image clarity, the motion energy across each condition (each row of [Figure 2-9](#)) was normalised between 0-1 and the conditions and sensors are depicted at the same spatial scale in the leftmost column and bottom row respectively. Note, the spatial frequency of the broadband carrier signal and DS filter are matched at 0.75 c/deg.

Initial inspection reveals the motion energy of the high-pass condition to be (unsurprisingly) concentrated in the high SF channels. However the pattern of motion energy in the broadband condition is more complex. To understand the distribution of motion energy in the broadband conditions, it is important to note that the spatial and temporal frequencies are independently coded in many V1 neurons (Foster, et al., 1985; Priebe, et al., 2006; Tolhurst & Movshon, 1975) - when there is a mismatch between the SF of a stimulus and the sensor, the speed tuning of the neuron is lost and the motion energy (in this SF channel) will be greatest when the temporal frequency of the DS filter and the stimulus is matched. For a rigidly translating band-pass (or low-cut) stimulus such as ours, this results in component motion (occurring at slower speeds) only being captured in the high-SF channels in accordance with Equation 1.25.. To highlight this point, Figure 2-10c plots the difference between the temporal frequency tuning of the DS neurons and the peak

temporal frequency tuning of the stimulus. Note how the peak of motion energy in Figure 2-10c closely follows the zero temporal frequency difference.

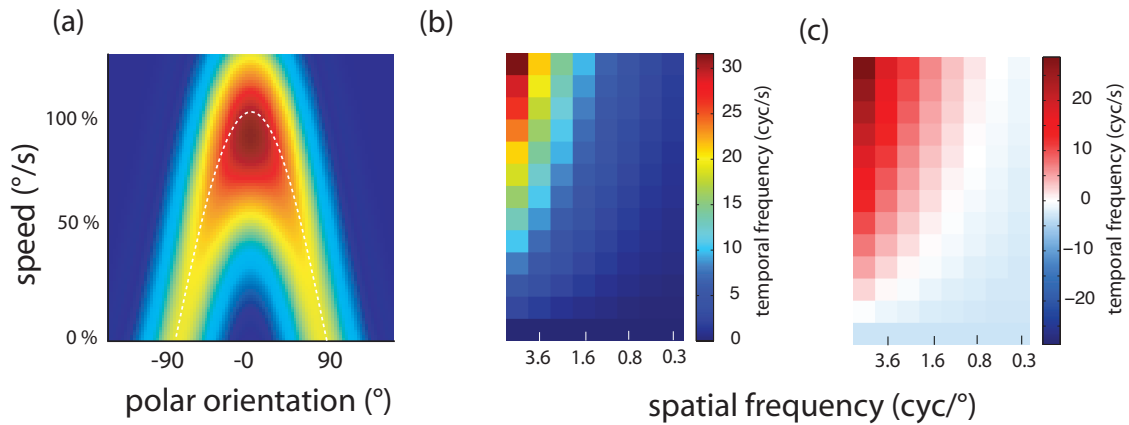


Figure 2-10 (a) hypothetical motion energy of a rigidly translating isotropic stimulus plotted in the speed-direction space used in Figure 11. The x-axis depicts the angular separation between the veridical object direction and the direction tuning of the DS filters whilst the y-axis plots the speed tuning of the DS filters as a percentage of the object speed. (b) Plot of the changing temporal frequencies used as a function of the spatial frequency tuning of the DS filters. (c) The temporal frequency tuning of the DS filters minus the peak temporal frequency of the stimulus. Note that the pattern of motion energy shown in Figure 11 closely follows the peak temporal frequency tuning of the stimulus.

Comparison of Figure 2-9 (a) and (b) shows that the effect of scrambling is to dramatically increase the bandwidth of the signal in the lower SFs indicating that scrambling leads the motion sensors to detect spurious correlations at low SFs. Finally, the directional bandwidths are sharper in the high-pass conditions, reflecting a lower superposition of signals across SF channels. This suggests the low frequency component of the broadband stimuli leads to 'masking' of the high frequencies and provides a plausible explanation for the higher psychophysical thresholds observed in the broadband conditions.

Implications for models of global motion processing

The changing nature of the signal across SF channels highlights the independence of the spatial and temporal tuning of V1 neurons (Foster, et al., 1985; Priebe, et al., 2006; Tolhurst & Movshon, 1975) while simultaneously showing that stimulus variables such as orientation and speed are not independently coded in area V1 (see; Mante & Carandini, 2005). It should be noted that the distinct pattern of motion signals across SF channels is determined by the low-cut SF profile of our stimuli. In contrast, if the stimulus were fractal and isotropic, the full expression of component motion may be found within each SF channel. However in naturally occurring stimuli the SF profile is likely to vary between a broadband and a band pass profile and is unlikely to be isotropic. Accordingly the broadband integration of signals across spatial frequencies observed in global motion studies (Bex & Dakin, 2002; Simoncelli & Heeger, 1998) appears necessary to capture the full expression of component motion (occurring across a range of speeds and orientations) despite the increased vulnerability to noise that such broadband integration brings (Bex & Dakin, 2002).

Experiment 4 Number, Density or Area

In Experiment 3 above, the following variables all co-varied; the number of elements, the density of elements and the region of integration, (I will term these variables *number*, *density* and *area*). This means that one variable cannot be manipulated independently of the other two variables; for instance changing the number of elements within a given area alters the density of stimulus. If an experimenter wants to isolate the influence of each variable (s)he must generate three functions; in each function one variable (either *number*, *density* or *area*) is held constant whilst the other two variables are manipulated. The relative slope of each of the three functions can then be used to ascertain the relative role of each variable. For instance, if the number of elements is held constant and the function remains flat as a function of area then it can be concluding that *number* is the primary determinant of performance.

Dakin et al. (2005a) applied this strategy in combination with an Equivalent Noise paradigm to explore the influence of *number*, *density* and *area* in the integration of moving random-dot patterns. The results showed that the *number* of elements was the primary factor driving performance and revealed only a very minor role for *density* that was attributed to correspondence noise. The equivalent noise analysis demonstrated that the observers' were performing like an ideal-observer whose sampling efficiency was equal to the square root of the total number of samples present. This section asks whether the pattern of data revealed in response to spatial-

frequency band-pass random-dot stimuli holds for the contoured stimuli used in the present experiment. While the integration of a single dot element can be very accurate; a precision of $\sigma \sim 3^\circ$ when embedded in noise (Watamaniuk & McKee, 1998), by contrast the integration of individual contoured elements is relatively poor (precision $\sigma > 25^\circ$, see [Figure 2-7](#)) Accordingly, the need to integrate across space is greater in the contoured stimuli. I propose that observers' difficulty in ascertaining the two-dimensional motion of individual contoured elements stems from the motion stream being unable to resolve the ambiguities associated with the 'aperture problem' and being biased by the orientation content of the stimulus, (i.e. the contoured elements often resemble a straight edge). In this respect, research has suggested that locally 1D and 2D motion signals may be treated differently by the visual system; Amano et al. (2009) who demonstrate that the integration of 2D stimuli (plaids) is constant with an averaging scheme, both in terms of the *direction* and *speed* of motion estimates but that the integration of locally 1D elements (Gabor) was consistent with a more complex integration rule that is able to ascertain both the 2D *speed* and *direction* from locally 1D signals (except for Type II stimulus classes). Moreover the percept of locally 1D elements is demonstrably unstable under some conditions, switching from a global percept to a local percept with decreasing time and the influence of various non-motion parameters such as contrast or the shape of occluders (Lorenceanu & Shiffrar, 1992; McDermott, et al., 2001).

Accordingly I have reason to expect the pattern of results demonstrated in (Dakin, et al., 2005a) may be different in response to stimuli of greater local ambiguity and I apply the co-variance technique documented in (Dakin, et al., 2005a) to the contoured stimuli used thus far in this chapter.

Subjects/Apparatus

Subjects and apparatus were identical to previous experiments in this chapter.

Stimuli

To simultaneously probe area, number and density, 7 conditions were tested. The number of moving elements was either 4, 16 or 64, the radius of the viewing area was either 1.5° , 2.9° or 5.9° and the density was either 0.13, 0.44 or 2.2 apertures per degree squared. The conditions are depicted in the inset of [Figure 2-11](#). The representation of the conditions in a grid allows us to highlight which conditions are consistent with a single dimension being held constant. By looking across the diagonal column one can examine data in which the density of elements was held constant; by looking across the horizontal column one can see data when area is held constant and by looking across the vertical column the data corresponds to when number is held constant. Here I collected data for both *scrambled* and *unscrambled* conditions (described above).

Procedure

A method of constant stimuli was applied; on the first experimental run all conditions began with broad upper and lower bound on the maximum cue size presented ($\pm 64^\circ$) and the function was sampled at 17 points. The thresholds varied considerably and it was necessary to adjust the range as the data was collected; after the first run the upper and lower bounds were adjusted such that the range was extended to ~ 2.5 times the estimated standard deviation.

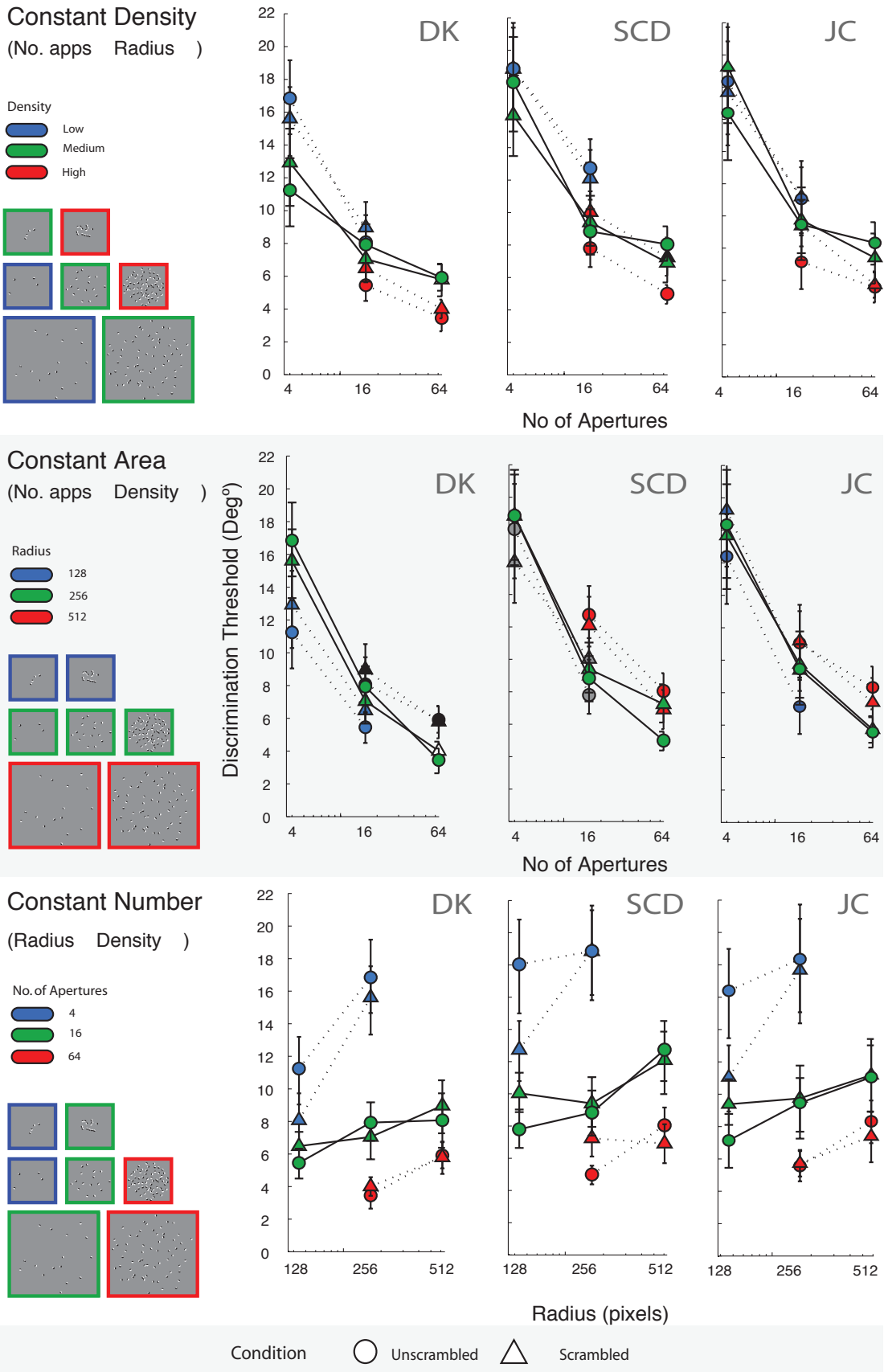


Figure 2-11 Exploring the role of area, number and density. Seven conditions are employed as denoted by insets on the left-hand side. As the factors of area, number and density co-vary the only way to isolate the influence of each factor is to co-vary two factors and hold the other

constant. In row one *density* is held constant, in row two *area* is held constant, in row three *number* is held constant. When number is held constant the changes in threshold are modest indicating that number is the primary determinant of performance, however unlike previous global motion tasks (Dakin, et al., 2005a) the role of density or area is still a strong moderator of performance.

Results

The results of Experiment 4 are shown in [Figure 2-11](#). The flattest function occurs when the number of elements is held constant. This demonstrates that *number* is the primary determinant of performance, consistent with the findings of Dakin et al. (2005a) and Barlow & Tripathy (1997) using band-pass random dot stimuli. When *number* is held constant, the function is the flattest of the three, but there is still a small increase in thresholds with stimulus area; performance gets worse with increasing *area* (i.e. decreasing *density*) demonstrating that either *area* (and therefore the *density*) of elements is an important factor in the integration of contoured elements. This effect is strongest in the four-element condition but also present in the 16 and 64 aperture conditions. The data for the unscrambled and scrambled conditions do not appear to differ substantially.

Discussion

The finding that *density/area* plays a role in the integration of contoured stimuli can be contrasted with the conclusion of previous studies (H. Barlow & Tripathy, 1997; Dakin, et al., 2005a) in which density was considered to play a role due to correspondence noise (i.e. when local elements overlap the number of 'false matches' increases; Qian, Andersen, & Adelson, 1994). Correspondence noise is unlikely to play a role in the present paradigm

because elements do not overlap. In particular the work of Dakin, et al., (2005b) may be compared to the present study because the dimensions of *number*, *density* and *area* were covaried in their study; that work demonstrated a small decrease in both sampling efficiency and internal noise (as inferred from an equivalent noise analysis) when *number* was held constant and the area of integration was increased. The overall effect on performance is very small as the two effects move in opposite directions. In the present work, I do not examine the data in terms of sampling efficiency and internal noise but in terms of the absolute thresholds of observers and it is found that thresholds increase with increasing *area*. This indicates a facilitatory role of *density* in our paradigm and our stimulus, although it is unclear whether this effect is due to a decrease in internal noise or an increase in sampling efficiency and it is not clear whether the effect is mediated by the low-frequencies in the stimulus. However, given that sampling efficiency was found to decrease in Dakin, et al., (2005b) and correspondence noise cannot play a role in our current paradigm because elements never overlap (e.g. Qian, et al., 1994) it may be that density improves the sampling efficiency of both stimulus classes. To provide a firmer test, a paradigm is needed in which the orientation bandwidth of the signal may be smoothly varied to produce a continuous transition between a 1D and 2D motion signal. I suggest that an adaptation of the Global-Gabor array in which the sinusoidal carrier is replaced with a rigidly translating band-pass spatial-frequency noise stimulus in which the orientation bandwidth may be manipulated in the Fourier domain would be appropriate. The paradigm employed by Dakin, et al., (2005b) can then

be used to infer the sampling efficiency and internal noise across the *number*, *density* and *area* functions, i.e. equivalent noise functions (based on discrimination thresholds) should be generated for each of the seven conditions used in the present experiment and for low, medium and high orientation variance band-pass filtered noise stimuli.

Recent research in neurophysiology suggests that an effect of density may be mediated by the response properties of MT pattern selective cells; Majaj, Carandini & Movshon (2007) demonstrate that when the component gratings are separated in space (but still constrained within the receptive field of an MT cell), the MT cell will respond to the component motion rather than the pattern motion. Thus the 'pattern' selectivity of such neurons is contingent upon the proximity or degree of overlap between the component gratings. The data from Majaj et al. (2007) are however insufficient to determine whether the 1D velocities need to be locally overlapping or simply closer in space to achieve 'pattern' selectivity, but an extension of the paradigm did reveal that as the number and density of the pseudo-plaids was increased (but not overlapping) 'pattern selectivity' returned (M. Jazayeri & A. J. Movshon, 2007), in support of the notion that MT 'pattern' cells may provide the neurological site for the effect of density noted.

(3) Experimental Chapter No.2

The aperture problem in natural scenes

The majority of studies probing the 'aperture problem' have used highly constrained stimulus classes often containing just two orientations. In contrast natural scenes contain a variety of different textures, end points and contours. The purpose of the present study is to reveal which components of natural images drive subjects' performance in a motion task. To this end, I introduce a novel variant of the image classification paradigm (Eckstein & Ahumada, 2002; Gosselin & Schyns, 2001). Broadly speaking, the aim of the classification image paradigm is to identify which aspects of a stimulus drive performance on a particular task. Such techniques work on the principle that if the information in a scene is important to the task at hand, then degrading the information through the application of additive (reverse correlation) (Eckstein & Ahumada, 2002) or multiplicative (Bubbles) (Gosselin & Schyns, 2001) noise will impair performance. Image classification techniques sum all the noise fields weighted by the observer's responses, an operation that is formally equivalent to performing a reverse correlation procedure (Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005), to generate a "perceptive field" that maps the relationship each part of the stimulus to the observers' response.

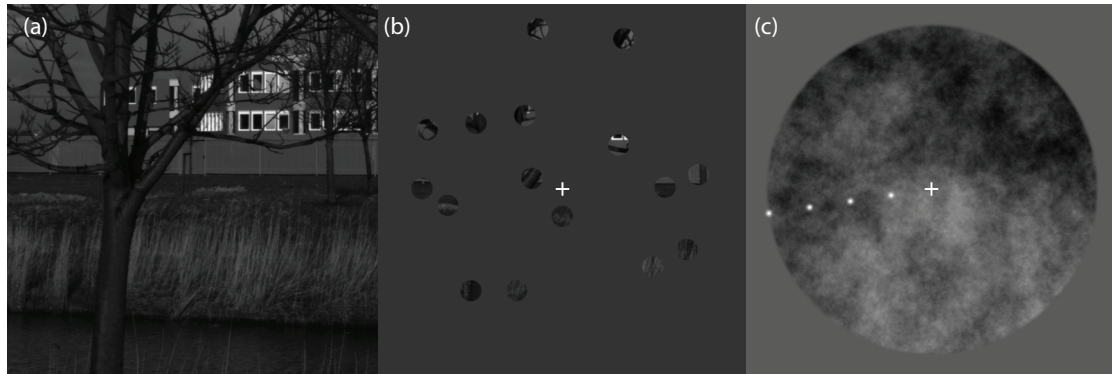


Figure 3-1 I measured subjects' ability to estimate the direction of motion of rigidly translating natural stimuli viewed through 16 apertures. (a) A linear greyscale natural image from the Van Hateren (van Hateren and van der Schaaf 1998) image set (b) A sample frame from the movie stimulus presented to subjects. (c) The test phase. Observers manipulated the orientation of a line composed of 4 Gaussian patches that radiated from the centre of the display (i.e. at fixation) to the edge of the potential viewing area until it matched the perceived direction of the translating natural scene. A phase-randomized version of the stimulus was presented during the test phase and between trials to mask the structure and onset/offset of the natural image.

The aim of the present work is to identify which features in natural images influence subjects' perceived direction of motion. As discussed previous studies have shown that the apparent global direction of motion depends on the local orientations present in the stimulus. Natural scenes have much broader orientation content than many artificial stimuli used to probe the 'aperture problem' and it is not known how naturally occurring textures and contours influence observers' perception of motion. In this chapter I introduce a novel variant of the reverse correlation paradigm that allows one to examine how natural scene statistics affect the apparent direction of motion.

The reverse-correlation paradigm used in the present study differs from the reverse-correlation paradigms described above: In this work subjects viewed a natural image that was rigidly translated in a random direction on each trial (Figure 3-1a). The translating natural image was viewed through an opaque

mask, punctured by 16 randomly positioned apertures (Figure 3-1b) and the observer's task was to indicate the direction of perceived motion using a method of adjustment (Figure 3-1b). On each trial a continuous error signal was generated (the angular separation between the reported direction of motion and real direction of motion). A reverse-correlation paradigm was then used to relate observers' distributions of errors to the underlying stimulus statistics. Observer errors were analyzed as a function of (1) the absolute direction of motion, (2) the orientation structure of the natural scene exposed to the observer on each trial (specifically the mean orientation and the orientation variance of the natural scenes viewed through each aperture on each trial) and (3) as a function orientation structure of aperture pairings.

Methods

Psychophysics

Subjects

The procedures complied with the tenets of the Declaration of Helsinki. Three psychophysically experienced observers (DK, SD, JG) each with normal or corrected-to-normal vision took part in all experiments.

Apparatus

Stimuli were generated on an Apple iMac computer running MATLAB (MathWorks) using functions from the Psychtoolbox (Brainard, 1997; Pelli, 1997). Stimuli were displayed on a Dell, Trinitron CRT with a spatial and temporal resolution of 1080 X 768 pixels and 85 Hz respectively. The display

was viewed at a distance of 97cm such that 64 pixels subtended 1 degree of visual angle. The video signal from the computer's graphics card was first passed through an attenuator (Pelli & Zhang, 1991) and was amplified and copied (using a line-splitter) to the three guns of the monitor to give a pseudo 12-bit monochrome image. Monitor linearization was achieved by recording the relationship between the signal and the monitor luminance (measured using a Minolta LS 110 photometer), to create a linearization lookup table that was passed to the Psychtoolbox internal colour lookup table.

Stimuli

Stimuli were natural images selected from the linear Van Hateren ".iml" image set (van Hateren & van der Schaaf, 1998). The mean luminance of the stimuli was 40 cd/m² and the root-mean-square contrast of the image prior to occlusion was fixed at 0.20. No local contrast normalization procedure was used. The native resolution of the Van Hateren images is 1536*1024 pixels; images were presented at this resolution. Due to the use of apertures, only a subset of the full image was ever presented - a region contained within a radius of 256 pixels (4°) from the centre of the original image.

Motion was generated using operations built in to the computer's graphics card (NVIDIA GeForce accessed via OpenGL) that allowed for sub-pixel resolution via linear interpolation. On each trial, a full size image was passed to the graphics card buffer. By shifting the source coordinates of the image on each frame of the movie, a percept of rigid translation was generated.

The speed of translation was one pixel-per-frame and lasted 32 frames, corresponding to a speed of $1.33^\circ/\text{sec}$, a total distance of 0.5° and a duration of 0.3765 seconds. During each trial/movie the centre of the image was constrained to pass through the point of fixation on the middle frame of each movie. Between trials, a static, phase-scrambled version of the natural scene was placed within the viewing area to mask the onset and offset of the movie stimulus to mask the presence of after-images and to maintain a fixed display contrast. The observer's response initiated the next trial.

The translating natural scene was viewed through 16 apertures each with a radius of 0.25° and whose edges were smoothed with a raised cosine over 0.05 arcmin. The apertures were presented at random locations (but avoiding overlaps) within a 4° radius from the point of central fixation (figure 1b). Thus during each frame 16% of the full area was visible to the subject.

Procedure

On each trial the underlying natural image was translated in a random direction (0° - 360°). After presentation of the stimulus movie a mask image appeared. Subjects then indicated the perceived direction of the movie they had just seen by manipulating the orientation of a probe. The probe was constructed from four evenly spaced 2D Gaussian elements that radiated from the fixation point to the circumference of the global aperture (Figure 1c). Subjects took as long as required to manipulate the probe (using the computer's mouse) until it was aligned with the perceived direction of

motion. Subjects were asked to maintain fixation at all times upon a dot presented in the middle of the stimulus.

Conditions

Two images were used in the study (No. 44 & 206 of the Van Hateren set) as shown in [Figure 3-2](#) (a&b)

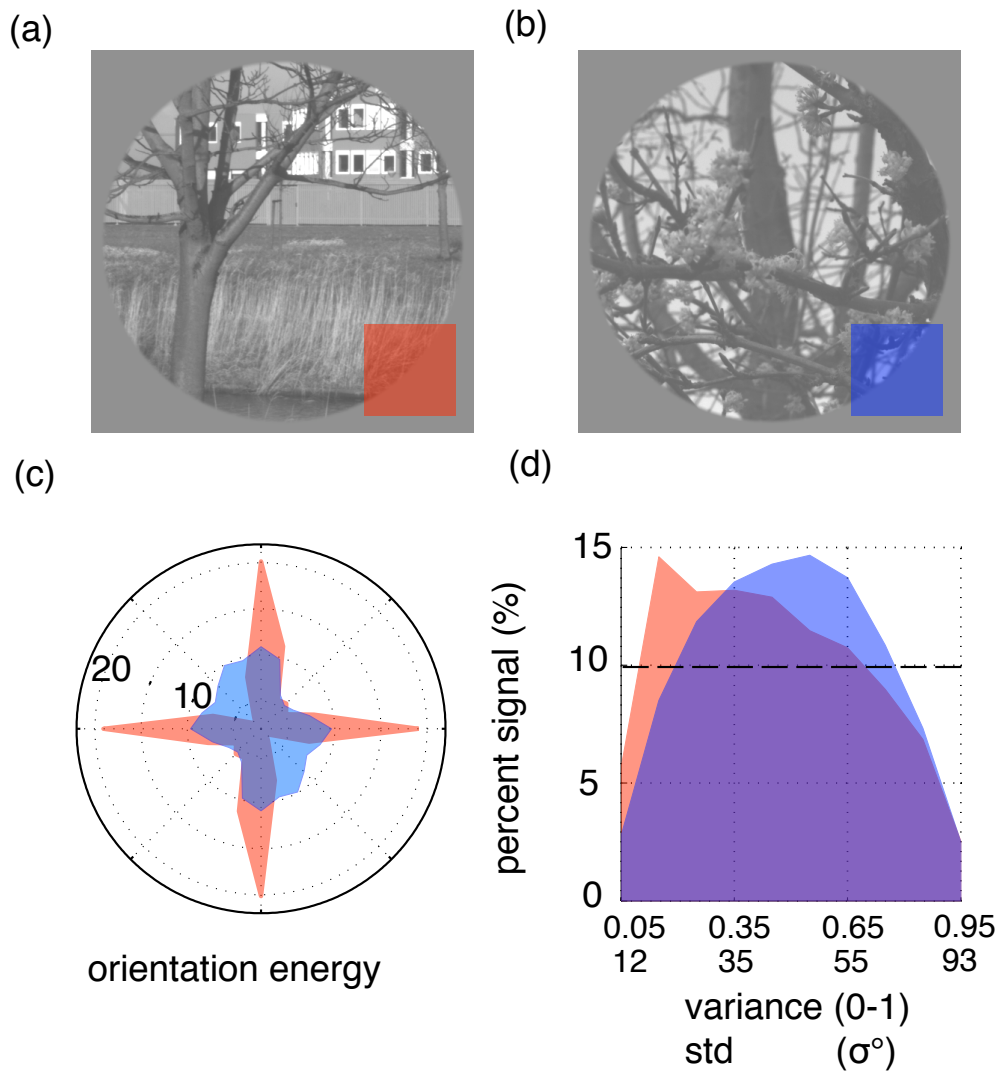


Figure 3-2 (a) image 44 (b) image 206 of the Van Hateren image set (c) Orientation energy as a function of the absolute orientation (see test for details). (d) The percent of pixels with a specified circular variance. Ten circular variance bins were used between 0 and 1, as such the expected number of pixels within each bin would be 10%.

