Wallace, Julian Michael (2003) *The efficiency of visual transparency.* PhD thesis.

http://theses.gla.ac.uk/5543/

The Efficiency of Visual Transparency

Julian Michael Wallace

Department of Psychology

University of Glasgow

Acknowledgements

Declaration

I declare that this thesis is my own work, carried out under the normal terms of supervision.

Abstract

This thesis examines the phenomenon of visual transparency in a novel application of the efficiency approach. Transparency provides a useful stimulus to probe the visual mechanisms that underlie the visual surface representation, introduced in Chapter One. Previous research has found that there is a cost in processing visual transparency defined purely by motion or stereo cues. This has been interpreted in terms of visual mechanisms constraining the recovery of transparency. However, the cost for transparency may reflect the increased complexity of the stimuli. To address this issue I computed the efficiency for motion and stereo defined transparency tasks by comparing human performance with that of the ideal observer. The efficiency approach has two key advantages over traditional psychophysical measures: 1) it provides a performance measure normalised relative to the available information, 2) it is an absolute measure and can be compared directly across diverse tasks. I provide a review of the efficiency approach in Chapter Two. In Chapter Three, I present a study of the efficiency for speed discrimination of transparent random dot stimuli and comparable non-transparent random dot stimuli, as a function of the speed ratio and the dot density of the stimuli. In Chapter Four, I present a study of the efficiency for depth discrimination of transparent and non-transparent random dot stereograms, across a range of disparity ratios and dot densities. In Chapter Five, I present an extension of the efficiency approach to the motor domain, for the smooth pursuit of high-density transparent and non-transparent random-dot stimuli. Finally, in Chapter Six I provide physiologically plausible accounts of the findings.

Table of Contents

Chapter One: General Introduction

1.1   The Problem of Vision

Vision science aims to understand the mechanisms of the visual