Like most natural scenes these images contained anisotropic orientation structure with relatively greater energy on the cardinal (horizontal and vertical) axes (see [Figure 3-2c](#)) (Switkes, Mayer, & Sloan, 1978). It is possible that these image-based anisotropies or sensitivity based anisotropies (Campbell, Kulikowski, & Levinson, 1966) could affect performance as a function of the direction of motion. To examine this question, the images were either translated at their original orientation, or were randomly rotated between 0°-360° prior to translation.

Subjects DK and JG completed at least 3000 trials on each condition, whilst subject SCD completed at least 3000 trials for both images, but not the random rotation conditions. In total we ran more than 34,000 trials.

Observers' error

On each trial, the signed angular separation between the real direction of motion θ_{2D} and the perceived direction θ_{per} was calculated using Equation 3.1. Negative and positive angular separations denote errors in the perceived direction that are respectively, clockwise and anticlockwise of the true direction of motion.

$$\theta_{err} = \tan^{-1}(\sin(\theta_{2D} - \theta_{per}), \cos(\theta_{2D} - \theta_{per}))$$

Equation 3.1

In each section of the results, error histograms were compiled. To do so, errors between -90 and +90° were binned at one degree intervals (errors greater

than $|90^\circ|$ were excluded). To relate observers' errors to the stimulus, separate histograms were compiled and the input to each histogram was weighted according to the presence or absence of particular stimulus features. The calculations used to do this are described at the beginning of each results section.

Once compiled, the mean and variance of each histogram was calculated and used as estimates of observers' bias and precision. The mean error $\bar{\theta}_{err}$ was calculated using the four quadrant arctangent of the sum of the weighted sine's and cosines (Equation 3.2) where θ represents the error of each bin and W_θ the weighting given to each error bin.

$$\bar{\theta}_{err} = \text{atan2}\left(\sum_{\theta} \sin(\theta)W_{\theta}, \sum_{\theta} \cos(\theta)W_{\theta}\right)$$

Equation 3.2

The variance V_{err} in each error histogram was then calculated using Equation 3.3 & Equation 3.4.

$$R^2 = \frac{\sum_{\theta} (\sin(\theta)W_{\theta})^2 + \sum_{\theta} (\cos(\theta)W_{\theta})^2}{\sum_{\theta} W_{\theta}^2}$$

Equation 3.3

$$V_{err} = 1 - R$$

Equation 3.4

The variance term V_{err} (between 0-1) is then converted into a more conventional circular standard deviation σ term (Mardia & Jupp, 1972).

$$\sigma_{err} = \sqrt{-2\ln(1 - V_{err})}$$

Equation 3.5

Bootstrapping

The estimates of observers' bias and precision reported throughout the paper are plotted with 95% confidence intervals. The confidence intervals were estimated using a bootstrapping operation: We assumed that each trial was independent, and 1024 bootstrapped data sets were compiled by re-sampling (with replacement) from the total number of trials. For each re-sampled data set the error histogram were recompiled and the bias and precision of observers' recalculated to generate 1024 estimates. The estimates were sorted from low to high and the 26th and 998th estimates were used as the upper and lower 95% confidence intervals.

Results

Absolute direction of motion

Data Analysis

In this section we relate observers' performance to the absolute (2D) direction of motion. To do so, separate histograms of observers' errors are generated as a function of the absolute direction of motion, at one-degree intervals

between 0-360°. The error signal on each trial is entered into every histogram, but is weighted by a circular/wrapped Gaussian ($\sigma=6^\circ$) function of the angular separation between the 2D direction (on each trial) and the histogram's direction tuning. Once the error histograms have been compiled, the mean and standard deviation of each histogram are taken as estimates of observers' bias and precision (column two and three, Figure 3-3). An analogous procedure is used to compile the number of reported and presented directions, and the ratio of the reported to presented directions is shown in column 1 of Figure 3-3. A bootstrapping procedure is used to generate 95% confidence intervals (as described in the methods section).

Results

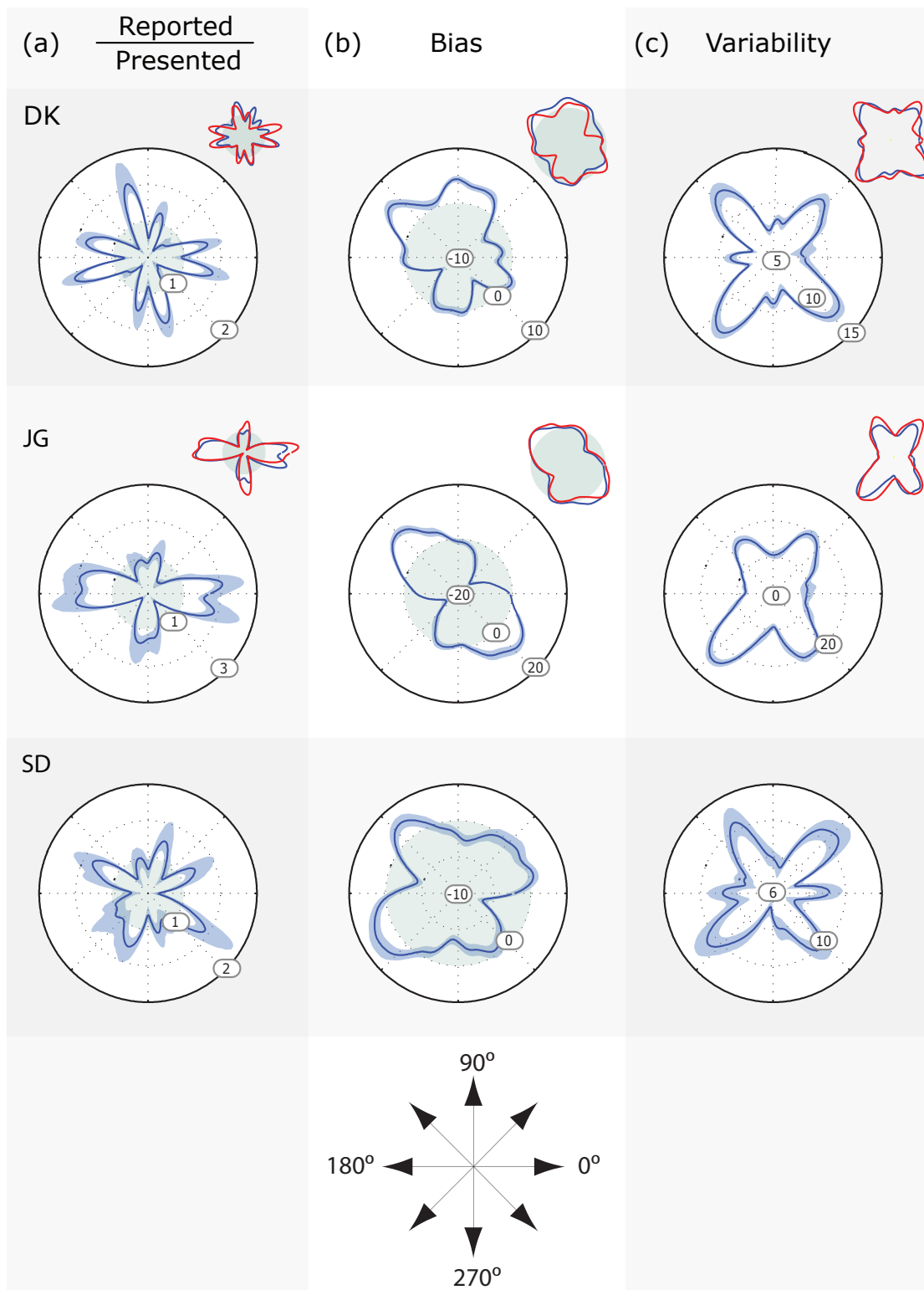


Figure 3-3 Analysis of the reported 2D direction as a function of the presented 2D direction. Data for each subject are presented separately in each row. In column one the ratio of reported to presented directions is shown as a function of the presented direction. In column two and three observers' bias and variability are shown as a function of the 2D direction. The green regions in column one denoted the expected ratio of reported to presented directions (one). The green region in column two denotes a bias of zero. The insets show the pattern of direction estimates for the canonically oriented natural scenes (blue) and the randomly

rotated natural scenes (red). The results are highly anisotropic; this pattern is not due to the stimulus anisotropies as the pattern is similar across both the canonically and randomly oriented conditions.

The ratio of reported to presented directions is plotted in the first column of Figure 3-3. The results demonstrate that all three observers' infrequently report oblique directions (45° , 135° , 225° and 315°). There is also a smaller dip in the frequency that cardinal directions are reported. Intuitively the data appear to reflect two effects previously noted in the literature, one pushing responses away from the cardinal directions towards the oblique directions (Rauber & Treue, 1998) and a second larger effect that pushes responses away from the oblique directions, towards the cardinals (Loffler & Orbach, 2001).

The second column of Figure 3-3 shows bias as a function of the direction of motion. The pattern of bias is idiosyncratic, but stable for each observer. The pattern of bias is nearly identical for the canonical and randomly oriented conditions (blue and red lines; inset). This demonstrates that the pattern of bias as a function of direction is not stimulus led, but is an internal function of each observer. It is not clear what factors may cause the biases in perception, but it is worth noting that experimental procedures that seek to measure bias for specific absolute directions may be confounded by observers' idiosyncrasies. It is for this reason that we use random directions in this reverse correlation experiment and collapse across the dimension of absolute direction when computing observers' response statistics in the next two sections of this thesis.

The third column of Figure 3-3 depicts precision as a function of the direction of motion. The 'oblique effect' is a loss in precision around the oblique angles (45°, 135°, 225° or 315°) (Dakin, et al., 2005a; Gros, et al., 1998) and the oblique effect is clearly present in the current data. A weaker effect is also present in the data with subjects DK and SD exhibiting a small decrease in precision around the cardinals. This effect is consistent with subjects being unwilling to report cardinal directions – an effect that would normally manifest itself as an increase in the precision of a discrimination task that utilized a cardinal direction as a decision boundary (M. Jazayeri & J. A. Movshon, 2007).

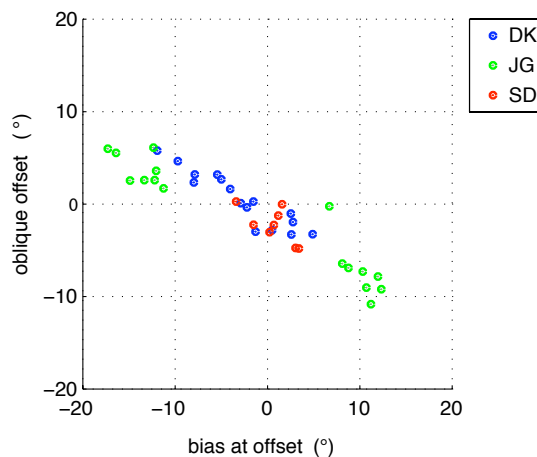


Figure 3-4 Scatter plot of the centre of mass of each quadrant of observers' precision against the bias measured at this angle. Results show a negative correlation ($R = -0.9$) indicating that the oblique effect is to be found in the perceived direction of motion, not the actual direction of motion.

In the present data, the "oblique" is not always centred upon the oblique directions. To examine whether observers' idiosyncratic biases influences the

location of the oblique effect, we first estimated the location of the oblique effect in each quadrant. This was achieved by taking the centre of mass of each quadrant of the variability statistics. This estimate was then subtracted from the nearest oblique direction (i.e. 45°, 135°, 225° or 315°), to estimate the extent that the oblique effect was offset from the true oblique directions. The oblique offset was then paired with the bias statistic (Figure 3-3, column 2) at the estimated location of the oblique effect. This process was repeated for each quadrant, for each condition and for each subject, to generate 40 offset-bias pairings. Figure 3-4 shows a scatter plot of bias versus oblique offset and reveals a strong negative relationship ($R = -0.952$, $p < 0.0001$). The near one-to-one relationship between the pairings demonstrates that it is the reported direction, not the actual direction that determines where observers' responses are most variable. Thus the bias in direction estimates around oblique directions depends on the reported and not the physical direction. This mirrors earlier findings for elevated thresholds around perceived (not physical) oblique directions and orientations (Heeley & Buchanan-Smith, 1992; Meng & Qian, 2005).

Scene Statistics

In the next two results sections we examined observers' errors as a function of the exposed orientation statistics of the natural scenes. The aim was twofold; firstly, we wanted to examine what the impact of orientation variance was on performance; To elaborate, the majority of studies probing motion perception use either locally ambiguous stimulus (e.g. translating bars), or locally unambiguous motion stimuli (e.g. translating dots). By examining observers'

errors as a function of the orientation variance of the exposed natural scenes, we can examine the relative impact of naturally occurring texture and edges upon observers' ability to compute 2D motion. Secondly, we wanted to examine the impact of the orientation of each element, relative to the 2D direction of motion. In a theoretical sense, only two differently oriented surfaces are required to compute 2D motion and it should not matter what the orientations on the components are. However, the literature on the 'aperture problem' clearly demonstrates that observers' are unable to correctly compute 2D motion under a variety of conditions and that this inability is linked to the orientation content of the stimuli (Kaoru Amano, et al., 2009; Bowns, 1996; Burke & Wenderoth, 1993; Loffler & Orbach, 2001; Mingolla, et al., 1992; Yo & Wilson, 1992). Accordingly, we wanted to examine the impact of the orientation of naturally occurring contours on observers' ability to compute 2D motion and to establish the capacity of the motion stream to overcome the 'aperture problem' given the heterogeneous orientation structure of natural scenes. Note, we are not interested in the absolute orientation of the natural scene elements, but the orientation of each element, relative to the 2D direction (as defined in the introduction).

Unlike the majority of studies probing the 'aperture problem', the exact orientation content of the scene was not under direct experimental control and was estimated using a biologically inspired model of orientation processing; Specifically, the two Van Hateren images used in the present study were convolved in the Fourier domain with a bank of log-Gabor filters tuned to 12 evenly spaced orientations between the polar orientations 0° and 165° (Figure 3-5 a&b). The peak spatial frequency of the log-Gabor's were 5.333 cycles per degree with a bandwidth of 0.65 octaves (ratio between the

centre spatial-frequency and the standard deviation of the log-Gaussian function) and their orientation bandwidth was 22.6° (half-width at half-height). The scene statistics were then computed on a pixel-by-pixel basis by taking the *sum*, *mean* and *variance* of the filter responses (Appendix; Equation 3.10, Equation 3.11 & Equation 3.12). The resulting *sum*, *mean*, and *variance* for image 44 of the Van Hateren image set is shown in [Figure 3-5](#) d, e & f, respectively.

We were not interested in the absolute orientation of each element; instead we were interested in the orientation of each element relative to the 2D direction of motion. Accordingly, on each trial, the mean orientation of each pixel was converted into a relative orientation term (Appendix; Equation 3.13); where relative orientation is the angular separation between the mean orientation of a pixel and the 2D direction, across the half circle. The relative orientation term was a number between $\pm 90^\circ$, where 0° denotes angles parallel to motion, $\pm 45^\circ$ angles oblique to the 2D direction, and $\pm 90^\circ$ angles orthogonal to motion. This metric is pictorially represented in the results section, via a standardised 2D direction (red arrow) a black line, oriented relative to the red arrow and a blue arrow denoting the 1D velocity stemming from the oriented element.

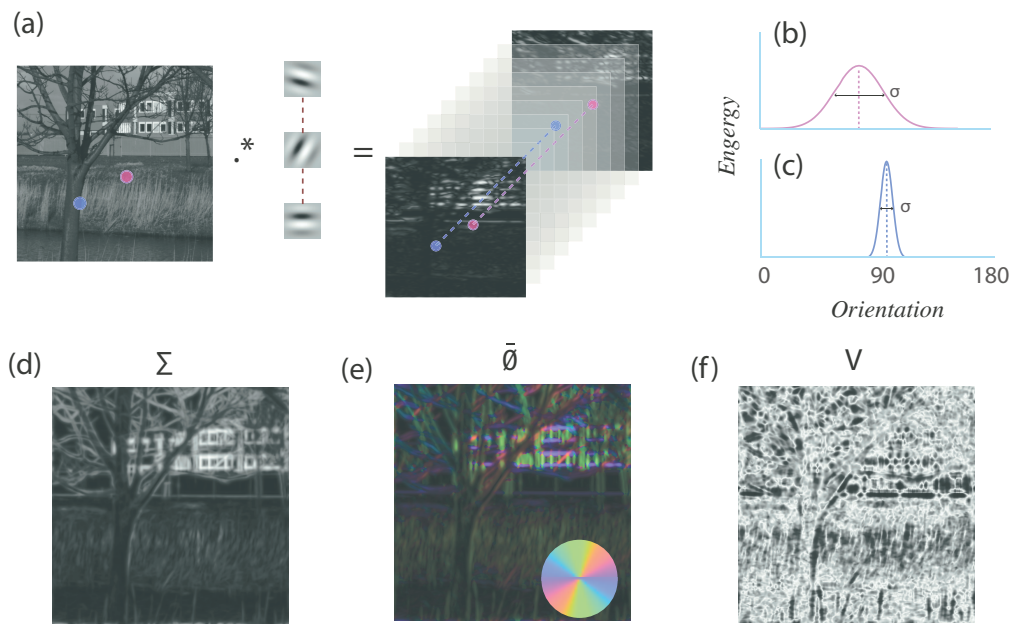


Figure 3-5 (a) A linear greyscale natural image from the Van Hateren (van Hateren and van der Schaaf 1998) image set was convolved with a set of log-Gabor tuned to one of 16 orientations. (b, c) example orientation energy for the pink and blue spatial regions in (a). The distribution of orientation energy at each pixel was classified in terms of (d) the sum of the energy across orientations (e) the mean orientation and (f) the orientation variance

Results

Relative orientation and orientation variance

Data Analysis

In this section, observers' errors were related to the underlying orientation statistics of the exposed natural scenes. To do so, the orientation statistics of the natural scenes were estimated (as described in the preceding section Scene Statistics). Specifically, the pixels exposed by the apertures on each trial, were assigned a relative-orientation and an orientation variance term. To relate performance to the orientation statistics, a number of error histograms were compiled. Each histogram was tuned to both the orientation variance and relative-orientation of a pixel: The relative orientation space was finely sampled between -90° to 89° at one degree intervals, whilst the orientation variance space was crudely sampled with three bins corresponding to low, medium and high, orientation variances. On each trial, each exposed pixel was binned in the error histogram corresponding to its conjoint relative orientation and orientation variance; importantly, the weight given to each error signal, corresponded to the sum of the orientation energy at that pixel. In this manner heterogeneous populations of errors were compiled that related to the expose natural image on each trial. Finally, a smoothing operation was applied across the relative orientation dimension ($=6^\circ$), before the mean and variability of the error histograms were used to estimate of the observers' bias and precision.

Results

Absolute direction of motion

Figure 3-6 plots response bias (column 1) and precision (column 2) as a function of the relative-orientation of each exposed pixel, where the relative orientation is the angular separation between the mean orientation of a pixel and the 2D direction. The abscissa denotes this function; the red arrow indicates a standardised 2D motion vector, the black line denotes the relative orientation of an element and the blue arrows denote the local (1D) direction of motion orthogonal to each contour orientation. Data is plotted separately for regions of high orientation variance (textures; blue line), medium orientation variance (green line) or low orientation variance (edges; red line). The pattern of data for edges (red lines) shows that observers' errors are modulated by the orientation content of the stimulus; observers' are more precise than average when the orientation of edges is orthogonal or parallel to the 2D motion vector. In contrast, when the orientation of contour elements is oblique to the 2D motion vector, observers' are less precise than average and are more biased. This effect, which I term the *relative-oblique* effect, is modulated by orientation variance and is absent for textured regions (blue lines). The results indicate that observers' suffer from the 'aperture problem' in natural scenes with local orientations oblique to the global (2D) direction of motion generating biases of between 2-5° and increasing variability by 20%-25% relative to local orientations that are orthogonal or parallel to the 2D motion vector.

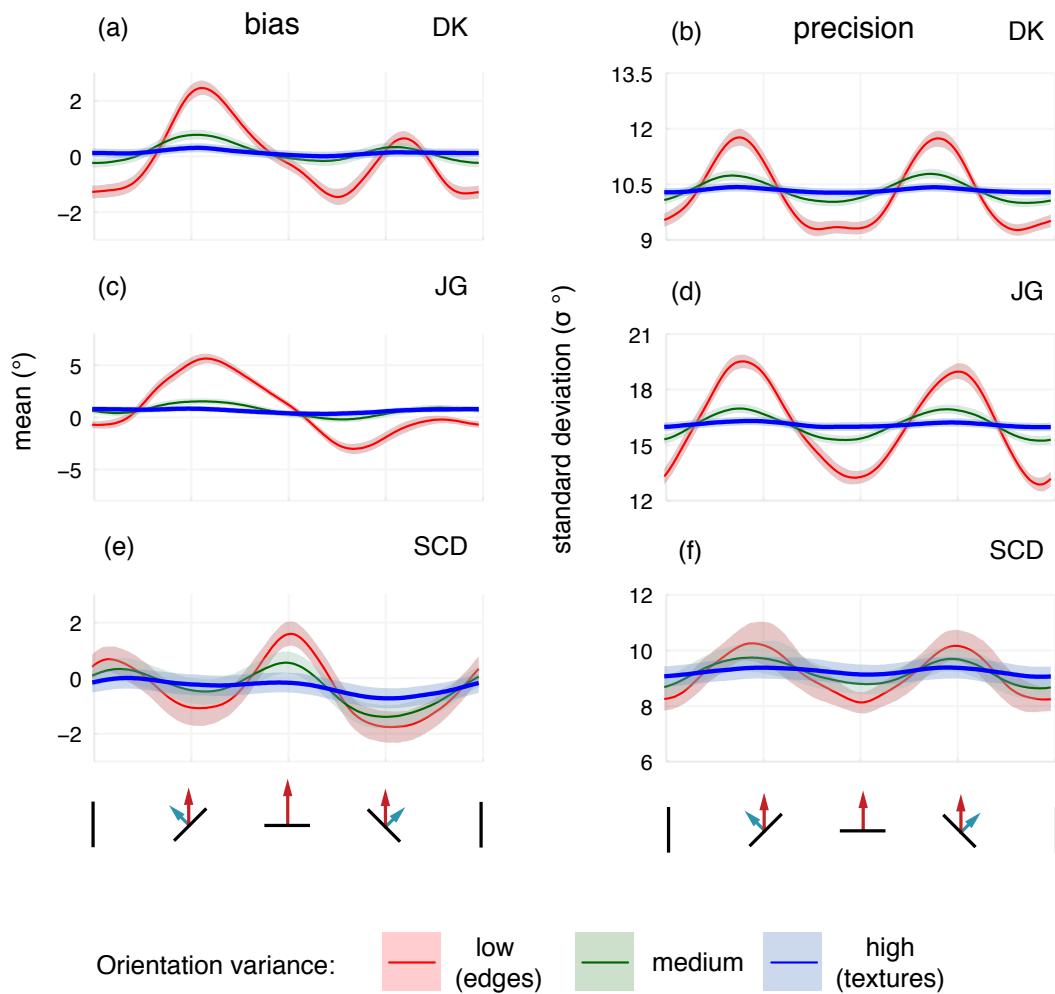


Figure 3-6 Bias and variability as a function of the relative orientation of pixels in the exposed areas for 3 observers (DK, JG and SD). Areas of high local orientation variance (Blue regions) are relatively flat across the dimension of relative orientation; in contrast, areas of low orientation variance (red data) exhibit a periodic dependence on the relative orientation of the pixels presented. The data for intermediate levels of orientation variance (green data) lie between these points. Typically the bias is orthogonal to the direction of motion but there is some idiosyncrasy in the pattern of bias data. In contrast observers' pattern of precision is stable across all observers: Precision is low when pixels are oriented obliquely to the direction of motion but high when pixels are oriented orthogonal or parallel to motion.

Results

Second-order orientation statistics

Data Analysis

In the preceding section we examined observers' errors as a function of the first order orientation statistics of the natural scenes. In this section we will examine observers' errors as a function of the conjoint relative orientation statistics of aperture pairings (second-order orientation statistics). To do so, the orientation statistics were not processed on a pixel-by-pixel basis; instead the orientation energy was collapsed across all pixels that pass under an individual aperture, during each trial, before calculating the orientation statistics for the aperture (to reduce computational costs). On each trial there were 128 unique aperture pairings and observers' errors were compiled as a function of the conjoint relative orientation of each aperture pairing. Unlike the preceding section, observers' errors were weighted by the orientation variance of each aperture, not the sum of the orientation energy. This procedure allowed us to use all aperture pairings, but reduced the impact of high orientation variance patches for which the mean orientation statistic is less meaningful. In total 179² histograms were compiled (179 across each relative orientation dimension, again corresponding the relative orientations between -90 to 89° at one-degree intervals). Finally, a two dimension Gaussian function ($\sigma_{x,y}=6^\circ$) was used to smooth across the two relative orientation dimensions, before the mean and standard deviation of each error population was calculated.

Results

In the previous section it was demonstrated that the relative orientation of low-variance regions of the natural scene, strongly influences the perceived direction motion. In this section, we extend the analysis to examine how observers' bias and precision varies as a function of the conjoint relative orientation of elements across space i.e. we ask what the impact of relative orientation A is, in the presence of relative orientation B.

Observers' pattern of bias and variability is plotted in [Figure 3-8](#), to help the reader understand the space used and to relate the findings to the literature, [Figure 3-7](#) schematically illustrates the full range of second-order conjoint orientations. The abscissa and the ordinate of [Figure 3-7](#) denote the relative orientation of aperture A and B respectively; where the orientation of an edge (black line) is depicted relative to a standardised 2D motion (red arrow). The conjoint relative orientation of each aperture pairing is denoted by the two-dimensional coordinate in this space. A line of symmetry runs through the coordinate system from the lower left to the upper right ([Figure 3-7c](#); purple dashed line) and we note that the results were computed separately for each side of the line of symmetry. At all points along the line of symmetry the local direction of motion, within each aperture pair is identical, along the pink dashed line the local direction of motion within each apertures is the mirror opposite on the abscissa and ordinate. [Figure 3-7\(d\)](#) denotes regions of Type I (green) and Type II (grey) configurations of local motions

and Figure 3-7(e) denotes regions in which motion in the abscissa is faster than motions in the ordinate (blue), and vice-versa (green).

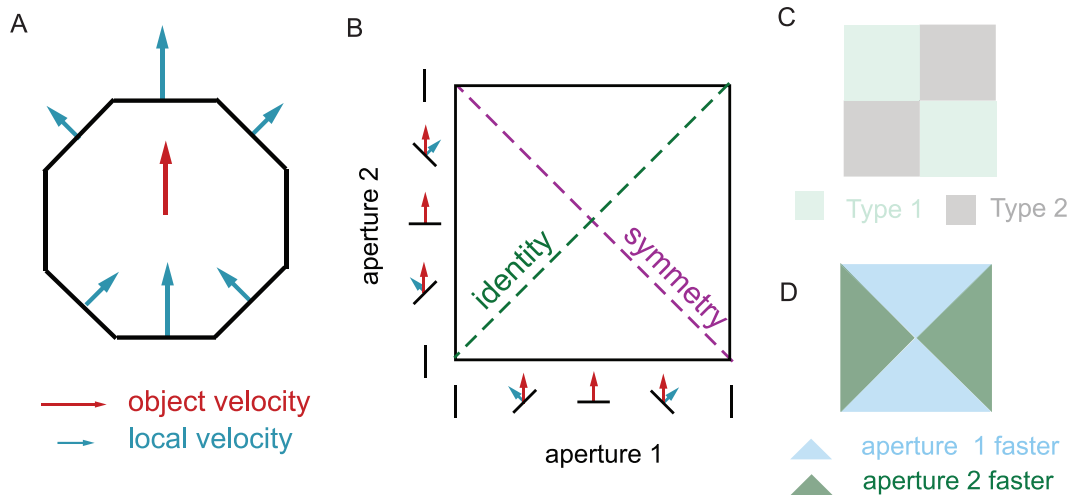


Figure 3-7 Representation of the second-order statistics used in Figure 3-8. (a) An example stimulus moving vertically upwards. The large red arrow depicts the global direction of motion, whilst the blue-arrows depict the local (1D) motions orthogonal to each contour orientation (b) Schematic representation of the complete set of pair-wise relations. Along the green dashed-line aperture pairings have identical orientations and along the purple dashed line, the apertures have mirror-reversed orientations. In (c) the areas of green denote Type I pairings (local (1D) motions fall on either side of the global (2D) direction of motion) whilst grey regions denotes Type II pairings (local (1D) motions fall the same side of the global (2D) direction of motion). (d) Blue denotes regions in which the local motions are faster in aperture one than aperture two, whilst the converse is true for green regions.

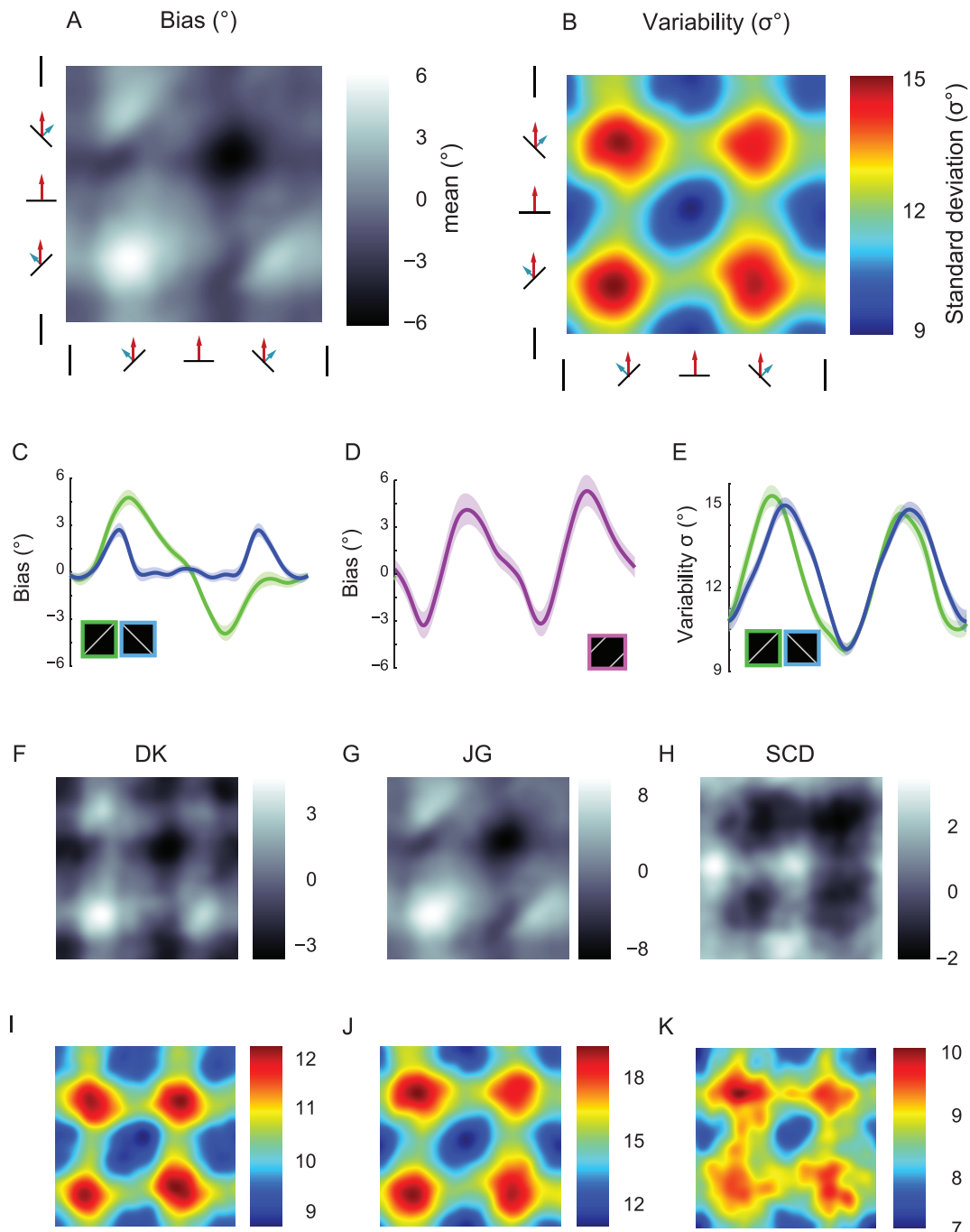


Figure 3-8 Observers' pattern of (a) bias and (b) precision as a function of the second-order relationships among aperture pairings for 3 observers (DK, JG and SD). (a) Light (positive) regions denote anti-clockwise bias and dark regions denote clockwise bias. (b) Warm regions denote high variability, whilst cool regions denote low variability. (c, d, e) One-dimensional slices from left to right through (a) and (b) as denoted by the insets. (c) Observers' bias for apertures

with identical relative orientations (green) and for mirror-symmetric relative-orientations (blue). (d) Non-symmetric Type I pairings, note how the bias tends to be in the direction of the fastest component of motion. (e) Same as (c) but for the variability statistics, there is no improvement for symmetric or identical aperture pairings. (f-h) bias and variability (i-k) statistics plotted individually for each subject.

Data in [Figure 3-8](#) depicts the (a) bias and (b) precision statistics as a function of the conjoint relative orientation of aperture pairings. To highlight the trends in our data and allow the reader to examine the 95% confidence intervals, [Figure 3-8\(c,d&e\)](#) plot one-dimensional slices (denoted by the inset). In [Figure 3-8\(c\)](#) the blue line depicts the bias of observers' when the local (1D) motions exposed by aperture pairings are symmetrically opposite the global (2D) motion vector and the green line depicts the bias of observers' when the local (1D) motions are identical. In line with previous findings, the bias for symmetric pairing is low (Bowns, 1996; Yo & Wilson, 1992) whilst for Type II pairings the bias has a greater magnitude and is towards the direction of local motion (Bowns, 1996; Burke & Wenderoth, 1993; Mingolla, et al., 1992; Rubin & Hochstein, 1993; Yo & Wilson, 1992). The magnitude of the bias for Type II pairing is reduced when the angular separation between the two components is increased, again consistent with research on plaids (Bowns, 1996; Burke & Wenderoth, 1993). It remains unclear whether this is due to the motion stream being better able to individuate motion signals when they are further apart in velocity space (as shown in motion transparency) (Braddick, Wishart, & Curran, 2002; Greenwood & Edwards, 2006a) or whether it simply

reflects the fact that as separation between the orientations in the stimulus increases then one or orientations move closer towards the (informative) static or orthogonal components of motion.

Examining the variability statistics (Figure 3-8b) there appears to be little impact of opposite pairings as performance is invariably noisy when the orientation of the stimulus is oblique to motion, regardless of their relative signs.

Discussion

Psychophysics

The aim of the psychophysics section was to investigate whether the results of studies investigating the perception of 2D motion using constrained stimulus classes (e.g. plaids or bars) generalise to the perception of motion when the stimulus is composed of naturally occurring textures; In the context of the 'aperture problem', natural scenes are different from plaid or bar stimuli because their orientation bandwidth is broad and the luminance and contrast vary independently across the stimuli (Mante, Frazor, Bonin, Geisler, & Carandini, 2005). Specifically, most studies on the 'aperture problem' have exposed observers to moving scenes containing just two discretely defined orientations, as this is the minimum number needed to uniquely specify a 2D motion. In contrast, naturally occurring textures contain a broader distribution of orientations and the aim of the psychophysics was to investigate the impact of naturally occurring contours and textures upon 2D motion perception. The reverse-correlation paradigm demonstrates that when low-

orientation variance elements (i.e. contours) are exposed to the observer, performance varies as a function of the orientation of the exposed elements relative to the 2D motion vector; specifically observers' estimates of direction are relatively imprecise and biased towards the direction of 1D motion when elements are orientated obliquely to the 2D direction. In contrast, observers' are relatively unbiased and precise when low variance elements are oriented parallel or orthogonal to motion. In other words, observers' are unable to discount the orientation structure of the natural scene when making directional judgements. This pattern of responses is consistent with psychophysical paradigms examining the perceived direction of translating bars; when a translating bar is oriented oblique to the 2D motion vector, observers' are biased towards the direction orthogonal to the bars orientation (Loffler & Orbach, 2001), particularly at short time periods (Lorenceanu, et al., 1993).

The majority of studies that examine observers' ability to solve the 'aperture problem' use stimuli composed of two orientations, as this is the minimum number needed to uniquely specify a 2D motion vector. To parallel this research, the reverse correlation paradigm is extended to examine observers' error distributions as a function of the second-order relationships between the conjoint orientations/directions that are exposed by pairs of apertures. In [Figure 3-8](#) I estimate observers' bias and variability over the full range of Type I & II combinations of orientations/directions. Consistent with previous studies I reveal that when the distribution of local (1D) motions is biased to one side of

the global (2D) direction (Type II) that observers' are biased towards the direction of local motion (Kaoru Amano, et al., 2009; Mingolla, et al., 1992; Wilson, et al., 1992; Yo & Wilson, 1992) and that the bias is reduced when there is a greater angular separation between the two orientations (Bowns, 1996; Burke & Wenderoth, 1993). The results also demonstrate that when the local motions are non-symmetrically either side of the global (2D) motion, observers' are biased in the direction of the fastest local motion, a finding not previously demonstrated.

Appendix

Scene statistics

In the results section the orientation statistics of the natural scenes were estimated by convolution of the natural scenes with a bank of log-Gabor filters tuned to directions between 0-165° at 15° intervals. The log-Gabor filters were constructed in the Fourier domain (Field, 1987) and the natural scenes were transformed into the Fourier domain using Matlab's **fft2** function. The product of the log-Gabor and the natural scene was calculated in the frequency domain and the results transformed back to the spatial domain using Matlab's **ifft2** function. This procedure is equivalent to performing convolution of the filter and the natural scene in the spatial domain.

Each log-Gabor G was constructed in the Fourier domain and was defined by Equation 3.6.

$$G = R(f_{xy})O(\theta_{xy})$$

Equation 3.6

Where $R(f_{xy})$ specifies the spatial frequency profile of the sensor and $O(\theta_{xy})$ the orientation tuning of the sensor, where f_{xy} denotes the spatial frequency of each point in the Fourier domain and θ_{xy} the orientation of each point in the Fourier domain.

$R(f_{(x,y)})$ is defined in Equation 3.9 where f_{peak} is the filters central frequency and σ is the ratio between the filters central frequency the standard deviation of the log-Gaussian, set to 0.65.

$$p(f_{x,y}) = e^{-\left(\frac{\ln(f_{x,y}/f_{peak})^2}{2\ln(\sigma/f_{x,y})^2}\right)}$$

Equation 3.7

$O(\theta_{xy})$ is defined in Equation 3.8 and is an angular Gaussian function, where ϕ (defined in Equation 3.9) is the angular separation between the orientation tuning of the sensor θ_{peak} and the orientation of each pixel in the Fourier domain.

$$O(\theta_{xy}) = e^{-\left(\frac{\phi^2}{2\sigma_\theta^2}\right)}$$

Equation 3.8

$$\phi = \left| a \tan 2\left(\sin(\theta_{xy} - \theta_{peak}), \cos(\theta_{xy} - \theta_{peak})\right) \right|$$

Equation 3.9

Once the energy at each orientation had been calculated the *sum* of the orientation energy, the *mean absolute orientation* $\bar{\theta}$ and the *orientation variance* was calculated. This was done on a pixel-by-pixel basis for Results section II and on an aperture-by-aperture basis on Results section III. The mean orientation $\bar{\theta}$ is calculated by Equation 3.10, where θ is the orientation of each filter and E_θ the filter output.

$$\bar{\theta} = \frac{1}{2} \tan^{-1} \left(\sum_{\theta} \sin(2\theta) E_{\theta}, \sum_{\theta} \cos(2\theta) E_{\theta} \right)$$

Equation 3.10

The orientation variance was calculated from Equation 3.11 & Equation 3.12.

$$R^2 = \frac{\sum_{\theta} (\sin(2\theta)E_{\theta})^2 + \sum_{\theta} (\cos(2\theta)E_{\theta})^2}{\sum_{\theta} E_{\theta}^2}$$

Equation 3.11

$$V = 1 - R$$

Equation 3.12

On each trial, the mean orientation of a pixel or aperture was converted to a relative orientation term by calculating the angular separation between the 2D direction and the mean orientation of a pixel

$$\theta_{relative} = \tan^{-1} \left(\frac{\sin(\theta_{2D} - \bar{\vartheta})}{\cos(\theta_{2D} - \bar{\vartheta})} \right)$$

Equation 3.13

(4) Global Motion Model

In this section I show how the pattern of observers' *bias* and *variability* may result from an interaction between the trial-by-trial spatiotemporal anisotropies in the stimuli (Adelson & Bergen, 1985) and a template-model of global motion (e.g. Nowlan & Sejnowski, 1995; Perrone, 2004; Simoncelli & Heeger, 1998) that is 'optimally' tuned for isotropic stimuli.

Although the orientation structure of natural scenes can be described statistically, the motion stream does not know what the orientation structure of a particular stimulus will be at any moment. I propose that a reasonable global motion strategy is for velocity-tuned global motion (GM) sensors to integrate across all signals consistent with that global motion (i.e. a cosine across speed and direction). That human observers adopt such a strategy is supported by evidence from Schrater et al. (2000) who show least masking for detection of a moving noise pattern when the signal energy is evenly distributed across all orientations.

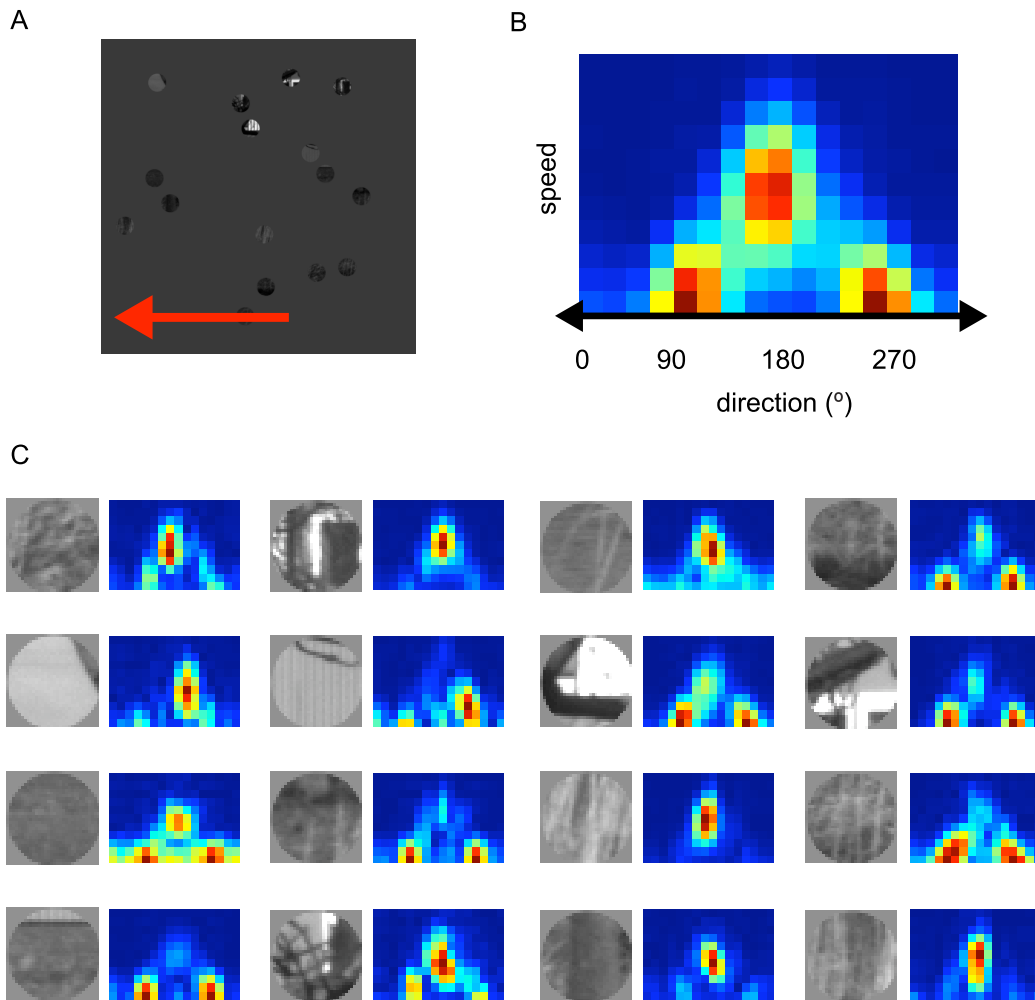


Figure 4-1 (a) Example trial in which the image rigidly translates in the leftward (180°) direction (b) sum of motion energy across all the apertures. (c) Motion-energy within the individual apertures

This chapter tests whether the pattern of psychophysical data results from a mismatch between trial-by-trial anisotropies in the orientation structure of the stimulus and a template model of area MT in which each GM sensor is optimized for an isotropic stimulus. I first estimated 1D motion using directionally selective (DS) motion-energy filters (Adelson & Bergen, 1985) as described in the introduction. The peak spatial and temporal frequencies tiled the full range of motion signals that the present global motion stimuli

could elicit (within one spatial-frequency channel) and the response of the motion sensor bank to an example trial is shown in Figure 4-1, encouragingly the pattern of activity is able to capture to cosine relationship between the speed and direction of 1D motions although the energy is not evenly distributed around the cosine. Averaging the response of DS filters to natural scenes allowed us to generate a series of GM sensor weighting profiles that were optimized to the stimulus and the filters configuration at hand (see Methods, below). GM filters were constructed over a broad range of speeds and directions. Selection of the final global motion estimate was achieved via a winner-take-all algorithm; such a procedure is equivalent to a maximum likelihood estimate of direction for a given set of GM filters.

The inspiration for the model was threefold; first, the observation that response variability was greater for orientations oblique to the global direction of motion was consistent with the motion stream fitting a cosine to the pattern of motion energy, but not with the motion stream computing an IOC (Adelson & Movshon, 1982) solution - for which each orientation is equally informative. Second, the pattern of bias was typically in the direction of 1D motion and maximum likelihood estimators (MLE) are demonstrably prone to such systematic biases (Webb, Ledgeway, & McGraw, 2007). Thirdly, our psychophysical data demonstrate that orientations parallel and orthogonal to the 2D velocity are informative, suggesting that models that only integrate over one temporal frequency channel (Rust, Mante, Simoncelli, & Movshon, 2006; Webb, et al., 2007) would be insufficient to account for the data at hand.

The global motion stage is similar in principle to Simoncelli and Heeger's (Simoncelli & Heeger, 1998) MT model, in that I employ a population of simulated MT neurons each with separate velocity tuning. The main difference between our approach and that of Simoncelli & Heeger (1998) is that the global motion sensors in our model are derived empirically whilst Simoncelli and Heeger (1998) derive their weighting functions mathematically. The data I report here are not able to differentiate between the models and a comparison between the two models is made in the conclusion.

Methods

A two-stage global motion model putatively representing the motion areas V1 and MT of primate motion stream is used to model the psychophysical data. The aim is to explore whether a global motion (GM) optimized for an isotropic stimulus class, exhibits the same pattern of bias and precision as our observers' when presented with stimuli that were anisotropic on a trial-by-trial basis.

Local 1D motion sensors (V1)

In the previous chapter I examine observers' ability to estimate the direction of motion as a function of the underlying orientation structure of the apertured image (during each trial). Orientation is used as an estimate of the direction of 1D motion in each aperture – a helpful simplification since it

allows us to reduce the variable of stimulus velocity to one-dimension. However, using orientation as a measure of 1D motion does not capture all the properties of V1 direction selective cells and a complete model of global motion processing must include a biologically plausible 1D motion stage. To this end I implemented a motion energy model of V1 directionally selective (DS) neurons (Adelson & Bergen, 1985). In order to capture the full range of 1D motions, DS sensors were tuned to 16 evenly spaced directions around the clock (0-337.5° in 22.5° steps) and a broad range of speeds (from static to 200% of global speed at 10% intervals) where a DS sensor's pseudo-speed tuning is defined by the ratio between its (peak) temporal and spatial frequency tuning.

Three simplifications were incorporated into the model:

(1) The spatial sampling was determined by the aperture positions on each trial. I assume zero motion energy response at all locations except at the centre of the aperture, allowing us to perform multiplication and not convolution of DS templates with the stimuli

(2) Only one SF channel was modelled

(3) All DS sensors had identical spatiotemporal envelopes ($\sigma_{(x, y)} = 7$ pixels, $\sigma_t = 7$ frames).

I chose to pre-determine the ME filter positions in order to reduce the contribution of aperture edges that would otherwise introduce an isotropic signal across the static and slow temporal frequencies. This is problematic if

one wishes to model the psychophysical results, because the static component of motion (from the parallel orientations) led observers to make precise and unbiased 2D direction judgements and this static/parallel signal would be weakened by the isotropic static signal from the aperture edges. The relatively fine temporal frequency selectivity of the motion energy filters used means that the signal aperture edges would be less problematic for those sensors tuned to greater pseudo-speeds/temporal frequencies. The problem of aperture edges, would be even greater if the properties of the motion-energy filters used in this study conformed to the known properties of V1 DS cells (e.g. Foster, et al., 1985), and this issue has been discussed in (Johnston, et al., 1992). By providing the model with knowledge of the aperture positions we are able to circumvent the problem for the static aperture edges, but only by providing the model with extra information. I note that there is ample evidence that the motion stream can ignore the influence of aperture edges using both binocular and monocular cues (McDermott, et al., 2001; Shimojo, et al., 1989), although there is no firmly established mechanism through which this is accomplished.

Restricting the analysis to one spatial-frequency channel presents a problem for the model, since the recovery of stimulus speed can only be achieved via a broadband integration of signals across spatial frequency channels due to the independence of spatial and temporal-frequency tuning observed in V1 directionally-selective cells (Foster, et al., 1985; Tolhurst & Movshon, 1975) and also present in the motion-energy model (Adelson & Bergen, 1985). However

natural scenes exhibit two properties that minimize this problem; First, natural scenes have a broadband and approximately $1/f$ amplitude spectrum; Second, images contain structure (edges) that contain information that is phase aligned across spatial frequency bands (Attneave, 1954; HB Barlow, 1961). These properties, combined with the relatively narrow temporal frequency tuning of the motion-energy sensors employed (relative to the temporal frequency relative to the bandwidth of V1 DS cells; Foster, et al., 1985; Perrone & Thiele, 2002; Tolhurst & Movshon, 1975), result in the speed tuning of our model being 'fit for purpose'. This point is justified by the precise estimates of velocity the model produces which produces estimates of direction with a precision $\sigma \approx 3^\circ$ in response to the natural scenes used in the psychophysical data. This greatly outperforms our human subjects. Figure 10 highlights the motion-energy generated by an example trial and a clear cosine pattern of motion-energy is apparent as a function of speed versus direction.

Global motion sensors (MT)

The global motion (GM) stage utilized sensors tuned to a wide range of speeds and directions. The weighting profiles were derived from the response of the 1D motion stage to drifting natural scenes. The scenes were randomly selected from the Van Hateren image set (van Hateren & van der Schaaf, 1998). Specifically a number of templates were created by averaging the 1D motion energy distribution elicited by stimuli travelling at a range of directions. During each trial the natural scene translated in a random direction but at a

specified global speed, the motion-energy derived on each trial was then phase-shifted to a standard direction (0°) and summed with the other trials. In total 21 templates were produced from zero speed (static) to 2 pixels per frame at 0.1 intervals. GM sensors tuned to the full range of directions at 0.1° intervals were constructed by phase shifting the averaged template. The resulting templates are homogeneous as a function of direction. Considerable complexity could be added to the model by generating separate templates for each direction of motion.

Model Details

1D motion sensors were convolved with the stimulus in the space and time domain (i.e. not in the Fourier domain). Sensors were centred upon the middle of each aperture and the middle frame – motion DS sensors and movies had dimensions of 0.5° , 0.5° & 0.37 seconds in x, y, and t (32 by 32 pixels, by 32 frames). Motion energy sensors were constructed from Equation 1.22 and had a peak sensitivity to structure at 4 c/deg. The spatiotemporal envelope was kept constant across all DS sensors (x, y = 0.2 arc min, t = 0.1 seconds - 7 by 7 pixels by 7 frames). This was beneficial because the directional bandwidth was kept constant at around 45° (half-wave at half height, as measured from the response to spatial frequency matched sine-wave gratings) and the maximal sensor response was identical across all speeds and directions.

Global motion integrators were tuned to speeds from 0% to 200% of the actual object speed (1pixel per frame=1.33°/s) at 1% intervals and to directions around the clock at 0.1° intervals. Such a fine spacing was needed since it sets the limit on the precision of the winner-takes-all algorithm used to select the 'winning' GM sensor.

No noise, normalization or gain was incorporated into the model because I wished to explore the 'noise' generated by convolving the GM sensors with anisotropic motion energy profiles without the complexity introduced by such mechanisms.

Results

The model was tested with both artificial-stimuli and with replicas of the stimuli used in the psychophysical trials. Testing across both stimulus classes allowed us to assess which features of the model output are due to the underlying mechanisms of the model, and which were the results of stimulus anisotropies.

Artificial Stimuli

The global-motion sensors perform optimally when presented with isotropic stimuli. Accordingly I was interested in how the model performed when presented with anisotropic motion energy profiles. To generate artificial stimuli and relate the analysis to the psychophysical data of Section III I took the motion energy profile for a rigidly moving object [Figure 4-2](#) whose component motions are represented by the white dashed line. Imbalances in motion energy were added in a pair-wise manner along the cosine to allow us to

relate the models behaviour to the results of Section III. In Figure 11(b) two Gaussian energy profiles have been constructed lying along the cosine (at -70° and $+40^\circ$ away from the veridical direction) that defines the global motion.

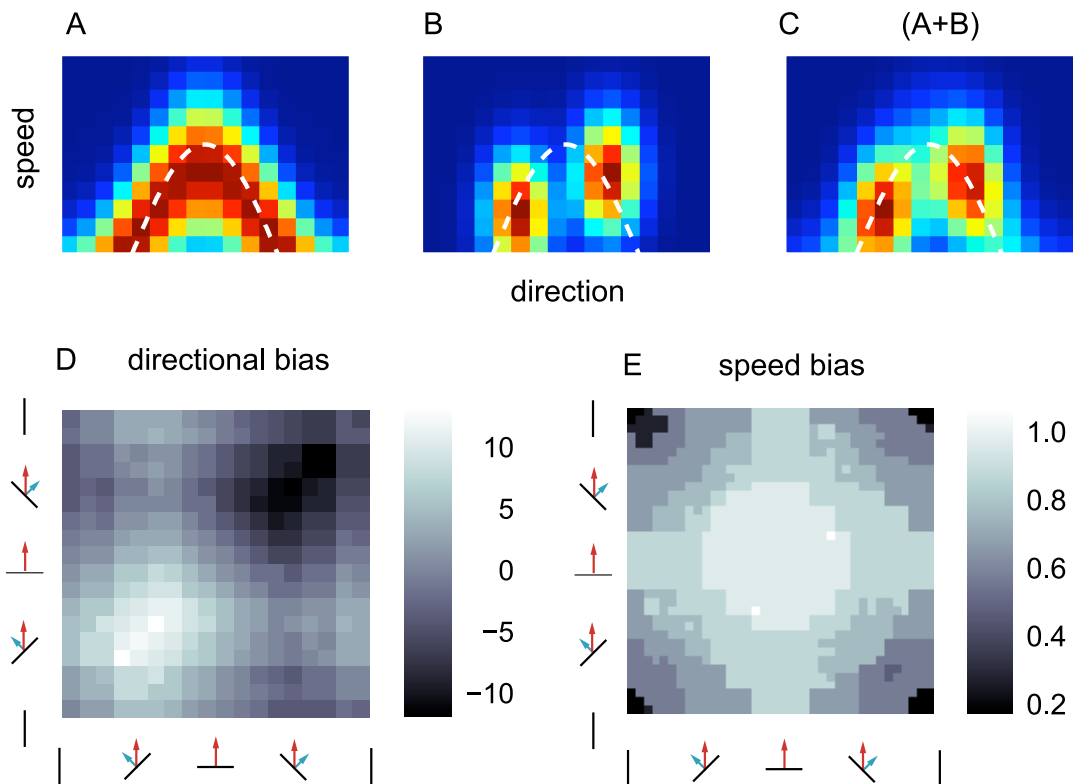


Figure 4-2 The global motion model was tested with anisotropic motion energy distributions. The input to the global motion stage was generated by taking the motion energy for an isotropic stimulus in (a) and superimposing additional signals along the orientation structure of the object (denoted by the white-dashed arrow). Signal additions were in a pair wise manner to allow us to produce model estimates (d & e) that can be compared to the data in Section III (d) model direction estimates (e) model speed estimates.

Figure 4-2(d, e) depicts the direction and speed estimates of the model. Each datum corresponds to only one trial, but in a noiseless model this provides a measure of the underlying biases of the system. The results reveal that the model's estimates of direction and speed vary systematically with the motion

energy imbalances. Encouragingly, the bias results are in good agreement with the psychophysical data with the directional estimates being drawn towards the motion energy imbalance for Type II combinations and towards the fastest component of motion for Type I combinations. Speed estimates become biased towards slower speeds as the component motions move away from the orthogonal orientation. Two factors lead to this bias in the speed estimates: The first is simply that as the motion energy moves away from the orthogonal orientations it shifts to progressively lower temporal frequencies and the 'winning' template is shifted towards lower speeds. The second reason is less immediately obvious and results from the 1D velocities of a faster moving object being spread over a greater range of temporal-frequencies/speeds than a slower moving object. Given that the total motion-energy is constant as a function of speed in our derived templates (Figure 4-3e) the motion-energy (or feed-forward weighting) must be more concentrated for templates tuned to slower speeds. To elaborate, two global motions travelling in the same direction, but with different speeds (Figure 4-3 a, b) overlap substantially in the low-to-static temporal frequencies, but are distinct at high temporal frequencies (orthogonal to motion) – thus if the orientations orthogonal to motion are not well represented there is little to disambiguate competing speed estimates, in this case the stronger weightings of the slower global motion templates (Figure 4-3c) bias global-motion estimates towards slower speeds.

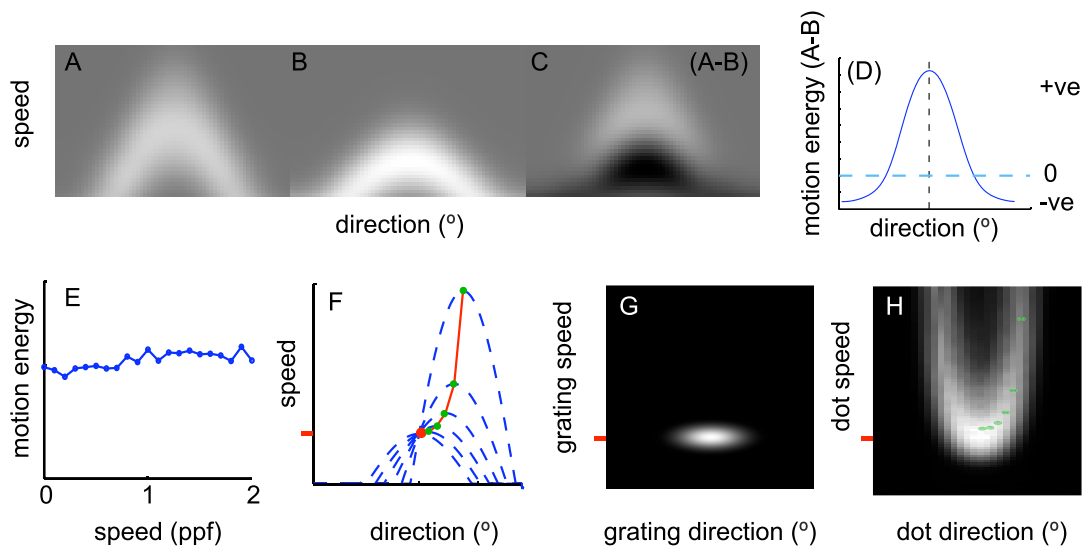


Figure 4-3 Motion energy for a SF band-pass dot moving at (a) 2 pixels per frame or (b) 1 pixel per frame. The difference between (a) and (b) is plotted in (c). (d) The absolute motion energy difference of (c) collapsed across the speed dimension. (e) Demonstrates that the sum of the motion energy over both speed and directions is largely constant regardless of the underlying object speed. However as (a) is spread over a greater range of speeds, the concentration of motion-energy across each direction must be greater in (b). In (d) the greatest difference between the two signals is to be found in the direction of global motion but that (b) has greater motion energy in the overlapping low-speeds – this aspect leads to a global motion stage being biased towards low-speeds for stimuli with a weak orthogonal component of motion. (f, g & h) give insight into the pattern of motion energy shown in the row above. In (f) I plot a series of global motions (green dots) whose component motions (blue lines) pass through the velocity tuning of a DS sensor denoted by the red dot. In (g&e) I plot the response of the DS sensor in (f) to a grating stimulus (g) or dot stimulus (h). Note how the profile of (g) closely follows the red line of (f) but the motion energy in (h) falls with increasing speed. This is because the component motions which pass through the receptive field of the DS sensor (red dot, f) are more finely spread.

Natural Scenes

The global-motion model was next tested with the stimuli used during the psychophysical experiments. This allowed me to repeat the second-order analysis of Section III, replacing the observers' directional response errors with those of the model. The patterns of bias and variability generated by the model are shown in [Figure 4-4](#) and are in good qualitative agreement with

the observers' data. Results also highlight unexpected anisotropies present in the observers' response - but not in the models predictions to the artificial motion energy profiles.

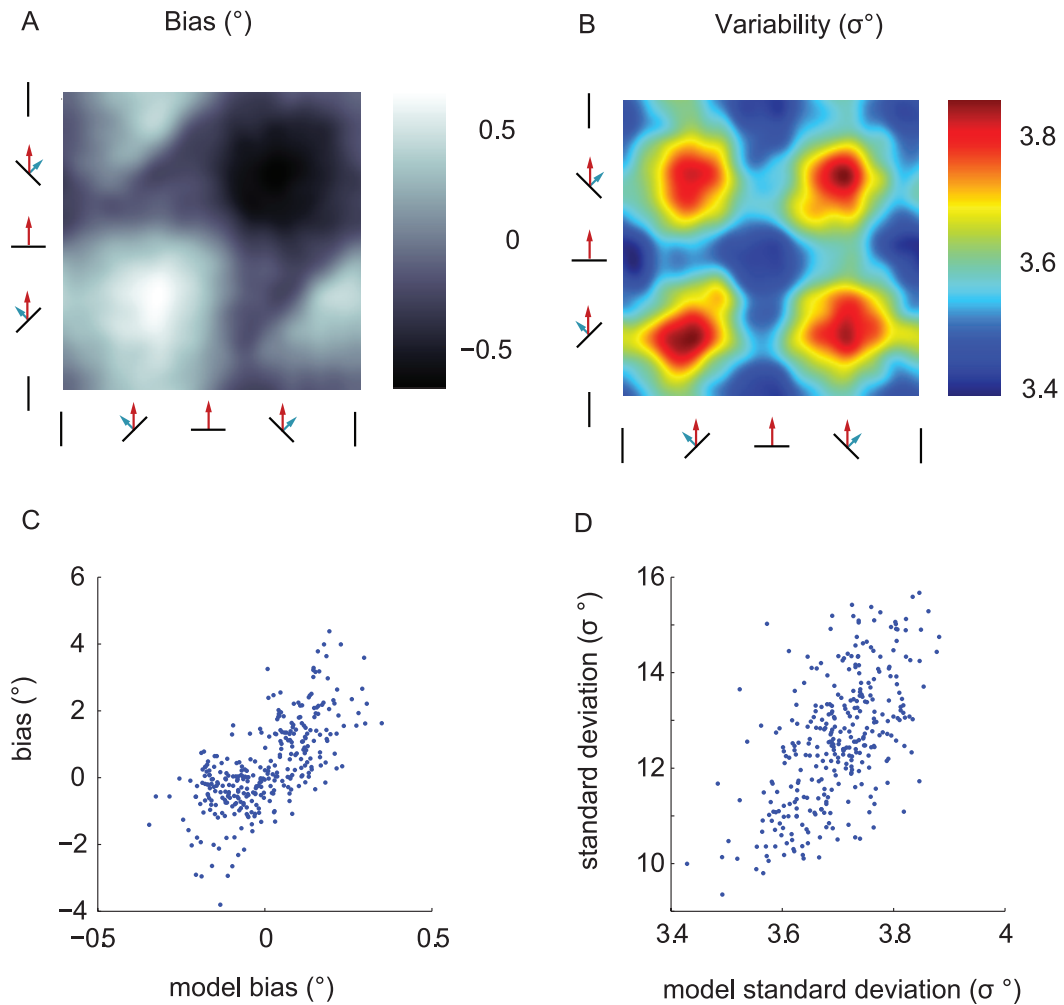


Figure 4-4 (a) Modal bias (b) and model variability as a function of the second-order relative-orientations of aperture pairings (c, d) Scatter plots of the model against the observers' bias (c) and variability (d).

To provide a more robust statistical analysis, the model and observer bias and variability were recalculated using larger bins (10°) but no further smoothing operation. This generated 441 independent measures of bias and variability

and a Pearson's correlation between the model and the observer bias revealed strong and significant correlations ($R = 0.72$ $p < 0.00001$ and $R = 0.72$, $p < 0.00001$ respectively).

Despite success in predicting the observer bias and variability, a correlation of the model errors on a trial-by-trial basis was very weak ($R = 0.025$), but significant at $p < 0.005$ ($N=34,000$). Thus while I am able to model the statistical properties of observers' *bias* and *precision*, I am unable to model observers' trial-by-trial variability.

Discussion

In this chapter, I show that a simple template model of global-motion processing (putatively reflecting visual processes that occur in area MT of the primate brain) optimally tuned for isotropic stimuli exhibits a psychophysically realistic pattern of errors (biases and precision) when confronted with anisotropic natural stimuli. One of the key motivations behind the model is the observation that the pattern of psychophysical data is consistent with global-motion being computed by fitting a cosine to the 1D motion energy distribution. Under such a model, the parallel and orthogonal components of motion are highly informative. To elaborate: the fastest component of motion is the signal closest to (if not identical to) the global velocity. Thus the more strongly this component of motion is represented at the 1D (motion energy) level, the more the model will be drawn to the veridical direction.

Interestingly, because the speed of 1D motion can never exceed that of the global motion and the receptive field profiles of lower speed tuned MT cells have relatively higher weightings in the low-temporal frequency component of the signal. Consequently, the model exhibits a bias for slower speeds (Figure 4-2e) similar to the model of Weiss et al., (Weiss & Adelson, 1998; Weiss, et al., 2002), but without the need for an explicit prior.

At the other end of the temporal spectrum, the static (parallel) component of motion is only consistent with two global motions that are 180° apart, thus for a cosine-fitting model this signal is very informative. Existing evidence that the motion stream is able to utilize the static component of motion comes from studies of randomly refreshed Glass patterns, which exhibit a consistent static signal, but a noisy and isotropic motion signal. When presented with such stimuli, the observers' percept is bimodal, switching between the two directions of motion predicted by the static component of motion (Ross, Badcock, & Hayes, 2000); furthermore, the inclusion of just a small percentage of coherently moving dots (~10%) can stabilize the motion percept (Ross, 2004). These empirical observations are consistent with the key role of static temporal frequencies in the present global-motion model.

The pattern of bias in the model and psychophysical data in response to natural scenes is smaller in magnitude than that observed in response of artificial scenes containing just two orientations (K. Amano, et al., 2009; Mingolla, et al., 1992; Rubin & Hochstein, 1993; Yo & Wilson, 1992). To test

whether the model also produced large errors in response to stimuli with a more constrained orientation structure I developed a means of moving from a simple stimuli composed of just two orientations to a stimuli with an isotropic orientation structure. The approach was an adaptation of the approach illustrated in [Figure 4-2](#). Instead of using a stimulus with a 50-50 balance between the isotropic and plaid stimuli, the percent of each signal was varied from 0 to 100%. The results are shown in [Figure 4-5](#). When the percent of the signal is 100% anisotropic the system is biased by over 20° towards the 1D direction similar to that found in the psychophysical literature (Kaoru Amano, et al., 2009; Mingolla, et al., 1992; Rubin & Hochstein, 1993). However when the percentage of the isotropic signal is increased, the error steadily falls to zero at 100% isotropic. Thus, the model performs well on stimuli with broader orientation content, but poorly on stimuli with highly constrained and limited orientation content.

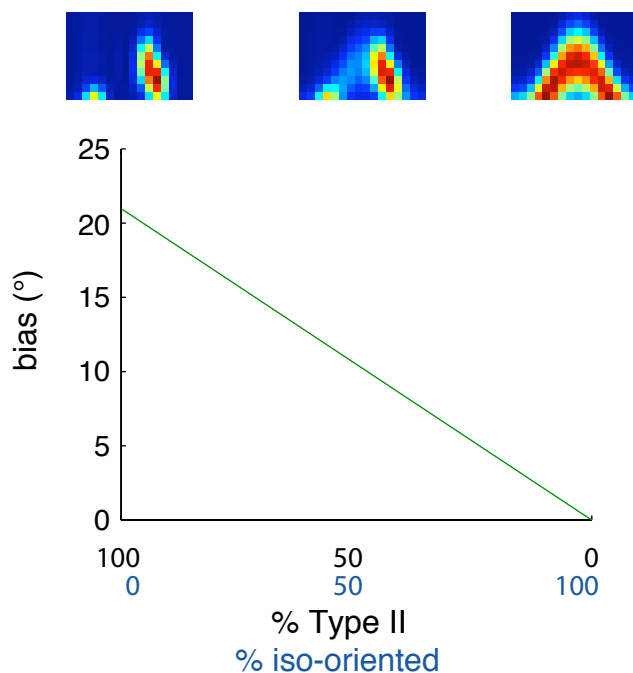


Figure 4-5 The model is tested on a range of stimuli composed of two orientations in a Type II configuration (100% Type II, 0% isooriented) to a signal composed of an isotropic signal (0% Type II, 100% isooriented). Increasing the percentage of the isotropic signal improves the model precision.

It should be noted that the only source of noise in the model is the direction and speed bandwidth of the 1D motion sensors - if the 1D motion were represented as discrete points then a fitting procedure would always produce a veridical answer, provided two or more orientations are present. This suggests that models operating on discrete representation of the stimulus are designed to an inappropriate level of abstraction for modelling the human visual system.

Although the model provides a good estimate of observer bias and variability as a function of the second-order orientation statistics, the model is unable to capture observers' trial-by-trial variability. This likely reflects a number of factors. The first is simply that the relative-orientation of aperture patches may not be the main cause of observers' trial-by-trial variability. Note, that the least variability was observed in the cardinal directions with a directional standard deviation of around $\sim 8-10^\circ$. Given that the observed biases as a function of relative-orientation (Section II&III) are smaller (around $3-6^\circ$), it may be that observers' stochastic response variability simply swamps the predictable variance caused by motion-energy imbalances. A second contributory factor could be that the model operates in a homogeneous manner as a function of direction, whereas the psychophysical data exhibits a number of anisotropies such as the oblique effect (Dakin, et al., 2005b;

Gros, et al., 1998), cardinal attraction (Loffler & Orbach, 2001) or reference repulsion (Rauber & Treue, 1998) that are not implemented in the model.

A third reason why the model fails to capture trial-by-trial variation may be the lack of any gain control mechanisms in the model, which may serve to alter the relative energy responses across the natural scenes. This is particularly pertinent because both the psychophysical data and the model demonstrate how imbalances in the energy across the orientation structure of the scene can lead to systematic errors in performance. Previous studies have shown that the use of natural stimuli can alter observers' response characteristics as a function of the underlying spatiotemporal frequency leading to the suggestion that heterogeneous factors are 'whitened' in response to natural scenes (Bex, et al., 2005). More specifically, the role of contrast saturation and gain control has been shown to be important in shaping the response properties of MT 'pattern' selective cells. For example, Rust et al. (Rust, et al., 2006), measured the response of MT pattern-selective neurons to plaid stimuli whose components spanned the full range of second order orientation relations (for one speed) that were consistent with the cell's velocity tuning. When the component gratings were both oriented in a manner consistent with the global velocity and fell either side of the global direction, MT cells responded strongly. When the component gratings moved in the same direction, the overall energy in the stimulus was identical but the response of MT pattern cells response was less strong. Rust et al. (Rust, et al., 2006) were able to model this property via a contrast saturation function

working within each directional channel or cell. This is important because the present study demonstrates that the impact of imbalances in motion energy is both to bias the percept of motion and to increase response variability – a contrast saturation function in combination with an un-tuned normalization function (Heeger, 1992a; Rust, et al., 2006) may serve to reduce this bias.

Experimental Chapter No. 3

Testing the global motion model

The previous chapter revealed that a model which fits *cosine* templates to the local motion energy (Adelson & Bergen, 1985) distribution could predict observers' patterns of *bias* and *variability* as a function of the second-order combination of orientations in natural scenes. However I was unable to model observers' performance on a *trial-by-trial* basis and a number of possible explanations were put forth. Given that the correlation between the observers' and the model's *first* and *second* moments were both *strong* ($R \sim 0.7-0.8$) and *significant* ($p < 0.0001$) we can conclude that the results were not spurious. I present two competing theories that could account for the inability to predict observers' error on a trial-by-trial basis; firstly that a number of theoretically predictable factors (e.g. contrast gain (Rust, et al., 2006) or anisotropies in V1 or MT response profiles (Dakin, et al., 2005b; Gros, et al., 1998)) were not incorporated into the model; or secondly that the lead cause of observers' variability was stochastic (i.e. not driven by the stimulus) and was thus unpredictable.

To explore this issue I ran a new experiment using the same task described in the natural scenes previous chapter but using (a) a more constrained stimulus class to allow greater control over the stimulus and (b) a "double-pass" technique to estimate the proportion of stochastic and deterministic variability in the data. The stimulus used was a global-Gabor array similar to

that used in a number of previous studies (e.g. Kaoru Amano, et al., 2009; Lorenceau & Zago, 1999). Unlike the natural scenes study, in this paradigm the exact configuration of the stimulus under direct experimental control allowing me to control for factors such as contrast and luminance (by keeping them constant across each element) and orientation (by directly specifying the orientation of each element). The basic configuration was of four Gabor elements that were randomly distributed within a 4° radius from fixation. The orientations of the Gabor's were either random or evenly distributed from 0° to 180° and the speed of each element was configured such that the motion from each Gabor was consistent with an underlying rigid global 2D translation. The rationale behind using *random* and *even* orientation distributions is that observers' can be biased by the orientation content of the stimulus under Type II conditions (Kaoru Amano, et al., 2009; Bowns, 1996; Burke & Wenderoth, 1993; Loffler & Orbach, 2001; Mingolla, et al., 1992; Rubin & Hochstein, 1993; Wilson & Kim, 1994; Yo & Wilson, 1992). As the sign and magnitude bias will vary with the orientation configuration of the global-Gabor array, randomizing the orientation structure on a trial-by-trial basis is predicted to increase in the total variability of observers' error distributions and importantly the proportion of variability that is determined by properties of the stimulus, rather than stochastic factors such as neural noise.

The aim of this experiment is to determine the proportion of stimulus-led (i.e. orientation led) errors that a model of 2D motion can account for. In other words, I predict that randomizing the orientation structure will increase the

amount of stimulus-led variability in observers' errors and the test of the model is to what extent the model is able to capture such stimulus-led variability. In order to estimate the proportion of the variability that is stochastic (unpredictable) or stimulus-led (predictable), a double-pass technique was employed; in this approach multiple presentations of the same stimulus are interleaved within each run. The proportion of predictable and unpredictable variability can then be estimated by calculating the correlation (R) between the observers' errors on stimulus retrieval; R^2 is an estimate of the percentage of the variability that is predicted by the stimulus retrievals. The same approach can be taken to estimate the proportion of the observers' variability that can be predicted by the model. Finally, the R^2 of the observer on stimulus retrievals and the R^2 of the model and observer can be compared to measure what degree of the observers' predictable variability the model can account for.

The initial testing used four Gabor elements, however two subjects were unable to perform under these conditions so the number of elements was increased to sixteen. To keep the number of orientations used to a minimum only four orientations were used in the sixteen-element condition (i.e. each orientation was repeated four times). In total two observers (DK and SD) completed testing using four elements and three completed testing using sixteen elements (DK, JG & PB). The reason for the failure of two observers to perform in the sixteen-element condition is the subject of a later equivalent noise experiment.

Methods

Subjects/Apparatus

Four subjects completed the testing. The apparatus was identical to the system used in the earlier Natural Scenes chapter.

Stimuli (experiment 1)

Stimuli consisted four or sixteen Gabors randomly positioned within a circular region (4° radius, centred on fixation). In the four-elements condition the orientations were either randomly chosen or evenly spaced at 45° intervals. In the even condition a random orientation-offset was added such that the absolute orientations of the elements were unpredictable on each trial (but the 45° separation was preserved). The speed of each Gabor was then specified to be consistent with a randomly chosen global 2D motion. The speed of the global 2D velocity was always 1.333 degree per second but the direction of motion was randomised each trial. In the sixteen-element condition, the same procedure was used except that each of the four Gabors was copied four times. The mean luminance of the stimuli was 40 cd/m² and the root-mean-square contrast of the each Gabor element was fixed at 0.20.

Procedure (experiment 1)

Observers' were presented with a set of drifting Gabors, which appeared for 0.3765 ms, before being replaced with a spatial-frequency matched noise mask. When the mask was present, observers were asked to manipulate the

orientation of a row of four Gaussian dots on a radius until it matched the perceived "overall" direction of motion of the preceding stimulus (a procedure identical to that in the natural scenes chapter). Runs consisted of interleaved trials of the *even* orientation and *random* orientation conditions and stimulus retrials. Each trial from each condition was repeated twice during each run. The correlation between the observers' data on the *first* and *second* pass was used to estimate the proportion of variance that was predictable and stimulus driven and the proportion of variance that was stochastic and unpredictable.

Data Analysis

Observers' *bias* and *precision* were calculated for each condition in an identical manner to the experiments on natural scenes. To estimate the proportion of variance that was either stochastic or stimulus-led a Pearson's correlation coefficient between errors on the first and second pass was computed and the R-score was used to infer the proportion of predictable and unpredictable variability using Equation 4.1 & Equation 4.2. A bootstrapping procedure was used to estimate 95% confidence intervals.

$$\sigma_{predictable} = R^2 \sigma_{total}$$

Equation 4.1

$$\sigma_{unpredictable} = (1 - R^2) \sigma_{total}$$

Equation 4.2

Results

Psychophysics

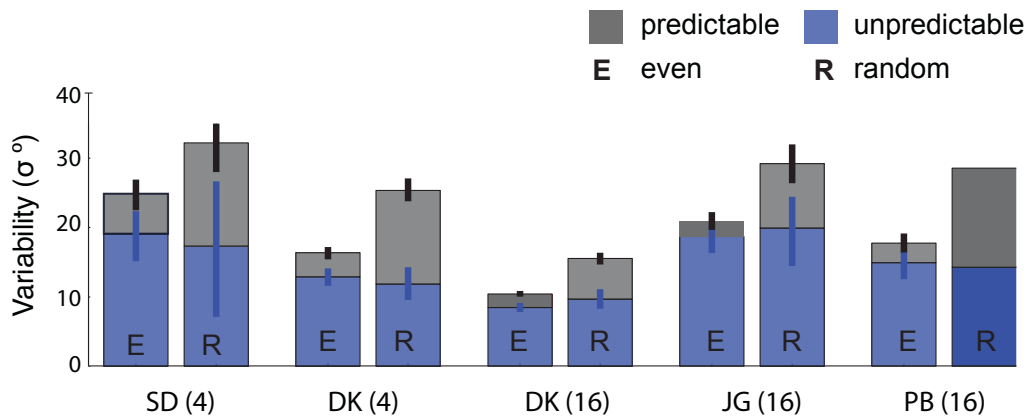


Figure 4-6 Observers' precision for the even and uneven orientation conditions. In all cases the precision is least for the uneven orientation condition. The data is broken into predictable (grey) and unpredictable (blue) proportions (see methods). The majority of the increase in variability in the uneven condition is potentially predictable and stimulus led. Error bars are 95% confidence intervals.

Observers' performance in the *even* and *uneven* conditions is shown in the stack plots in Figure 4-6. The full height of each stack is the total variability of observer responses whilst the red and blue sections denote estimated proportion of the variability that is either stochastic/unpredictable (blue) or stimulus-led/predictable (red). To estimate the proportion of the variability that was stimulus-led, a double pass technique was employed, whereby each trial was shown twice and the error on the first and second pass was correlated. The correlation for subject DK on the random orientation and 16-element condition is shown in Figure 4-7. The resulting R-score was 0.62, the R^2 term is known as the coefficient of determination and describes the percentage of the total variance that is predicted by variable A on variable

B. The proportion of variability that was either stimulus-led or unpredictable was then estimated from the R^2 score using Equation 4.1 & Equation 4.2 and The blue and grey regions of Figure 4-6 correspond to the values from Equation 4.1 & Equation 4.2.

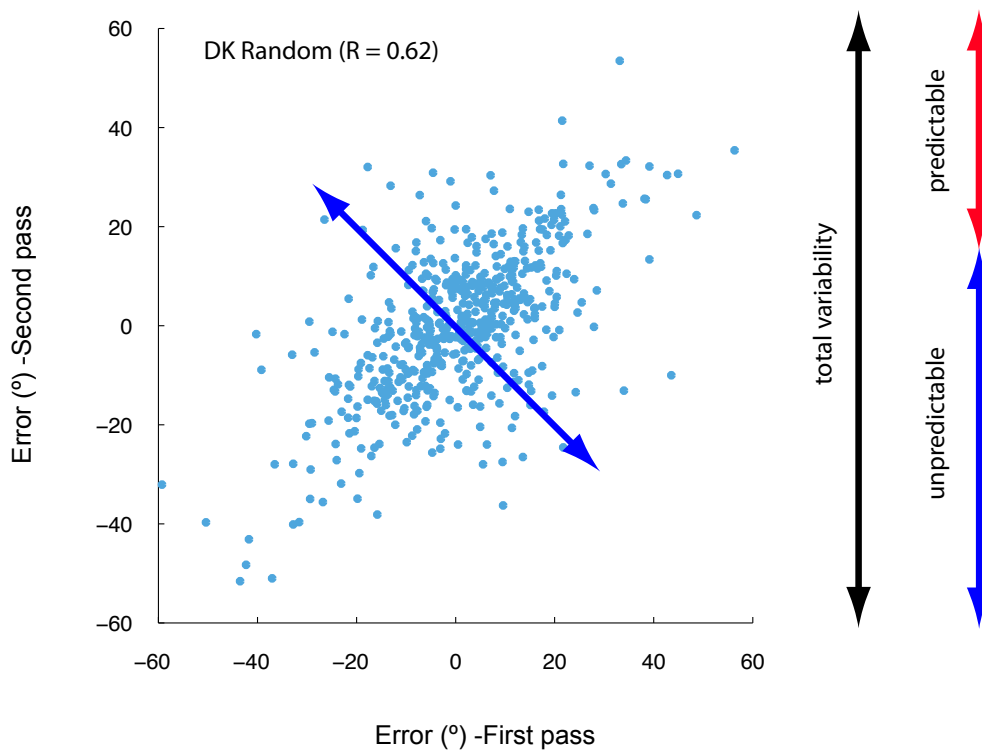


Figure 4-7 Correlation for the error on the first and second pass for subject DK in the 16 element condition. The black arrow depicts the total variability of the errors on the second pass. The blue arrow depicts the estimated proportion of variability that is left *unexplained* by the error on the first pass. Finally, the red arrow depicts the estimated variability that is *explained* by the error on the first pass (black minus the blue arrow)

The results of the analysis show that for all subjects the total variability is greatest in the random orientation condition and this increase in variability can be attributed to a corresponding increase in the percentage of stimulus-led variability. This is encouraging because it suggests the model should (theoretically) be able to capture the additional variability caused by randomising the orientation structure of the stimulus.

The next section extends the psychophysics of this section to an Equivalent Noise paradigm by adding direction noise. The aim of the EN section is examine what proportion of observers' variability is due to internal noise or poor sampling efficiency and to examine how the estimates of both properties affect observers' self-consistency. If poor sampling efficiency is the main contributory factor to observers' stochastic noise then it might be possible to increase observers' self-consistency by improving the binding between elements (e.g. by simulating an occluder, such that the signal appears to be passing under apertures).

Equivalent noise

Subjects JG and PB could only perform at a very poor level ($\sigma \sim 55^\circ$) in the four elements condition suggesting that they were unable to 'bind' the spatially disparate 1D motions together (an interpretation supported by subjects' comments). This idiosyncratic behaviour (two subjects were able to perform, two were not) is not entirely surprising considering the inherent ambiguity of the stimulus; i.e. there is no information in the stimulus that

conclusively determines whether elements should be bound or not. In this regard the stimulus is theoretically similar to previous work (Lorenceanu & Shiffrar, 1992; McDermott, et al., 2001) that explores the inherent ambiguity in stimuli composed of locally ambiguous motions whose overall configuration is consistent with a two-dimensional motion. Like mine, this study used four 1D elements that could be perceived either as independent motions or as a coherent global motion. Studies using such stimuli reveal that the percept of motion is bi-stable and may be altered by non-motion cues (McDermott, et al., 2001). Moreover a comparable study by Amano et al. (2009) went to great lengths to ensure that the individual Gabor elements were integrated as a whole, by lowering the contrast of the stimuli (Lorenceanu & Shiffrar, 1992) and presenting the Gabors in the parafovea (Takeuchi, 1998). It has also been noted that stimuli containing a large number of orientations stabilises the percept' (Kaoru Amano, et al., 2009). Clearly the ability to bind spatially locally 1D motion cues is a limiting factor when using global-Gabor stimuli and the aim of this section is to examine observers' sampling efficiency and to determine whether poor sampling efficiency may account for the failure of observers' JG and PB to perform in the 4-element condition. We also examine whether the estimates of observers' sampling efficiency is correlated with observers' self-consistency. This is of interest, because sub-optimal sampling efficiency could be due to a random or predictable under-sampling of the stimulus (e.g. just using the Gabor elements near fixation).

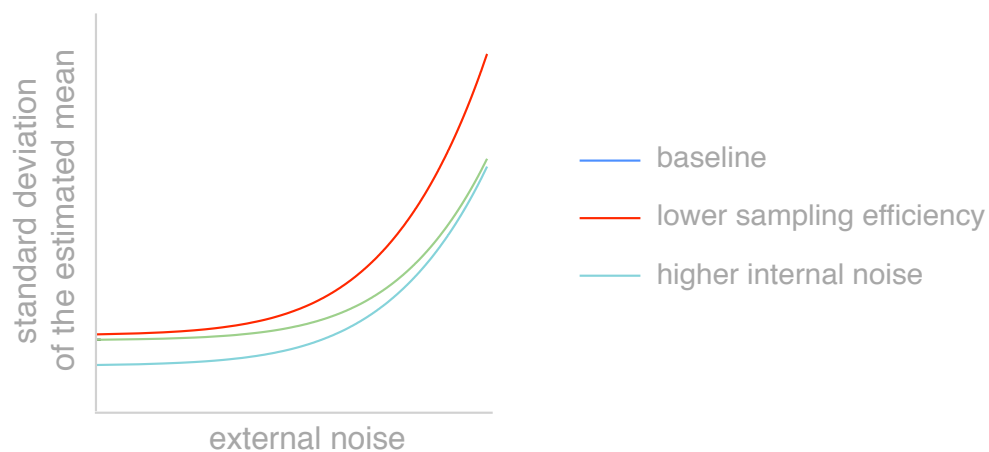
The notion that sampling efficiency (the number of motion samples used to calculate direction) may vary between subjects is plausible given the above discussion, but the interpretation is not consistent with the literature on the integration of elements in random-dot-displays (Dakin, et al., 2005b). In this work, the sampling efficiency of observers was found to be roughly equal to the square root of the total number of samples present in the stimulus. However the use of random dot stimuli is different from locally ambiguous line stimuli; in isolation the latter are ambiguous within $\pm 90^\circ$ and accordingly there may be a greater requirement to integrate information across space. Furthermore recent research has supported the idea that 1D and 2D signals are processed in a different manner by the motion-processing stream (Kaoru Amano, et al., 2009; Bowns & Alais, 2006).

The hypothesis that observers PB and JG integrate over fewer elements than observers DK and SD cannot be directly tested on an individual measure of a system's variability because both internal noise and sampling-efficiency cannot be uniquely specified from a single threshold data point. An equivalent noise (EN) technique can be used to get around this problem, the technique works by adding controlled levels of external noise (designed to mimic the effect of additive internal noise) into a system. By doing so the variability of a system can be measured at various noise levels to generate an equivalent noise function; as the influence of internal noise is *additive* but the influence of sampling efficiency is *multiplicative* both parameters are needed to uniquely specify an equivalent noise function.

To illustrate this point [Figure 4-8](#) plots three EN functions defined by the standard EN equation on *loglog* axis. The influence of internal σ_{int} external noise σ_{ext} and sampling-efficiency n on the variance of a system can be described by Equation 4.3 for simple averaging operations.

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

Equation 4.3



[Figure 4-8](#) Equivalent noise functions as defined by Equation 4.3. The blue is a baseline function from which we can compare the influence of increase internal noise (green line) and of decreasing sampling efficiency (red line). The impact of the sampling efficiency term is multiplicative and shifts the function up or down. The influence of changing internal noise is additive and shifts the function at low external noise levels, with performance converging at higher noise levels.

If the blue line is used as a standard, then we can compare the role of increasing the internal noise (green line) and decreasing sampling efficiency

(red line); increasing the internal noise of a system increases the system's variability at low-external noise levels but not at high. The impact of decreasing the sampling efficiency is to shift the function upwards, i.e. uniformly elevating thresholds.

Methods

Subjects/Apparatus /Procedure/Stimuli

Subjects, apparatus and stimuli were identical to the previous experiment (the number of elements was four for SD, DK performed in the both the four and sixteen element conditions and JG and PB used sixteen elements) except that four external noise levels were applied to the stimulus. The noise source was Gaussian-distributed directional variability and the pseudo-random noise distributions added to each stimulus had standard deviations of 0, 16, 32 and 64°. The noise was independently added to each element regardless of the number of elements.

Ideal observer

The standard equivalent noise equation used in many psychophysical studies is inappropriate for the present experiment because the model does not incorporate stimulus-constraints associated with the 'aperture problem'. Accordingly, I define an ideal-observer by adapting the geometric Solution described in the introduction. The geometric solution defined previously was ideal for noise defined evenly in the two dimensions of Velocity space. This section adapts the technique to be optimal for noise defined in just the direction dimension.

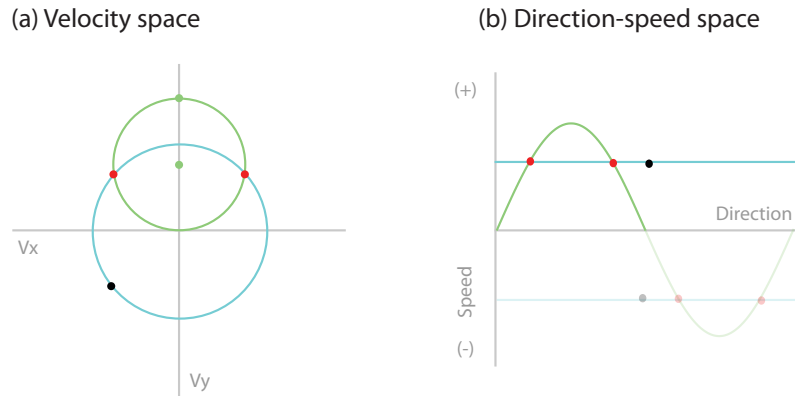


Figure 4-9 Schematic of an ideal observer that solves the 'aperture problem' for noise defined in direction only. The solution is derived from the geometric solution presented in the Introduction and is presented in both (a) Velocity space and (b) in the Speed-Direction space. The solution first attempts to estimate where on a potential 2D motion a 1D motion (black dot) may have arisen. As the noise is defined only in the direction dimension the problem reduces to finding the 1D directions on the cosine at the same speed as the 1D motion. In Velocity space this involves searching on a circle, whilst in the Velocity-Speed space this simplifies to searching along a line. In either case the cosine will be bisected in two places and the shortest distance is taken as the estimated 1D motion.

For noise defined purely in the direction-dimension, the ideal-observer should know the speed, but not the direction of the 1D velocities. For each data point (with a given speed) I can ask which orientations (from a potential fit) are likely to have given rise to these 1D motion. This is depicted in Figure 4-9 in Velocity space (a) and the Direction-Speed (b) space. In the Direction-Speed space the potential directions consistent with a 1D motion at a particular speed, correspond to the points of intersection between a line along the direction dimension (at the speed of each 1D motion) and a cosine defining the global motion in question. The line will bisect the cosine twice; the shortest angular separation is taken as the estimated 1D fit. This is formally defined below;

As we know the speed of each local element (ϕ_{local}) we can calculate the potential relative-orientation ($\theta_{relative}$) for a given global speed (ϕ_{global}) as below;

$$\theta_{relative} = \pm a \cos(\phi_{local} / \phi_{global})$$

Equation 4.4

By adding in the direction of global direction (θ_{global}) motion the relative-orientations can be converted into local directional signals.

$$\theta_{localEst} = \theta_{global} \pm \theta_{relative}$$

Equation 4.5

The angular separation between the 1D motion estimate and each 1D motion is taken

$$\theta_{sep} = \theta_{localEst} - \theta_{local}$$

Equation 4.6

and the minimum distance found

$$\theta_{min\ sep} = \arg \min(\theta_{sep})$$

Equation 4.7

Finally the root-mean-square error of the the minimum separation is calculated

$$rms = \sqrt{\frac{\theta_{\min \text{ sep}}^2}{n}}$$

Equation 4.8

To fit the psychophysical data, equivalent noise functions were generated from simulations of the experiments across sampling-efficiencies from 2 to 16 at 0.1 intervals and for internal noise standard deviations of 0° to 50° in one-degree intervals. For each point on the EN functions 10,000 sample trails were used and the variance of the system calculated for the resulting population of errors.

The fitting procedure for each EN fit minimised the root-mean-square difference between the psychophysical and ideal observers' variance across the entire EN function.

Results

Equivalent Noise

The estimated *internal noise* and *sampling efficiency* of the each subject is shown in [Figure 4-10](#). Equivalent noise functions were fit to both the even and random orientation conditions simultaneously to reduce the variance on the estimates equivalent noise fits, but individual fits revealed no significant differences. The results from the four-aperture condition are shown as circles and results from the sixteen-aperture condition are shown as squares. The estimates of internal noise are relatively consistent across subjects and conditions and there is a greater degree of variability in the *sampling*

estimates for our observers. Sampling efficiencies for DK and SD in the four-aperture condition are high at around ~80%. The sampling efficiency drops (but not significantly) for subject DK in the sixteen aperture condition. Sampling efficiencies for subjects PB and JG are lower (significantly so for JG). This suggests that the failure of PB and JG to achieve good performance on the four aperture condition was due to poor sampling efficiencies leading to a catastrophic failure of integration; in the limit, integration of one signal and simply estimating the two-dimensional direction orthogonal to a Gabor's orientation would perform with a precision of $\sigma \sim 54.4^\circ$ (standard deviation) without external noise. This value is considerably more than the internal noise estimates of any of our subjects ([Figure 4-10](#)) indicating that the noise was not just a function of a subject's internal stochastic noise (e.g. neural noise) but was exacerbated by an inability to overcome the ambiguities associated with the 'aperture problem'.

In [Figure 4-10](#) (bottom right) I plot the estimated sampling efficiency for all subjects against the R-Scores on the first and second pass. As the estimates of sampling efficiency were made across all conditions, the errors from all the conditions were appended before calculating the double-pass R-Scores. The results show that the R-score increases with increasing sampling efficiency. This indicates that imperfect sampling efficiency was a key source of stochastic variability in observers' data. This is of interest because it indicates that the imperfect sampling efficiency was not due to observers' preferentially

intergrating across some elements (e.g. near fixation) but was due to a random process.

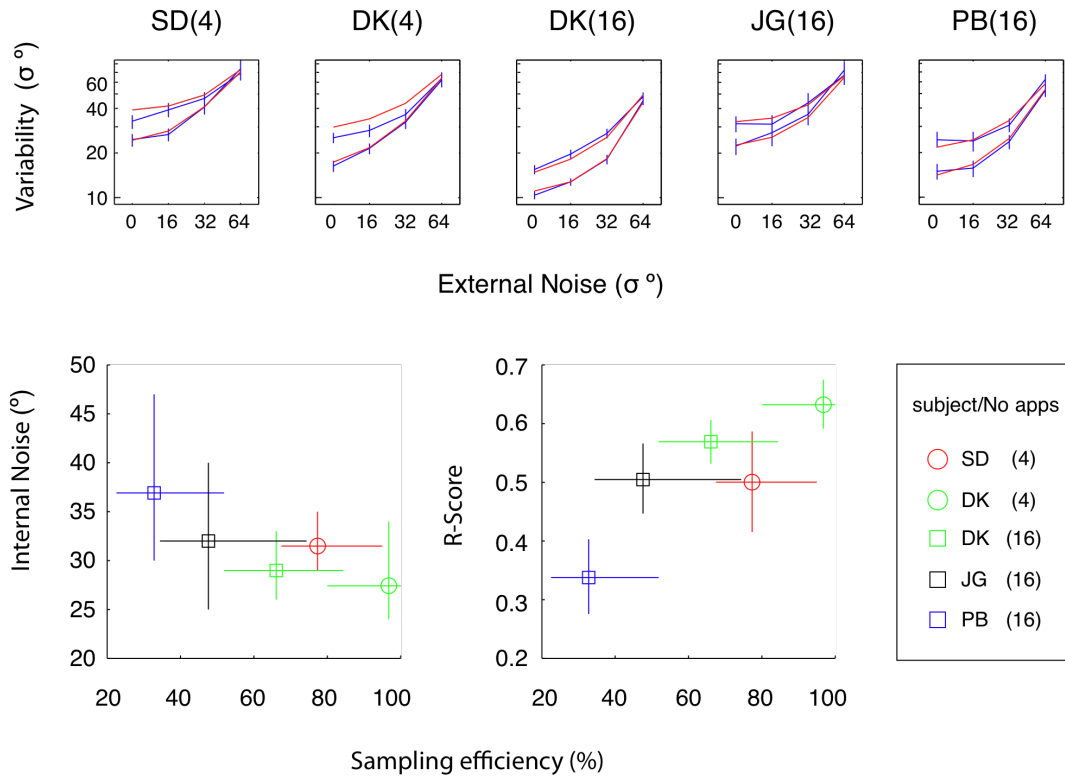


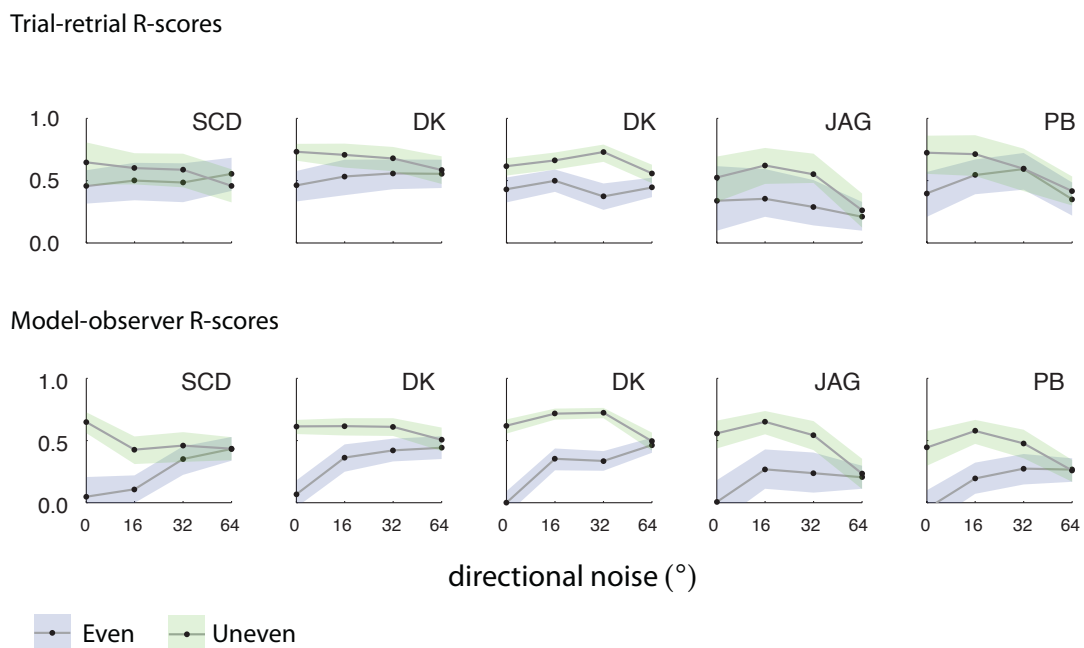
Figure 4-10 In the top row I plot the observers' EN functions (blue) and the ideal observer fits (red). The estimates of sampling efficiency and internal noise are shown in the bottom left and the sampling efficiency is plotted against the R-Scores in the bottom right. All error bars are 95% confidence intervals. The internal noise estimates are relatively stable across subjects at around 30° . Data for DK and SCD on the four elements condition show high sampling efficiencies ($\sim 80\%$). The sampling efficiency for DK is reduced in the sixteen-element condition ($\sim 66\%$). The sampling efficiency of subjects PB and JAG is less than for DK (significantly so for JAG). This suggests that the inability of PB and JAG to perform on the four-element condition was due to a lack of binding between elements. The estimated sampling efficiencies are correlated with the observers R-scores (collapsed across all conditions). This suggests that imperfect sampling efficiency is a key source of stochastic variability in observers' data.

Results

Testing the 2D model

The template-matching model presented in the preceding chapter was able to capture observers' *first* and *second* moments of their error distributions but was unable to predict observers' errors on a trial-by-trial basis. This may have reflected a number of factors (e.g. contrast gain; Rust, et al., 2006) not incorporated into the model or alternatively it may be that the majority of observers' response variability was stochastic, and thus unpredictable. The data in the natural scenes chapter is not sufficient to separate these two hypotheses. The present experiment uses a more constrained stimulus class to test the model in which the experimenter has direct control of all aspects of the stimulus. To estimate the proportion of variability that is *stochastic* and that which is *stimulus-led* a double pass technique was employed and the correlation between observers' performance when retested with the same stimuli was used to estimate the degree of *predictable* and *unpredictable* variability. The R-scores are shown in the first row of [Figure 4-11](#) as function of the external directional noise added to the stimulus. The results show that R-scores are always higher for the *random* orientation condition at low external noise levels but that this effect drops off at high external noise levels. The R-scores for the model-observer correlations are shown in the second row of [Figure 4-11](#). The R-scores in the *random* orientation condition appear very similar for both the *double-pass* correlations and the *model-observer* correlations, suggesting that the model can capture the majority of observers' predictable variability in this condition (the ratio between the R-

scores is plotted in [Figure 4-12](#). However, the model is unable to capture as much variability in the *even* orientation condition particularly at low external noise levels; this is unsurprising because the (noiseless) model makes no errors in the *even* orientation condition.



[Figure 4-11](#) R-scores for the observer on stimulus retrials (first row) and for the model and observer correlations (second row). The data for the even condition is presented with blue 95% error bars whilst the data for the uneven condition is presented with green error bars. The data show that for both the model and observer correlations, R-scores are higher for the uneven condition at low-external noise levels but that the effect drops away at high external noise levels. The effect is much more pronounced in the model-observer correlations primarily due to the fact that the model is unable to capture any of the data in the even condition at low-external noise levels (as the model produces no errors). In contrast, R-scores for the retrials and model-observer correlations appear very similar across the entire EN-function.

To examine the data further, [Figure 4-12](#) plots the ratio of the R^2 score of the observers' error correlations on stimulus retrials over the R^2 of the model-observer correlations. In [Figure 4-12\(a\)](#) the ratio of R^2 is plotted for the *even* orientation condition and in (b) for the *random* orientation condition, error

bars are excluded for clarity. To provide statistical verification of the pattern of results a straight line ($y=mx+c$) was fit to the data and estimates of both the constant (c) and gradient (m) are plotted in [Figure 4-12\(c\)](#). The plotted fits are to the original data and the 95% confidence intervals are estimated by bootstrapping ($n=1024$) the data with replacement from the full distribution of errors. The scatter plot depicts the gradient fit on the x-axis and the constant fit on the y-axis; fits in the *even* orientation condition are blue squares and the *random* orientation conditions are green circles. [Figure 4-12\(a\)](#) demonstrates that the model is not able to capture any of the stimulus-led variability at low-external noise levels but is able to capture a proportion of the variability at high-external noise levels. The corresponding fits of a straight line show a gradient significantly greater than zero for all subjects thus statistically verifying this finding. The constant is not significantly greater than zero, verifying that the model is unable to capture any data when no external directional noise is added to the stimulus. In contrast to the *even* orientation condition, the model is able to capture data at all external noise levels in the *random* orientation condition ([Figure 4-12b](#)) and the fit of a straight line ([Figure 4-12c](#)) demonstrates the proportion of variance captured is significantly greater than zero for all subjects and has a mean of around 0.8. The proportion of variance captured does not vary significantly as a function of the external noise as the gradient is never significantly greater than zero,

As the proportion of deterministic variability is not constant across conditions ([Figure 4-12d&e](#)), the R^2 score is arguably an unfair assessment of the model's

behaviour; at the extreme if there is no deterministic noise in observers' variability then the R-score of the model will also be zero, this logic can also be applied to the R^2 ratios. To overcome this criticism, [Figure 4-12](#) d&e plot the absolute stimulus-led variability the model is *unable* to account for. There appears to be no overall effect of adding external noise in either the even orientation (c) or random (d) orientation condition and the fit of a straight line (e) reveals the gradient is never significantly different from zero. The constant fit effectively states the degree of variability the model is unable to account for and is between zero to six degree of variability. This is encouraging because it demonstrates that the proportion of predictable variability not accounted for the model is constant across conditions. This demonstrates that the model is able to captures all the additional variability caused by *randomising* the orientation structure and adding external *directional* noise.

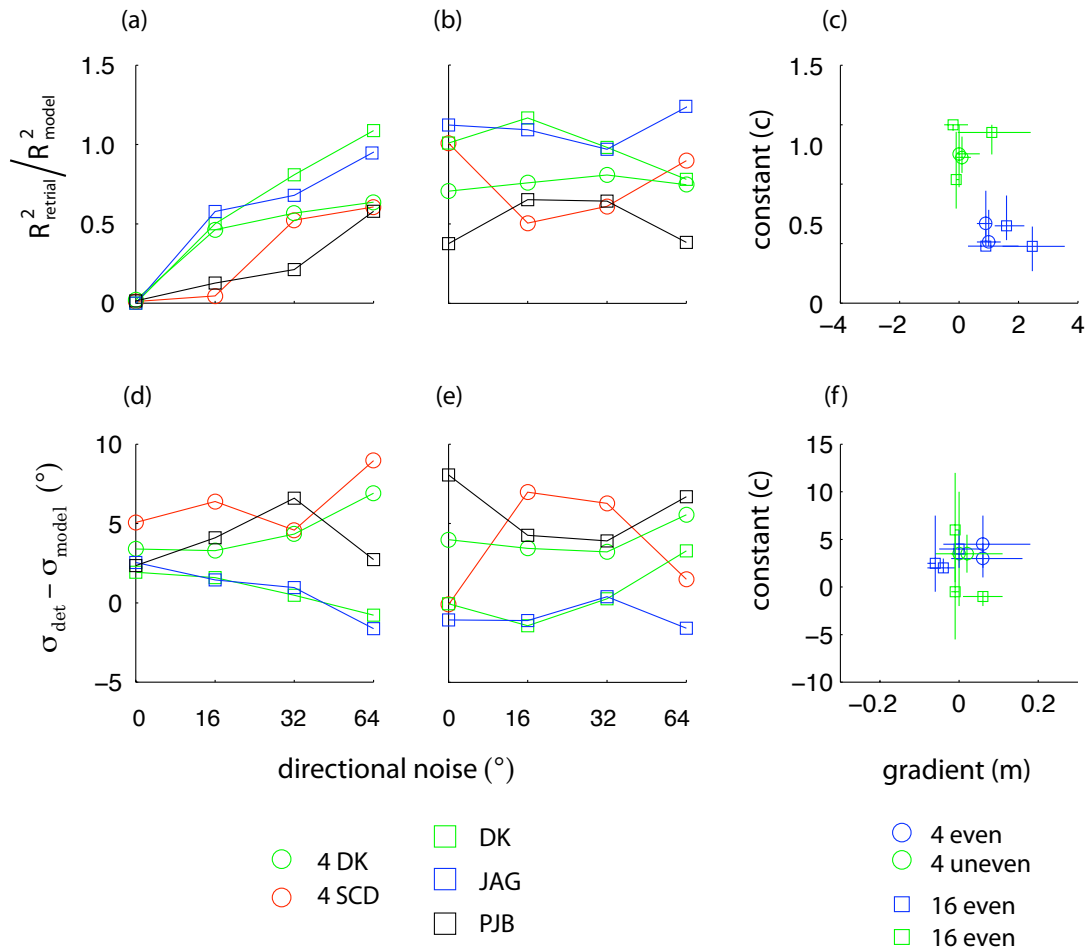


Figure 4-12 The ratio between the trial-retrial R-scores and the observer-model R-scores in the (a) even condition and (b) uneven condition. The total predictable variance unaccounted for by the model in (d) even condition and (e) the uneven condition. Although the ratio between the R-scores varies with the condition the total predictable variance unaccounted for by the model appears relatively constant, this suggests that the model is able to capture all addition variability due to adding external noise or randomizing the orientation structure of the stimulus.

Discussion

The aim of this chapter was to further assess the model presented in the proceeding chapters; a model that was able to capture observers' *first* and *second* moments of their error distributions but not the trial-by-trial variability. It was proposed that this may either reflect factors not incorporated into the model (such as contrast gain or sensor anisotropies) or that the majority of the variability in the data was stochastic. The paradigm in this chapter does not definitively answer this question, but it does provide a constrained test of the model in which all stimulus parameters are under direct experimental control and is able to estimate the degree of predictable and unpredictable variability in the error distributions via a double-pass technique. The approach of estimating the degree of predictable variance in the data has been used in neurophysiology (e.g. David & Gallant, 2005; Hsu, et al., 2004; van Hateren, et al., 2002) as a means of testing well-developed models of retinal and LGN processing. Such tests are considered a strong test of a model when applied to natural scenes; if a model can account for all the data in the rich environment of a natural scene (that the system has evolved to process), then the full functionality of data has been captured. The present paradigm was not applied to natural scenes but it does demonstrate the model's predictive power in estimating the additional variability due to external noise being added to the stimulus (arguably not errors, but estimates) and due to the randomisation of the orientation structure. The next stage then must be to test the model in response to natural scenes. In this respect, it should be noted that the degree of correlation between the observers' error on stimulus

retrials was poor in the *even* orientation condition (only ~25% of the variance was captured by observers' retrial data) and higher (~50%) in the *random* orientation condition. Due to the high orientation bandwidth of natural scenes, it is likely that the degree of stimulus-led variability will be closer to the *even* orientation condition than the *random* orientation condition and suggests that the inability to predict observer errors on a trial-by-trial basis was due to the low percentage of stimulus-led variability. Accordingly, if one wants to extend the work to natural scenes it would be necessary to increase the proportion of stimulus-led variability, perhaps by decreasing the number of apertures or the area of each aperture. However this raises an additional issue since it is likely that the high proportion of stochastic variability is primarily a function of incomplete sampling efficiency rather than internal stochastic factors such as neural noise. To elaborate, I have been able to demonstrate that altering the orientation structure of the stimulus is able to increase observers' errors even when there is sufficient information to solve the 'aperture problem'. Comparable errors are also likely to arise if the sampling efficiency is incomplete because orientation imbalances will occur even when the orientation of the Gabor element is evenly distributed. Such errors will be unpredictable if the sub-optimal sampling of a system is itself unpredictable but not if the sub-optimal sampling efficiency were predictable. This is a question that merits further investigation. For instance if observers' gave a greater weighting to those elements that occurred near fixation, this would increase the stimulus led variability of the system, not the stochastic variability. This question may be approached by applying a reverse

correlation paradigm that related the underlying stimulus, not to observers' errors, but to observers' trial-retrial correlations.

It should be noted that an attempt was made to explore the role of sampling efficiency and internal noise for global-Gabor arrays composed of either i) elements whose local 1D velocities did conform to a 2D vector or ii) those in which the speed and direction of Gabor elements was scrambled, however the equivalent noise estimates in the scrambled condition were so poor (due to a shallow EN function) that it was impossible to draw any conclusions.

(5) Conclusion

Summary

The thesis contained three experimental chapters and one computational chapter. The approach taken in the first two experimental chapters was to incorporate more naturalistic statistics into psychophysical paradigms that explore two-dimensional motion processing. Both chapters are motivated by the broad hypothesis that the visual system can only be fully understood in terms of the natural environment that it has evolved to process (Simoncelli & Olshausen, 2001). To this end the first chapter probes the role of 'natural' contour statistics in motion processing. Two aspects of natural contour statistics are identified that may influence the integration of local motion signals across space. The first is the spatially broadband and phase-aligned nature of natural contours (Attneave, 1954; HB Barlow, 1961) and the second is the second-order structure of contours across space (Geisler, et al., 2001). My results reveal that breaking the natural contour statistics does increase observers' thresholds in a two-alternative forced choice (2-AFC) direction discrimination task, but only when low spatial frequencies are present. Application of the motion energy model (Adelson & Bergen, 1985) across spatial-frequencies demonstrates that the low-frequency component of the signal is made more variable when the orientation statistics across space are scrambled. This demonstrates that the motion stream is unable to ignore the low-spatial frequency component of the signal, consistent with previous research (Bex & Dakin, 2002) and may reflect the fact that spatial frequency broadband integration is needed to recover speed information from sensors

that are sensitive to spatiotemporal frequency, not speed *per se* (Perrone & Thiele, 2001, 2002; Priebe, et al., 2006). Conversely, the work also demonstrates that disrupting the second-order relationship between each element of the high-pass stimuli did not influence observers' ability to judge the 2D motion of the contoured stimulus. This finding contradicts a finding by Lorenceau and Alais (2001) who demonstrate that the spatial arrangement of features can influence observers' ability to estimate 2D direction. The task in the two studies is broadly equivalent (both 2-AFC direction discrimination tasks) but the stimuli used in the two studies are different across a number of behaviourally relevant dimensions including the *number* (Dakin, et al., 2005a) and *density* of elements and the ambiguity of the 1D velocities (Kaoru Amano, et al., 2009). To elaborate I will first recap the stimuli configuration in the Lorenceau and Alais (2001) paper: The stimulus employed was a rotating stimulus composed of four moving bars (an example stimulus is shown in [Figure 2-1](#)) that were viewed through apertures such that only the 1D velocities were exposed to the observer. Theoretically the stimulus may be correctly perceived as moving either as a coherent structure rotating through space or as individual 1D velocities. The second-order spatial configuration of the elements was manipulated by changing the bar-orientations; the results showed that for 'closed' shapes such as a diamond observers' were better able to determine the coherent 2D motion than for 'open' stimuli such as a chevron. This data can be interpreted in two ways; one interpretation argues that that some spatial arrangements improve the degree of integration between local elements, which in turn improves threshold in a discrimination

task. However the research in this thesis suggests a different interpretation; when Lorenceau and Alais (2001) manipulated the second order structure, they also altered the *local* distribution of motions. In other words, changing shape alters the orientation of each element relative to the 2D motion vector and in turn alters the 1D velocity distribution. The research in this thesis has demonstrated that the orientation of elements relative to the underlying global motion plays an important role in determining 2D motion percepts; as such, the only way to isolate the influence of shape on motion is to manipulate the two factors independently.

Despite the above critique there is good reason to expect that shape/form is likely to play a role in the stimulus class employed by Lorenceau and Alais (2001) as the percept of such stimuli are notably bi-stable (McDermott, et al., 2001). Under these conditions, the inherent ambiguity of the local motion is much greater. This raises the broader question of how locally 1D and 2D signal are processed in the motion stream; recent research in both the psychophysical (Kaoru Amano, et al., 2009; Bowns & Alais, 2006; Lorenceau, et al., 1993) and neurophysiological (Majaj, et al., 2007) literature has suggested the motion stream may dynamically alter the nature of 2D motion estimation with respect to the ambiguity of local motion signal. Psychophysically, it has been demonstrated that locally 1D motion signals may be integrated in a manner that allows for the correct estimation of the *speed* and *direction* of motion under a number of conditions (K. Amano, et al., 2009; Lorenceau, et al., 1993), whilst the perceived speed of a series of

locally 2D elements is consistent with an averaging rule (Kaoru Amano, et al., 2009). This research supports the notion that the manner in which a local signal is processed depends on the relative ambiguity of the signal. In this respect, I present two lines of evidence that support the notion that 1D and 2D signals are subject to different patterns of integration; firstly the study probing the role of *number*, *density* and *area* for the contoured stimulus class can be contrasted against a study by (Dakin, et al., 2005a) who probed the role of *number*, *density* and *area* using band-pass filtered dots. The local ambiguity greatly differs in the two stimulus classes. While direction discrimination thresholds for individual contoured elements was at best $\sigma \sim 25^\circ$, but the ability to discriminate the 2D direction of an individual dot passing through noise is better than $\sigma = 3^\circ$ (Watamaniuk & McKee, 1998). In my data, increasing density was shown to improve performance in the 2-AFC task. In contrast, the work of Dakin et al demonstrate that increasing the density of elements led to a small increase in both sampling efficiency and internal noise (attributed to correspondence noise) estimates with equivalent noise analysis when the task was to integrate band-pass dot stimuli. Increasing sampling efficiency improves performance while increasing internal noise impairs performance, so the overall impact on thresholds of these competing effects was minimal. Contrasting the two experiments suggests that density plays a stronger role in the integration of locally 1D stimulus than locally 2D stimuli.

A second line of evidence in favour of differential processing of 1D and 2D motion signals comes from the data on sampling efficiency. In Dakin et al. 's (2005a) study, the sampling efficiency of band-pass random dot stimuli was approximately the square root of the total number of elements. In contrast, in the present data, the sampling efficiency was idiosyncratic but nearly always higher than the square root of the total number of elements, approaching 100% in the four-element condition and greater than four-elements in the sixteen-element condition (for two out of three subjects). This suggests that either that the square root law does not hold at low elements densities or that the sampling efficiency can be greater when the local elements are composed of 1D signals rather than 2D signals. To examine this question an additional experiment is needed in which the stimulus is smoothly varied between 1D and 2D, as this would allow for all experimental factors to be controlled and a definitive conclusion to be made.

Edges oriented oblique to the 2D motion vector produced both biased and highly variable responses from observers. In contrast, elements oriented orthogonal or parallel to motion led to unbiased and relatively precise responses with low response variability. This finding is consistent with a study employing translating lines (Loffler & Orbach, 2001); when the orientation of the translating line was oblique to the direction of motion, the majority of motion estimates were biased towards the direction orthogonal to the contour, but a minority were biased or in the opposite direction. The final population of errors was bi-modal with the two peaks of responses either side

of the veridical 2D direction. The present analysis was extended to examine the influence of pairings of orientations in natural scenes. This approach allowed the results to be compared to studies that only use two orientations to study the 'aperture problem' (e.g. Yo & Wilson, 1992). Broadly speaking, the results are consistent with the literature using constrained stimuli with just two orientations (Kaoru Amano, et al., 2009; Bowns, 1996; Mingolla, et al., 1992; Rubin & Hochstein, 1993; Wilson & Kim, 1994; Yo & Wilson, 1992); in both paradigms observers are biased towards the direction of local motion for Type II combinations of orientations. The present work extends the literature on the 'aperture problem' to detail observers' bias and variability over the full range of Type I and Type II orientation combinations; this analysis also reveals that observers are biased towards the direction of the fastest component motion for Type I combinations. A more general finding of the present research is that although the pattern of observer bias was consistent with that reported in the literature using much more constrained stimulus classes, the magnitude of the bias observed in response to the natural scenes was smaller than that in response to Type II stimuli composed of just two oriented elements (Kaoru Amano, et al., 2009; Bowns, 1996; Mingolla, et al., 1992; Rubin & Hochstein, 1993; Wilson & Kim, 1994; Yo & Wilson, 1992). This suggests that the motion stream was relatively well optimized to process natural scenes and it is likely that the relatively broad orientation bandwidth of natural scenes aids subjects in making motion judgments. This pattern of responses is consistent with models of 2D motion processing that sum across all possible 1D velocities that are consistent with a global 2D velocity (Perrone, 2004; Schrater, et al.,

2000; Simoncelli & Heeger, 1998; Watson & Ahumada, 1983) because such models assume isotropy in the orientation structure of a motion signal. In order to successfully relate observers' errors to the natural scenes, some assumptions are needed about how to estimate image features. The approach taken was to use biologically inspired models of local orientation (Daugman, 1980) and direction (Adelson & Bergen, 1985) processing. Encouragingly, application of the motion energy model revealed that the *cosine* pattern of 1D velocities as a function of *speed* and *direction* (predicted by Equation 1.1) was well captured by a bank of filters tuned across a range of different *directions* and *speeds*. This motivated a model of 2D motion processing based on the generation of templates for each 2D motion from the interaction between natural scenes and the responses of a standard motion-energy model. Application of the templates revealed that the model was able to capture the first (bias) and second (variability) moments of observers' error distributions in the natural scenes paradigm. Somewhat surprisingly, the model was unable to capture a high percentage of observers' variability on a trial-by-trial basis so an additional experiment was designed to directly test the model under more constrained conditions. The paradigm used a global-Gabor array in which the orientation of each element was under direct experimental control. Two conditions were tested under a number of external (directional) noise conditions; one in which the local orientation structure was *evenly* spaced and another in which the orientation distribution was *randomly* distributed. The results showed that performance was worse in the *random* orientation condition, consistent with

observers' being biased by imbalances in the local orientation structure of the scene. The paradigm included a double-pass technique designed to estimate the degree of stimulus-led and stochastic variability in the observers' response. The analysis showed that the increase in observers' response variability in the *random* orientation condition was stimulus-led and thus predictable. Accordingly, a plausible model of 2D motion processing should be able to predict the additional errors produced in the random orientation condition. Correlations between the observers' errors and the model of 2D motion processing were broadly consistent with the double-pass correlations in the *random* orientation condition, but were notably poorer in the *even* orientation condition. The level of variability the double-pass correlations allowed me to estimate how much stimulus-led response variability the model could be expected to capture and the residual unexplained variability. The analysis revealed that the residual variability the model could not capture was consistent across both conditions and all external noise levels. In other words, the model could capture all the additional variability induced by adding direction noise to the stimulus or by randomizing the orientation structure of the stimulus.

Biological plausibility

The argument that the 'aperture problem' is solved by a system which integrates across a pattern of local (1D) motion that is consistent with uni-

directional/rigid two-dimensional motion (i.e. a template model) was proposed by Watson & Ahumada (1983) who demonstrated that an iso-oriented object rigidly translating through space generates a plane in the spatiotemporal frequency domain (a class of model known as the F-plane model; Bradley & Goyal, 2008). This notion gained empirical support from Simoncelli & Heeger (1998) who demonstrated that a number of response properties of MT 'pattern selective' cells are consistent with a model that integrates across a plane in spatiotemporal frequency space. The *F-plane* model has also received psychophysical support in that masked detection is best when signal energy is evenly spread across a plane in spatiotemporal space (Schrater, et al., 2000). Simoncelli & Heeger (1998) went to lengths to point out that their model of motion processing in MT does not encode the 2D velocity of objects in each cell, but it is a population model in which the pooled responses of a series of complementarily-tuned MT cells encode for the global (2D) velocity (i.e. the output of an individual sensor is not enough to determine 2D motion).

It is worth noting at this stage that the receptive field profiles of the motion sensors in the Simoncelli & Heeger model (1998) are similar to receptive field profiles I derived from the interaction between the motion energy (Adelson & Bergen, 1985) model and natural scenes. I plot the weighting functions for both models in the Speed-Direction space in [Figure 5-1](#). The Simoncelli & Heeger model (1998) of MT 'pattern selective' cells works by calculating the (shortest) distance between the spatiotemporal frequency tuning of each sensor and a plane which defines the 2D velocity tuning of an MT sensor. The

weighting of each local sensor is inversely proportional to that distance, Simoncelli & Heeger (1998) leave the exact function open to be constrained by neurophysiological data. To compare the receptive field properties of the Simoncelli & Heeger model (1998) to the model presented in this thesis, I use the same bank of local filters throughout. The distance of each sensor with a spatial frequency profile (sf_x, sf_y, sf_z) is relative to a plane which defined the 2D motion (v_x, v_y) where $s = \sqrt{v_x^2 + v_y^2}$ is defined below.

$$d(sf_x, sf_y, sf_t) = \frac{|v_x sf_x + v_y sf_y + s_t sf_t|}{\sqrt{v_x^2 + v_y^2 + s_t^2}}$$

Equation 5.1

The distance was then converted to a weighting function by passing it through a Gaussian function.

$$w(sf_x, sf_y, sf_t) = e^{\left(\frac{-d(sf_x, sf_y, sf_t)}{2\sigma^2}\right)}$$

Equation 5.2

The weighting functions for both models are shown in [Figure 5-1](#) as a function of both the *speed* and *direction* of local motion, the standard deviation of the Gaussian function was hand chosen, nonetheless it is clear that both models produce a similar *cosine* pattern of weighting in the Speed-Direction space.

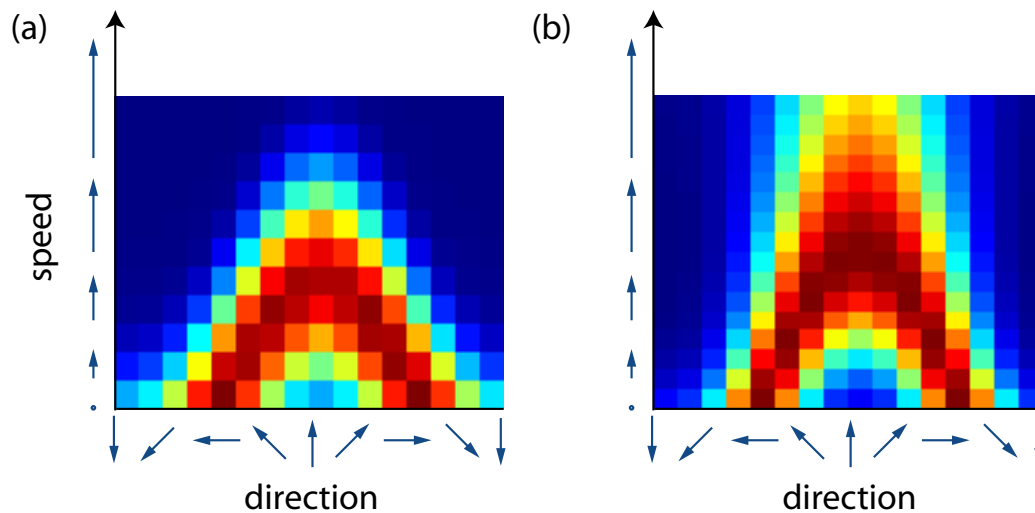


Figure 5-1 2D Motion Templates (a) motion energy profile derived from the response of the motion energy model to a rigidly translating natural scene at 1 pixel per frame (b) the receptive field profile of the Simoncelli & Heeger (1998) plotted using the same underlying filter configuration as in (a) and used throughout the thesis.

Model limitations

The model described in this thesis has been designed to model the data to hand i.e. compute 2D velocity from (effectively) a full field motion stimulus. This is a constrained task and I now discuss the limitations of the model in the context of natural vision. Although I tested the model upon natural scenes, all motions were confined to the fronto-parallel plane and a multitude of problems associated with natural vision were not tested. Moreover, the model was provided with information about the location of the apertures when modelling the psychophysical data. The psychophysical data is consistent with observers' being able to ignore the signal stemming from the apertures and the question of how this is achieved is open to question. It is possible that mechanisms of attention may allow observers' to selectively monitor the spatial region of integration (Burr, Baldassi, Morrone, & Verghese, 2009) or a

mechanism to remove the static signal may be inherent to the motion stream (Johnston, et al., 1992) or may result from the classification of visual stimuli as 'external' or 'internal' to the motion at hand (Shimojo, et al., 1989). In any case, the data presented throughout this thesis throws no light on this issue.

Other major motion issues the model is not designed to and will not cope with include the segregation of multiple moving objects. To accurately compute 2D velocity in the experiment used in this thesis the visual system integrates across space. This poses a particular difficulty in natural vision because of the risk of integrating across motion signals belonging to different objects, a problem known as superposition catastrophe (Lorenceanu, Giersch, & Series, 2005). In theory the present model may be extended to allow the extraction of multiple moving objects through a localised summation of activity from the 1D stage to the 2D stage and may also be expanded to allow the detection of transparency by allowing the read-out strategy to pick multiple winners. However these are large problems in their own right and are not examined in this thesis.

Predictable errors

Both the *F-plane* and cosine models are defined for 1D velocities which are defined discretely, in which case defining two or more points either by their *speed* and *direction* or their *spatiotemporal frequency* in \mathbf{x} , \mathbf{y} , and \mathbf{t} would always allow for a veridical solution to be found. However it is unlikely that any system that attempts to derive local motion signals from complex natural

environments would be able to derive noiseless estimates; specifically the primate visual system uses a wavelet-style system (Adelson & Bergen, 1985; Daugman, 1980; Gabor, 1946) to derive estimates of local motion from the environment (detailed extensively in the Introduction) and such sensors are known to achieve a compromise between spectral sensitivity and spatial localization (Daugman, 1980; Gabor, 1946). This compromise means that all encoded signals are subject to the limits imposed by a processor with a finite bandwidth, i.e. a single *orientation* or *direction* will maximally stimulate a sensor tuned to that stimulus parameter but will also activate sensors tuned to nearby parameters on each stimulus dimension. Theoretically, in a noise free environment, the veridical stimulus parameter could be recovered if a stimulus were defined as a discrete point along a single dimension (e.g. the orientation of a straight edge). However as soon as the complexity of the stimulus increases, the computational requirements to solve the problem increase exponentially (the curse of dimensionality). In complex natural environments, the problem is sufficiently difficult that a more general and less accurate solutions is required. Furthermore, when sensory noise is incorporated, the effective resolution of a system drops, in turn increasing the ambiguity of any sensor's response.

Psychophysics has long studied this class of problem and human observers exhibit a number of non-optimal patterns of responses when observers are asked to identify stimulus features along two or more relevant stimulus dimensions. In terms of motion, the ability to detect and identify a stimulus composed of two or more motion velocities comes under the heading of motion transparency. In such studies, the aim is to understand what factors limit the observer's ability to detect multiple velocities and the results show

that the probability of detecting multiple motions increases with increasing separation in *speed* (Greenwood & Edwards, 2006a), *direction* (Braddick, et al., 2002) and well as *binocularity* (Greenwood & Edwards, 2006b) and *space* (Greenwood & Edwards, 2009). A related finding is that the perception of direction can be biased by simultaneously presented motions (Marshak & Sekuler, 1979). This phenomena is know as *motion repulsion* and is reduced when there is less overlap between the 1D velocities of dot stimuli (Curran & Benton, 2003). The literature on *motion repulsion* and *motion transparency* demonstrates that the motion stream is unable to correctly estimate motions that are close on some stimulus dimensions (motion repulsion) and may fail to recover a single estimate of 2D motion when the underlying 1D signals are too close (motion transparency). Application of this logic to the aperture problem suggests that subjects' misperceptions of motion are likely to result from an inability to correctly recover the velocity of 1D velocities, rather than a non-optimal 2D pooling strategy. Evidence in support of this hypothesis comes from studies in which the angular separation between two 1D velocities (in a Type II configuration) was systematically altered (Bowns, 1996; Burke & Wenderoth, 1993). The results show that it is only when the two motions are close in velocity space that misperceptions occur and it can be argued that the local motion interfere when they are close in velocity space.

Another model of global motion that is able to predict observers' bias under a number of conditions is the Bayesian model of Weiss et al. (Weiss & Adelson, 1998; Weiss, et al., 2002). The model is a modification of the Intersection of Constraints (IOC) solution detailed in the Introduction, where the lines of constraint are Gaussian blurred to represent the uncertainty originating from the finite bandwidths of motion-energy filters (Adelson & Bergen, 1985;

Daugman, 1980; Graham, 1989). This means that 2D motion is represented across a population of activity representing the likelihood of a given stimulus parameter given the sensory input. The distribution of likelihoods is multiplied with a Gaussian low-pass prior that draws the likelihood distribution towards slower speeds. This has the effect of drawing the final estimate of motion away from the correct solution towards the local component motions and towards the predictions of an averaging model.

The Weiss model is feed-forward and biologically plausible in a theoretical sense, but lacks empirical support because area MT is commonly believed to compute 2D motion but is not thought to compute *lines of constraint*. It should be noted that the 'noise' or error of the system occurs at a late stage in this model. The strongest criticism of the model concerns the concept of *optimality*. Bayesian models are predicated on the notion that the incorporation of a prior increases the probability of correct motion-estimates given the low-temporal frequency bias of moving natural images (Dong & Atick, 1995) and the possibility of errors due to neural or other noise sources. The concept of 'optimality' is under-constrained, however. For instance, the behavioral "pay-off" of the sign of misestimates is not established. Moreover the *prior* of the Weiss model was not optimized for processing dynamic natural scenes, but rather with respect to the data to hand (i.e. it has not been demonstrated that the prior does produce optimal estimates given natural movie statistics/noise). Without a means of estimating the behavioral and environmental context the claim of optimality is hypothetical (Geisler & Ringach, 2009) and as such the use of a *prior* runs the risk of simply adding an additional parameter allowing empirical data to be better fit.

One area where Bayesian models have found particular application is in understanding the role of contrast in motion processing. For instance, it is known that contrast may cause human subjects to see slower speeds when the contrast of the element is low (Thompson, 1982). An additional finding in the literature on plaids and the 'aperture problem' concerns how observers may be biased by changing the contrast of one the component grating in a Type I stimulus (Stone, Watson, & Mulligan, 1990). Under such conditions the observer is biased towards reporting the direction of the higher contrast grating. Under these conditions, the template-matching model described in this thesis will maximize the product of the 2D sensor and the motion energy distribution and would not (qualitatively) capture this percept if it were not for the bias towards low-speed noted in the modeling chapter (see [Figure 4-2](#) & [4-3](#)). This bias arises from the sum of the receptive field of each 2D sensor being approximately equal. Sensors tuned to greater speed integrate over a greater number of local motion sensors (i.e. a greater range of temporal frequencies) and thus have lower weightings for each spatiotemporal frequency within their receptive field. This has the property that if only one local motion is present (i.e. the motion of a straight edge) the output of the 2D stage is at the lowest temporal frequency, ie in the direction orthogonal to the edge's orientation, consistent with the human percept (Wallach, 1935). If the slow-speed bias were not present then the percept would also be unstable when noise was added to the model, because an infinite number of 2D percepts are consistent an individual local motion within an aperture (as illustrated in [Figure 1-3](#)). The bias for slow speeds also has the effect of drawing the estimates direction of motion towards the component with greater contrast as shown in [Figure 5-2](#). Although I criticize the slow prior in the Weiss model on theoretical grounds, the bias or prior is also present in the 2D

model presented in the thesis and can be derived from the interaction between the motion energy model (Adelson & Bergen, 1985) and natural scenes.

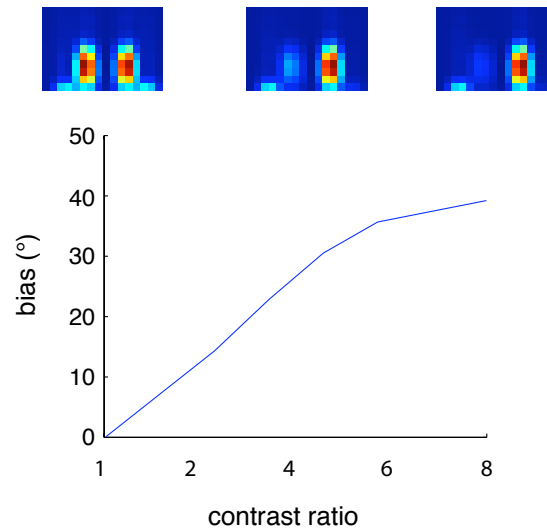


Figure 5-2 The bias of the template model for a Type I plaid with the local motion $\pm 60^\circ$ from the 2D direction; as the contrast ratio of the clockwise component compared to the anticlockwise component increases from 1:1 to 8:1 the direction estimate from model becomes biased towards the component with higher contrast

The concept of optimality is powerful because it can lead to a formal and intuitive means of understanding a system that may be more informative than understanding a system in terms of an arbitrary task (the behavioral relevance of most psychophysical is not always fully justified). However defining optimality is also problematic, as assumptions need to be made about the requirements needed to define when a system is optimal and when it is not. For instance a recent branch of computational neuroscience defines optimality in terms of the response of low-level visual systems to natural scenes. The approach attempts to derive low-level filters from the environment by incorporating fitting constraints (assumptions) about what features of neural coding are desirable e.g. the independence of neural

coding (Simoncelli & Olshausen, 2001), or neural energy constraints (Attwell & Laughlin, 2001). Interestingly the results from such studies are consistent with the early decomposition of the visual world embodying a number of complimentary constraints (Simoncelli & Olshausen, 2001) and it will be interesting to see if the same approach can be applied to higher level processing further up the visual stream. Indeed recent work by Olshausen & Cadieu (2007) has attempted to derive two-dimensional motion filters from natural scenes.

It can be argued that the bias (or prior) towards slower speeds in the 2D model presented in this *thesis* results from the filter parameters chosen. To elaborate the 'prior' is embedded within the feed-forward weightings between the local motion sensors and the global motion sensors. The weighting function from each local (1D) sensor to each global (2D) sensor was determined by the mean response of a local sensor to the rigid translation of a natural scene in the *speed* and *direction* of the desired global (2D) sensors tuning. It is possible to alter the mean response of each local (1D) sensor by changing the *envelope* parameters or incorporating a normalization term. Thus providing scope for the experimenter to 'fit' or refine the model to the data at hand. In defense of the model presented in this *thesis* the selection of local filter parameters was based on achieving a homogeneous amplitude response from each local motion sensor without the need for normalization (i.e. the peak response of a sensor to an optimal stimulus was identical for all sensors). Accordingly, the bias for slow speed exhibited by the model can be considered as an emergent property of the filter configuration. None-the-less it would be interesting to see if the approach taken by Olshausen & Cadieu (2007), utilizing a more rigorous

definition of optimality, will also reveal a set of global motion filters with a preference for slow speeds under some orientation configurations.

Biological realism

In the section Biological Plausibility I went to great length to argue that the cosine-fitting model presented in this thesis is theoretically very similar to of the model of MT pattern selective cells proposed by Simoncelli & Heeger (1998). In this section I will argue that both models lack a number of features known about V1 DS cells. An important feature of both models is a fine sampling of the temporal frequency domain using filters that are tightly tuned for temporal frequency. Neither property is consistent with the known properties of V1 DS cells. Firstly, the temporal frequency tuning of such cells is described as either sustained (low-pass) or transient (band-pass) i.e. do not smoothly sample the temporal frequency domain. Secondly, the temporal frequency tuning of both sustained and transient cells is broad, non-symmetric and overlap to a large extent (Foster, et al., 1985; Hawken, Shapley, & Grosof, 1996). In turn, psychophysical studies report two or at most three temporal frequency channel (Anderson & Burr, 1985; Hess & Snowden, 1992). These findings have led many authors to propose that speed is calculated by taking the ratio of the response of sustained and transient filters (Johnston, et al., 1992; Thompson, 1982). Unlike the approach in this work, the activity of ratio based sensors increases monotonically with stimulus speed, rather than generating sensors that are *tuned* to specific speeds.

The problem of incorporating more realistic properties into the 'tiling' models has been discussed in depth elsewhere (Perrone, 2004) and I will briefly review

the main points here. The main issue regards the mechanism designed to recover 1D speed for sensors tuned not to speed, but tuned independently to spatial and temporal frequency. In the Simoncelli & Heeger (1998) model the 1D velocity sensors are inseparable for spatial and temporal frequency, meaning the sensors are speed tuned and thus inconsistent with the properties of V1 DS cells (Foster, et al., 1985; Hawken, et al., 1996). In the present work I use sensors that are separable for spatial and temporal frequency, however the temporal frequency tuning of the filter is narrow and the fine sampling of the temporal frequency dimension means that speed can be recovered from a broadband integration of the motion signals. Clearly the cosine pattern of activity recovered by the motion energy filters in [Figure 1-14](#) & [Figure 4-1](#) would be much harder to identify from just two temporal frequency channels (sustained and transient) with overlapping temporal frequency profiles. In turn, the capacity of the model presented throughout this thesis to compute 2D velocity is likely to be severely compromised. Accordingly, in its present state the model does not represent a complete model of 2D motion processing in the primate brain. To recover speed tuning from V1 sustained and transient filters an additional stage is required to refine the speed tuning of the sensors and this approach has been taken by Perrone (2004). In this work an additional Weighted Intersection Mechanisms (WIM) operates between the VA and MT stages. The operation of the WIM stage is described in (Perrone & Thiele, 2002): In short, by modeling the sustained and transient filter responses as Difference-of-Gaussian functions, Perrone and Thiele are able to specify the shape of the transient function such that its product with the sustained functions is inseparable in the spatial and temporal frequency domain. The resulting WIM

sensors are speed tuned and have similar speed tuning properties to some cells identified in area MT (Lagae, et al., 1993; Maunsell & Van Essen, 1983b).

Given that the model presented throughout this thesis is not biologically realistic, it is reasonable to ask whether there are any general principles we can extract from the model. In the following section I will argue that it is the superposition of signals from 1D velocities that causes the model to produce errors. I speculate that the problem is not unique to the cosine-fitting model, but will present a problem for any other models that assume isotropy, including the model by Perrone (2004) that makes a stronger argument for biological plausibility. To make this point, I use a more constrained modeling approach to explore the influence of bandwidths in a cosine-fitting model of 2D velocity processing.

The first stage consists of a bank of hypothetical sensors that are sensitive to the velocity, normal to a contour's orientation. The response properties of each sensor is described by 2D Gaussian function (Equation 5.3) where \mathbf{s} and \mathbf{d} denote the 1D speed and direction tuning of a sensor, σ_s and σ_d denote standard deviations of a sensor in speed and direction, whilst the physical speed and direction, normal to a contour's orientation is denoted by ϕ_{1D} and θ_{1D} (respectively). A bank of filters is created that spans directions between 0-360° and to velocities between 0 and 2 unit distance, per unit time. The model stimuli will always have a speed of 1 unit distance, per unit time, but a random 2D direction. As such this configuration of filters is able to detect the full range of 1D velocities that the stimuli could elicit.

$$V_{1D}(s,d) = e^{-\left(\frac{(s-\phi_{1D})^2}{2\sigma_{sp}^2} + \frac{(d-\theta_{1D})^2}{2\sigma_{dir}^2}\right)}$$

Equation 5.3

The second stage of the model is a template matching stage. Each template assumes isotropy in the orientation structure of a stimulus and is generated by summing the response of filters to each velocity upon a cosine (at 1° intervals). By changing the amplitude and phase of the cosine, 2D velocity templates are generated for a range of speeds and directions. An individual template for upward motion is shown in Figure 5-3. Estimates of 2D velocity are obtained by multiplying the output of the sensor bank with each 2D velocity template. A winner-takes-all algorithm is then used to select the estimate of 2D velocity in keeping with the modelling approach used throughout this thesis.

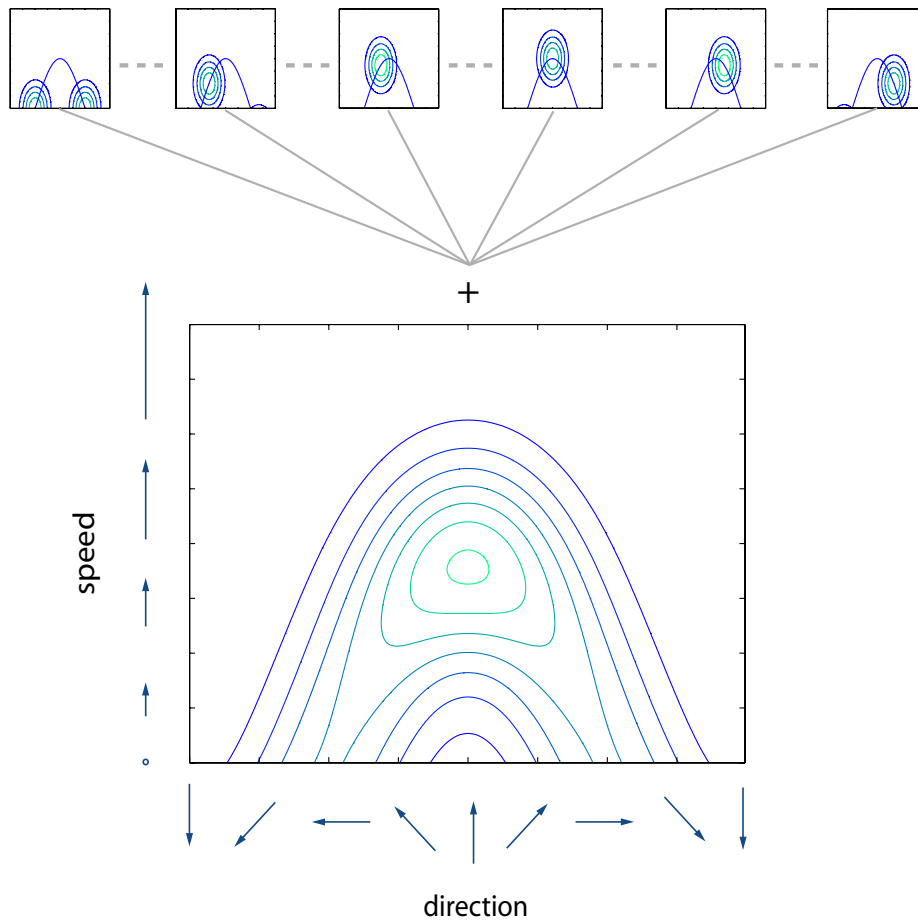


Figure 5-3 Templates for 2D velocity are generated by summing the activity generated by summed across the signals from 1D velocities that lie upon a cosine. Templates for different speeds and directions are computed by changing the phase and amplitude of the cosine.

To examine the influence of bandwidth, the standard deviation of the Gaussian was varied from [4, 8, 16, 32 & 64°] in direction and [0.04, 0.08, 0.16, 0.32, 0.64 & 0.128 unit distance, per unit time], in velocity. Because the templates were generated from the output of the 1D sensor bank, the

templates also inherited the increasing bandwidth size. Example templates are shown in the ordinate of Figure 5-3.

Initial testing of the model used just two orientations and a fixed 2D velocity. One of the orientations was fixed and was orthogonal to the 2D direction. The other was varied from -90 to $+90^\circ$ (i.e. parallel - orthogonal - parallel) to the 2D direction. The results are shown in Figure 5-4. Errors are generated for all bandwidths levels, and the sign of error is always towards the direction of 1D velocity. When both orientations are orthogonal, the model produces no errors. Errors are generated when the second component moves away from the orthogonal orientation. For the smallest bandwidth level, the maximal errors occurs for orientations near orthogonal, but occurs are more removed angles as the bandwidths size is increased. This pattern of results is consistent with the idea that it is the superposition of signals that causes the model to produce errors. Interestingly, when the bandwidth is $32^\circ 0.32 \text{ dt}^{-1}$, the point of maximal error occurs when the second component has a relative orientation of $\sim 45^\circ$, consistent with the results in the natural scenes chapter.

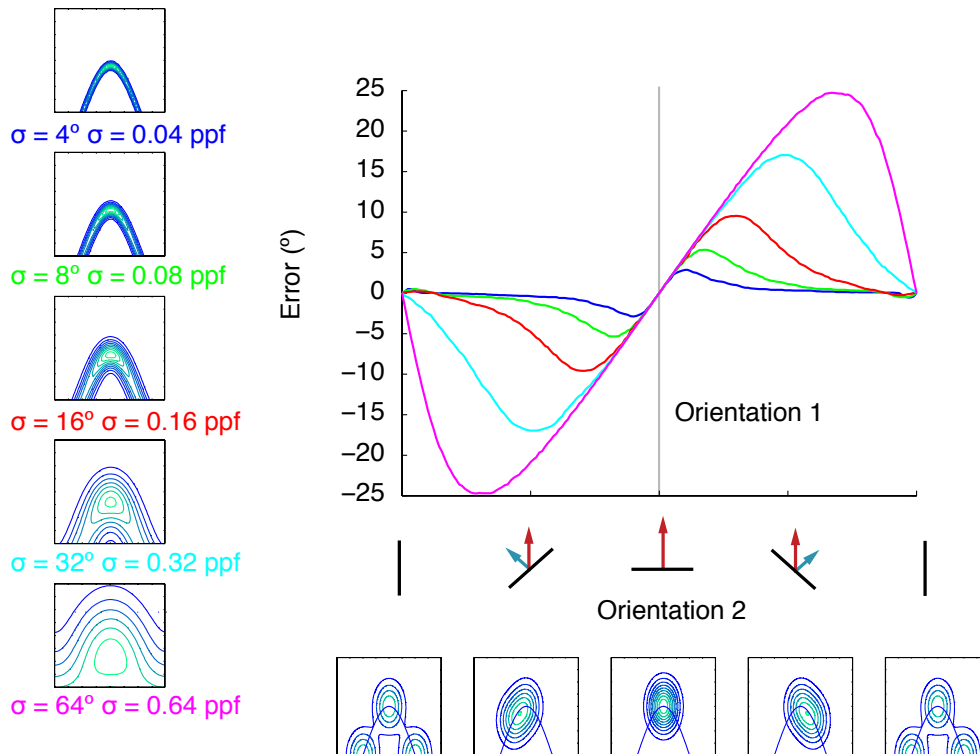


Figure 5-4 Performance of the cosine-fitting model given different bandwidths on the 1D velocity estimation. The stimulus is composed of two orientations, one is always fixed and is orthogonal to the 2D direction and the second is varied as denoted by the abscissa. The pattern of errors is color coordinated. The blue denotes errors in the smallest bandwidth condition; the errors are smallest in this condition and the maximal error occurs for orientations close to orthogonal. As the bandwidth is increased the magnitude of the errors increases and move to more oblique angles.

To illustrate how superposition may lead to the model to make errors Figure 5-5 plots the signals from two orientations, one oblique to the 2D direction and another orthogonal to the 2D direction. In (a) and the (b) the signals are plotted in isolation, note how both distributions of activity lie upon the blue cosine. In (c) the signals are plotted together; the superposition of the

signals generates a uni-modal distribution and the peak of energy lies inside of the cosine. This point is highlighted in (d) and (e) where the energy has been collapsed across speed and direction, respectively. In both cases the normalized sum of the two components (blue line) lies inside of the two components (pink and green). The best-fitting cosine (red) lies close to the peak of the uni-modal distribution.

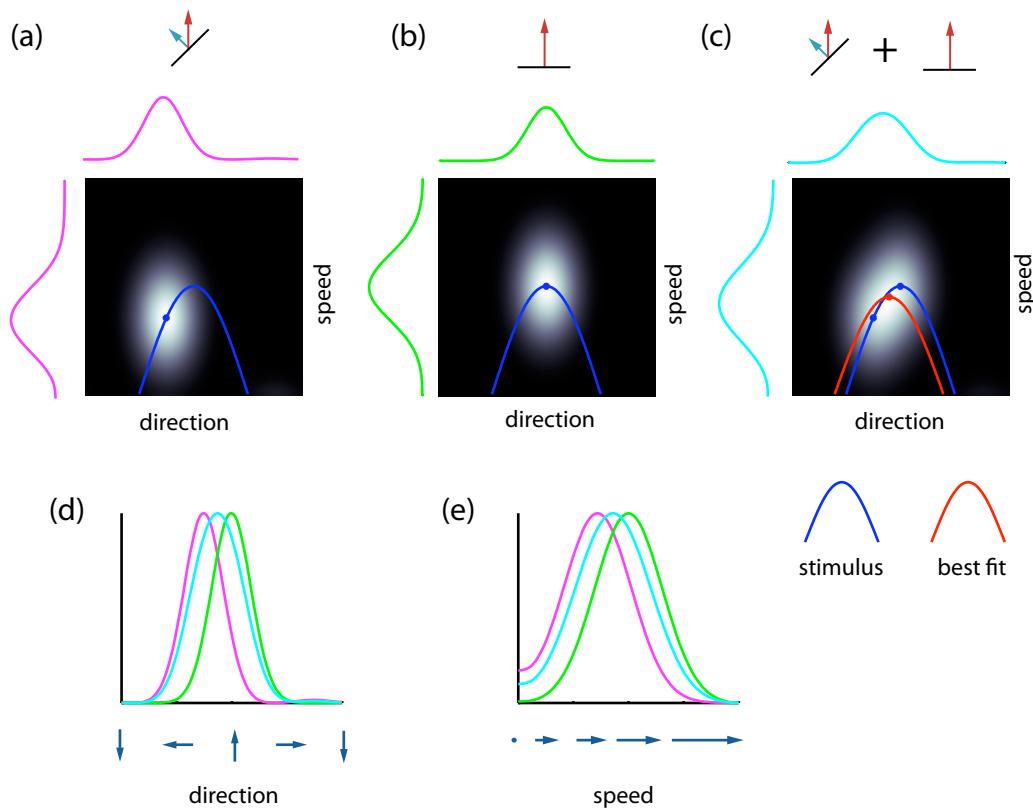


Figure 5-5 The signal for an orientation (a) oblique and (b) orthogonal to the 2D direction. (c) the combined signals from (a) and (b). The signals from both orientations (1D velocities) lie upon the blue cosine. In (d) and (e) the energy has been collapsed across the (d) speed and (e) direction dimensions. Note how the superposition of signals leads to a uni-modal distribution, with a peak that lies

inside of the blue cosine. The best fitting cosine (red) is close to the peak of energy in (c).

The aim of the final section was to identify what led the model to generate errors. I identify the superposition of 1D velocity signals as the cause of the errors. Superposition can distort the means of the component distributions by pulling the means inwards of the cosine. Even in the absence of external or internal noise this leads the model to produce errors. In a noisy system such errors would lead to systematic biases in the estimates of the model. The model used is hypothetical; the 1D velocity sensors are not designed to work on real stimuli, instead the model is designed to examine the *theoretical* influence of bandwidths on a cosine-fitting model of 2D velocity processing. I identify the superposition of signals as a cause of errors in the model. The problem of superposition is unlikely to just be a problem for the present model, for without a stage designed to disambiguate overlapping signals (such as a Gaussian-mixture-model) any model that inherits the 1D signal must account for the distortions caused by superposition. Models that assume isotropy in the orientation structure of the moving object such as Perrone (2004) and Simoncelli & Heeger (1998), are likely to produce a comparable pattern of errors.

References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A*, 2(2), 284-299.
- Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300(5892), 523-525.
- Ahumada, A., & Lovell, J. (1971). Stimulus Features in Signal Detection. *Journal of the Acoustical Society of America*, 49(6), 1751-&.
- Albright, T. D., & Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. *Exp Brain Res*, 65(3), 582-592.
- Amano, K., Edwards, M., Badcock, D. R., & Nishida, S. (2009). Adaptive pooling of visual motion signals by the human visual system revealed with a novel multi-element stimulus. *J Vis*, 9(3), 4 1-25.
- Amano, K., Edwards, M., Badcock, D. R., & Nishida, S. y. (2009). Adaptive pooling of visual motion signals by the human visual system revealed with a novel multi-element stimulus. *Journal of Vision*, 9(3), 1-25.
- Anderson, S. J., & Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Res*, 25(8), 1147-1154.
- Anderson, S. J., & Burr, D. C. (1989). Receptive field properties of human motion detector units inferred from spatial frequency masking. *Vision Res*, 29(10), 1343-1358.
- Ashida, H., & Osaka, N. (1994). Difference of spatial frequency selectivity between static and flicker motion aftereffects. *Perception*, 23(11), 1313-1320.
- Atick, J. J., & Redlich, A. N. (1992). What Does the Retina Know about Natural Scenes? *Neural Computation*, 19(9), pp. 2281-2300.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychol Rev*, 61(3), 183-193.
- Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab*, 21(10), 1133-1145.
- Bair, W., & Movshon, J. A. (2004). Adaptive temporal integration of motion in direction-selective neurons in macaque visual cortex. *J Neurosci*, 24(33), 7305-7323.
- Baker, C. L., Jr., Baydala, A., & Zeitouni, N. (1989). Optimal displacement in apparent motion. *Vision Res*, 29(7), 849-859.
- Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*, Cambridge, MA: MIT Press, pp. 217-234.
- Barlow, H., & Tripathy, S. P. (1997). Correspondence noise and signal pooling in the detection of coherent visual motion. *J Neurosci*, 17(20), 7954-7966.
- Barlow, H. B., & Hill, R. M. (1963). Selective sensitivity to direction of movement in ganglion cells of the rabbit retina. *Science*, 139, 412-414.
- Bartels, A., Zeki, S., & Logothetis, N. K. (2008). Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cereb Cortex*, 18(3), 705-717.
- Basole, A., White, L. E., & Fitzpatrick, D. (2003). Mapping multiple features in the population response of visual cortex. *Nature*, 423(6943), 986-990.

- Bex, P. J., & Dakin, S. C. (2002). Comparison of the spatial-frequency selectivity of local and global motion detectors. *J Opt Soc Am A Opt Image Sci Vis*, 19(4), 670-677.
- Bex, P. J., & Dakin, S. C. (2003). Motion detection and the coincidence of structure at high and low spatial frequencies. *Vision Res*, 43(4), 371-383.
- Bex, P. J., Dakin, S. C., & Mareschal, I. (2005). Critical band masking in optic flow. *Network*, 16(2-3), 261-284.
- Bex, P. J., Mareschal, I., & Dakin, S. C. (2007). Contrast gain control in natural scenes. *J Vis*, 7(11), 12 11-12.
- Bex, P. J., Simmers, A. J., & Dakin, S. C. (2001). Snakes and ladders: the role of temporal modulation in visual contour integration. *Vision Res*, 41(27), 3775-3782.
- Bex, P. J., Solomon, S. G., & Dakin, S. C. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *J Vis*, 9(10), 1 1-19.
- Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J Physiol*, 203(1), 237-260.
- Bowns, L. (1996). Evidence for a feature tracking explanation of why type II plaids move in the vector sum direction at short durations. *Vision Res*, 36(22), 3685-3694.
- Bowns, L. (2002). Can spatio-temporal energy models of motion predict feature motion? *Vision Res*, 42(13), 1671-1681.
- Bowns, L., & Alais, D. (2006). Large shifts in perceived motion direction reveal multiple global motion solutions. *Vision Res*, 46(8-9), 1170-1177.
- Braddick, O. J., Wishart, K. A., & Curran, W. (2002). Directional performance in motion transparency. *Vision Res*, 42(10), 1237-1248.
- Bradley, D. C., & Goyal, M. S. (2008). Velocity computation in the primate visual system. *Nat Rev Neurosci*, 9(9), 686-695.
- Brady, N., & Field, D. J. (2000). Local contrast in natural images: normalisation and coding efficiency. *Perception*, 29(9), 1041-1055.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, 10(4), 433-436.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci*, 12(12), 4745-4765.
- Burke, D., & Wenderoth, P. (1993). The effect of interactions between one-dimensional component gratings on two-dimensional motion perception. *Vision Res*, 33(3), 343-350.
- Burr, D. C. (1981). Temporal summation of moving images by the human visual system. *Proc R Soc Lond B Biol Sci*, 211(1184), 321-339.
- Burr, D. C., Baldassi, S., Morrone, M. C., & Verghese, P. (2009). Pooling and segmenting motion signals. *Vision Res*, 49(10), 1065-1072.
- Cai, D., DeAngelis, G. C., & Freeman, R. D. (1997). Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. *J Neurophysiol*, 78(2), 1045-1061.
- Campbell, F. W., Kulikowski, J. J., & Levinson, J. (1966). The effect of orientation on the visual resolution of gratings. *J Physiol*, 187(2), 427-436.
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *J Physiol*, 197(3), 551-566.
- Carandini, M. (2007). Melting the iceberg: contrast invariance in visual cortex. *Neuron*, 54(1), 11-13.

- Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., et al. (2005). Do we know what the early visual system does? *J Neurosci*, *25*(46), 10577-10597.
- Cass, J., Stuit, S., Bex, P., & Alais, D. (2009). Orientation bandwidths are invariant across spatiotemporal frequency after isotropic components are removed. *Journal of Vision*, *9*(12), 1-14.
- Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *J Vis*, *5*(9), 659-667.
- Cheng, K., Hasegawa, T., Saleem, K. S., & Tanaka, K. (1994). Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *J Neurophysiol*, *71*(6), 2269-2280.
- Cleary, R., & Braddick, O. J. (1990). Direction discrimination for band-pass filtered random dot kinematograms. *Vision Res*, *30*(2), 303-316.
- Cornsweet, T. (1970). *Visual Perception*: New York: Academic.
- Craik, K. (1966). *The Nature of psychology; a selection of papers, essays and other writings by the late K. J. W Craik*: Cambridge University Press.
- Curran, W., & Benton, C. P. (2003). Speed tuning of direction repulsion describes an inverted U-function. *Vision Res*, *43*(17), 1847-1853.
- Dakin, S. C., & Bex, P. J. (2003). Natural image statistics mediate brightness 'filling in'. *Proc Biol Sci*, *270*(1531), 2341-2348.
- Dakin, S. C., & Hess, R. F. (1998). Spatial-frequency tuning of visual contour integration. *J Opt Soc Am A Opt Image Sci Vis*, *15*(6), 1486-1499.
- Dakin, S. C., Mareschal, I., & Bex, P. J. (2005a). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vision Res*, *45*(24), 3027-3049.
- Dakin, S. C., Mareschal, I., & Bex, P. J. (2005b). An oblique effect for local motion: psychophysics and natural movie statistics. *J Vis*, *5*(10), 878-887.
- Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res*, *20*(10), 847-856.
- David, S. V., & Gallant, J. L. (2005). Predicting neuronal responses during natural vision. *Network*, *16*(2-3), 239-260.
- De Valois, K. K., De Valois, R. L., & Yund, E. W. (1979). Responses of striate cortex cells to grating and checkerboard patterns. *J Physiol*, *291*, 483-505.
- De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res*, *22*(5), 531-544.
- DeAngelis, G. C., & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J Neurophysiol*, *89*(2), 1094-1111.
- Dong, D. W., & Atick, J. J. (1995). Statistics of Natural Time-Varying Images. *Network-Computation in Neural Systems*, *6*(3), 345-358.
- Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *J Neurophysiol*, *65*(6), 1346-1359.
- Dumoulin, S. O., Dakin, S. C., & Hess, R. F. (2008). Sparsely distributed contours dominate extra-striate responses to complex scenes. *Neuroimage*, *42*(2), 890-901.
- Eagle, R. A., & Rogers, B. J. (1996). Motion detection is limited by element density not spatial frequency. *Vision Res*, *36*(4), 545-558.

- Eckstein, M. P., & Ahumada, A. J., Jr. (2002). Classification images: a tool to analyze visual strategies. *J Vis*, 2(1), 1x.
- Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nat Neurosci*, 8(12), 1643-1646.
- Felsen, G., Touryan, J., Han, F., & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS Biol*, 3(10), e342.
- Ferster, D., & Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annu Rev Neurosci*, 23, 441-471.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A*, 4(12), 2379-2394.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision Res*, 33(2), 173-193.
- Finn, I. M., Priebe, N. J., & Ferster, D. (2007). The emergence of contrast-invariant orientation tuning in simple cells of cat visual cortex. *Neuron*, 54(1), 137-152.
- Foster, K., Gaska, J., Nagler, M., & Pollen, D. (1985). Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *J Physiol.*, 331-363.
- Fredericksen, R. E., Bex, P. J., & Verstraten, F. A. (1997). How big is a Gabor patch, and why should we care? *J Opt Soc Am A Opt Image Sci Vis*, 14(1), 1-12.
- Gabor, D. (1946). *Theory of Communications* (Vol. 93).
- Gander, W., Golub, G. H., & Strebler, R. (1994). Least-Squares Fitting of Circles and Ellipses. *BIT* 34 pp. 558-578.
- Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res*, 41(6), 711-724.
- Geisler, W. S., & Ringach, D. (2009). Natural systems analysis. Introduction. *Vis Neurosci*, 26(1), 1-3.
- Gizzi, M. S., Katz, E., Schumer, R. A., & Movshon, J. A. (1990). Selectivity for orientation and direction of motion of single neurons in cat striate and extrastriate visual cortex. *J Neurophysiol*, 63(6), 1529-1543.
- Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Res*, 41(17), 2261-2271.
- Graham, N. V. S. (1989). Visual pattern analysers.
- Graham, N. V. S. (1989). Visual pattern analysers.
- Graziano, M. S., Andersen, R. A., & Snowden, R. J. (1994). Tuning of MST neurons to spiral motions. *J Neurosci*, 14(1), 54-67.
- Greenwood, J. A., & Edwards, M. (2006a). An extension of the transparent-motion detection limit using speed-tuned global-motion systems. *Vision Res*, 46(8-9), 1440-1449.
- Greenwood, J. A., & Edwards, M. (2006b). Pushing the limits of transparent-motion detection with binocular disparity. *Vision Res*, 46(16), 2615-2624.
- Greenwood, J. A., & Edwards, M. (2007). An oblique effect for transparent-motion detection caused by variation in global-motion direction-tuning bandwidths. *Vision Res*, 47(11), 1411-1423.
- Greenwood, J. A., & Edwards, M. (2009). The detection of multiple global directions: capacity limits with spatially segregated and transparent-motion signals. *J Vis*, 9(1), 40 41-15.

- Gros, B. L., Blake, R., & Hiris, E. (1998). Anisotropies in visual motion perception: a fresh look. *J Opt Soc Am A Opt Image Sci Vis*, 15(8), 2003-2011.
- Harris, J. M., & Drga, V. F. (2005). Using visual direction in three-dimensional motion perception. *Nat Neurosci*, 8(2), 229-233.
- Hawken, M. J., Shapley, R. M., & Gross, D. H. (1996). Temporal-frequency selectivity in monkey visual cortex. *Vis Neurosci*, 13(3), 477-492.
- Heeger, D. J. (1992a). Half-squaring in responses of cat striate cells. *Vis Neurosci*, 9(5), 427-443.
- Heeger, D. J. (1992b). Normalization of cell responses in cat striate cortex. *Vis Neurosci*, 9(2), 181-197.
- Heeley, D. W., & Buchanan-Smith, H. M. (1992). Directional acuity for drifting plaids. *Vision Res*, 32(1), 97-104.
- Hess, R. F., & Snowden, R. J. (1992). Temporal properties of human visual filters: number, shapes and spatial covariation. *Vision Res*, 32(1), 47-59.
- Hsu, A., Borst, A., & Theunissen, F. E. (2004). Quantifying variability in neural responses and its application for the validation of model predictions. *Network*, 15(2), 91-109.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160, 106-154.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol*, 195(1), 215-243.
- Jazayeri, M., & Movshon, A. J. (2007). Integration of sensory responses in coarse and fine motion discriminations. *Journal of Vision*, 7(9), 775-775.
- Jazayeri, M., & Movshon, J. A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, 446(7138), 912-915.
- Johnston, A., Benton, C. P., & Morgan, M. J. (1999). Concurrent measurement of perceived speed and speed discrimination threshold using the method of single stimuli. *Vision Res*, 39(23), 3849-3854.
- Johnston, A., McOwen, P. W., & Buxton, H. (1992). A biologically plausible scheme for measuring image velocity. *J. Physiol*, 452(288).
- Khawaja, F. A., Tsui, J. M., & Pack, C. C. (2009). Pattern motion selectivity of spiking outputs and local field potentials in macaque visual cortex. *J Neurosci*, 29(43), 13702-13709.
- Lagae, L., Raiguel, S., & Orban, G. A. (1993). Speed and direction selectivity of macaque middle temporal neurons. *J Neurophysiol*, 69(1), 19-39.
- Ledgeway, T. (1996). How similar must the Fourier spectra of the frames of a random-dot kinematogram be to support motion perception? *Vision Res*, 36(16), 2489-2495.
- Ledgeway, T., & Hess, R. F. (2002). Rules for combining the outputs of local motion detectors to define simple contours. *Vision Res*, 42(5), 653-659.
- Ledgeway, T., & Hess, R. F. (2006). The spatial frequency and orientation selectivity of the mechanisms that extract motion-defined contours. *Vision Res*, 46(4), 568-578.
- Ledgeway, T., Hess, R. F., & Geisler, W. S. (2005). Grouping local orientation and direction signals to extract spatial contours: empirical tests of "association field" models of contour integration. *Vision Res*, 45(19), 2511-2522.
- Li, B., Peterson, M. R., & Freeman, R. D. (2003). Oblique effect: a neural basis in the visual cortex. *J Neurophysiol*, 90(1), 204-217.
- Loffler, G., & Orbach, H. S. (2001). Anisotropy in judging the absolute direction of motion. *Vision Res*, 41(27), 3677-3692.

- Lorenceanu, J. (1998). Veridical perception of global motion from disparate component motions. *Vision Res*, 38(11), 1605-1610.
- Lorenceanu, J., & Alais, D. (2001). Form constraints in motion binding. *Nat Neurosci*, 4(7), 745-751.
- Lorenceanu, J., Giersch, A., & Series, P. (2005). Dynamic competition between contour integration and contour segmentation probed with moving stimuli. *Vision Res*, 45(1), 103-116.
- Lorenceanu, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Res*, 32(2), 263-273.
- Lorenceanu, J., Shiffrar, M., Wells, N., & Castet, E. (1993). Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Res*, 33(9), 1207-1217.
- Lorenceanu, J., & Zago, L. (1999). Cooperative and competitive spatial interactions in motion integration. *Vis Neurosci*, 16(4), 755-770.
- Majaj, N. J., Carandini, M., & Movshon, J. A. (2007). Motion integration by neurons in macaque MT is local, not global. *J Neurosci*, 27(2), 366-370.
- Mante, V., & Carandini, M. (2005). Mapping of stimulus energy in primary visual cortex. *J Neurophysiol*, 94(1), 788-798.
- Mante, V., Frazor, R. A., Bonin, V., Geisler, W. S., & Carandini, M. (2005). Independence of luminance and contrast in natural scenes and in the early visual system. *Nat Neurosci*, 8(12), 1690-1697.
- Mardia, K., & Jupp, P. (1972). Directional Statistics. *Wiley series in probability and statistics*, p.17.
- Marshak, W., & Sekuler, R. (1979). Mutual Repulsion between Moving Visual Targets. *Science*, 205(4413), 1399-1401.
- Masson, G. S., Rybarczyk, Y., Castet, E., & Mestre, D. R. (2000). Temporal dynamics of motion integration for the initiation of tracking eye movements at ultra-short latencies. *Vis Neurosci*, 17(5), 753-767.
- Matthews, N., & Qian, N. (1999). Axis-of-motion affects direction discrimination, not speed discrimination. *Vision Res*, 39(13), 2205-2211.
- Maunsell, J. H., & van Essen, D. C. (1983a). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J Neurosci*, 3(12), 2563-2586.
- Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol*, 49(5), 1127-1147.
- McDermott, J., Weiss, Y., & Adelson, E. H. (2001). Beyond junctions: nonlocal form constraints on motion interpretation. *Perception*, 30(8), 905-923.
- Meng, X., & Qian, N. (2005). The oblique effect depends on perceived, rather than physical, orientation and direction. *Vision Res*, 45(27), 3402-3413.
- Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *J Neurophysiol*, 55(6), 1308-1327.
- Mingolla, E., Todd, J. T., & Norman, J. F. (1992). The perception of globally coherent motion. *Vision Res*, 32(6), 1015-1031.
- Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature*, 355(6358), 344-346.
- Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). *The analysis of moving visual patterns*. New York: Springer.

- Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1986a). The analysis of moving visual patterns. In Chagas, C., Gattass R. & Gross, C. (Eds), *Experimental brain research supplementum II: Pattern recognition mechanisms*, pp. 117-151.
- Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1986b). The analysis of moving visual patterns.
- Movshon, J. A., & Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J Neurosci*, 16(23), 7733-7741.
- Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature*, 395(6705), 894-896.
- Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, 341(6237), 52-54.
- Nishida, S. (2004). Motion-based analysis of spatial patterns by the human visual system. *Curr Biol*, 14(10), 830-839.
- Nowlan, S. J., & Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates. *J Neurosci*, 15(2), 1195-1214.
- O'Brien. (1958). Contour perception, illusion and reality. *J. Opt. Soc. Am.*, 48,, 112-119.
- Olshausen, B., & Cadieu, C. (2007). Learning invariant and variant components of time-varying natural images. *Journal of Vision*, 7(9), 964-964.
- Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding v1? *Neural Comput*, 17(8), 1665-1699.
- Pack, C. C., & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409(6823), 1040-1042.
- Parker, A. J., & Hawken, M. J. (1988). Two-dimensional spatial structure of receptive fields in monkey striate cortex. *J Opt Soc Am A*, 5(4), 598-605.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis*, 10(4), 437-442.
- Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Res*, 31, 1337-1350.
- Perrone, J. A. (2001). *A closer look at the visual input to self-motion estimation.*
- Perrone, J. A. (2004). A visual motion sensor based on the properties of V1 and MT neurons. *Vision Res*, 44(15), 1733-1755.
- Perrone, J. A., & Thiele, A. (2001). Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nat Neurosci*, 4(5), 526-532.
- Perrone, J. A., & Thiele, A. (2002). A model of speed tuning in MT neurons. *Vision Res*, 42(8), 1035-1051.
- Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Res*, 33(7), 993-999.
- Priebe, N. J., Lisberger, S. G., & Movshon, J. A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *J Neurosci*, 26(11), 2941-2950.
- Qian, N., Andersen, R. A., & Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics. *J Neurosci*, 14(12), 7357-7366.
- Ramachandran, V. S., & Cavanagh, P. (1987). Motion capture anisotropy. *Vision Res*, 27(1), 97-106.

- Rauber, H. J., & Treue, S. (1998). Reference repulsion when judging the direction of visual motion. *Perception*, 27(4), 393-402.
- Rauber, H. J., & Treue, S. (1999). Revisiting motion repulsion: evidence for a general phenomenon? *Vision Res*, 39(19), 3187-3196.
- Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. A. Rosenblith (Ed.) *Sensory communication*, pp. 303-317.
- Rodman, H. R., & Albright, T. D. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Exp Brain Res*, 75(1), 53-64.
- Ross, J. (2004). The perceived direction and speed of global motion in Glass pattern sequences. *Vision Res*, 44(5), 441-448.
- Ross, J., Badcock, D. R., & Hayes, A. (2000). Coherent global motion in the absence of coherent velocity signals. *Curr Biol*, 10(11), 679-682.
- Rubin, N., & Hochstein, S. (1993). Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional objects. *Vision Res*, 33(10), 1385-1396.
- Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nat Neurosci*, 9(11), 1421-1431.
- Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nat Neurosci*, 8(12), 1647-1650.
- Salzman, C. D., & Newsome, W. T. (1994). Neural mechanisms for forming a perceptual decision. *Science*, 264(5156), 231-237.
- Schrater, P. R., Knill, D. C., & Simoncelli, E. P. (2000). Mechanisms of visual motion detection. *Nat Neurosci*, 3(1), 64-68.
- Shapley, R., & Perry, V. H. (1986). Cat and Monkey Retinal Ganglion-Cells and Their Visual Functional Roles. *Trends in Neurosciences*, 9(5), 229-235.
- Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Res*, 29(5), 619-626.
- Shipley, T. F., & Kellman, P. J. (1993). Optical tearing in spatiotemporal boundary formation: when do local element motions produce boundaries, form, and global motion? *Spat Vis*, 7(4), 323-339.
- Silveira, L. C., & Perry, V. H. (1991). The topography of magnocellular projecting ganglion cells (M-ganglion cells) in the primate retina. *Neuroscience*, 40(1), 217-237.
- Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Res*, 38(5), 743-761.
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu Rev Neurosci*, 24, 1193-1216.
- Snowden, R. J., Treue, S., & Andersen, R. A. (1992). The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. *Exp Brain Res*, 88(2), 389-400.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nat Neurosci*, 9(4), 578-585.
- Stone, L. S., Watson, A. B., & Mulligan, J. B. (1990). Effect of contrast on the perceived direction of a moving plaid. *Vision Res*, 30(7), 1049-1067.
- Stoner, G. R., & Albright, T. D. (1996). The interpretation of visual motion: evidence for surface segmentation mechanisms. *Vision Res*, 36(9), 1291-1310.

- Switkes, E., Mayer, M. J., & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis. *Vision Res*, 18(10), 1393-1399.
- Takeuchi, T. (1998). Effect of contrast on the perception of moving multiple Gabor patterns. *Vision Res*, 38(20), 3069-3082.
- Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Res*, 22(3), 377-380.
- Tolhurst, D. J., & Movshon, J. A. (1975). Spatial and temporal contrast sensitivity of striate cortical neurones. *Nature*, 257(5528), 674-675.
- Uka, T., & DeAngelis, G. C. (2003). Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *J Neurosci*, 23(8), 3515-3530.
- van Hateren, J. H. (1997). Processing of natural time series of intensities by the visual system of the blowfly. *Vision Res*, 37(23), 3407-3416.
- van Hateren, J. H., Ruttiger, L., Sun, H., & Lee, B. B. (2002). Processing of natural temporal stimuli by macaque retinal ganglion cells. *J Neurosci*, 22(22), 9945-9960.
- van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc Biol Sci*, 265(1394), 359-366.
- van Santen, J. P., & Sperling, G. (1984). Temporal covariance model of human motion perception. *J Opt Soc Am A*, 1(5), 451-473.
- von Grunau, M., & Dube, S. (1992). Comparing local and remote motion aftereffects. *Spat Vis*, 6(4), 303-314.
- Wallach, H. (1935). Uber visuell wahrgenommene Bewegungsrichtung. *Psychologische Forschung* (20), pp. 325-380.
- Watamaniuk, S. N. J., & McKee, S. P. (1998). Simultaneous encoding of direction at a local and global scale. *Perception & Psychophysics*, 60(2), 191-200.
- Watson, A., & Ahumada, A. (1983). A look at motion in the frequency domain. *Motion: perception and representation*, pp. 1-10.
- Webb, B. S., Ledgeway, T., & McGraw, P. V. (2007). Cortical pooling algorithms for judging global motion direction. *Proc Natl Acad Sci U S A*, 104(9), 3532-3537.
- Weiss, Y., & Adelson, E. H. (1998). Slow and Smooth: a Bayesian theory for the combination of local motion signals in human vision. *Technical Report 1624, MIT AI lab*, 1998.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat Neurosci*, 5(6), 598-604.
- Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Vis Neurosci*, 9(1), 79-97.
- Wilson, H. R., & Kim, J. (1994). Perceived motion in the vector sum direction. *Vision Res*, 34(14), 1835-1842.
- Yo, C., & Wilson, H. R. (1992). Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Res*, 32(1), 135-147.
- Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *J Neurosci*, 11(3), 641-649.

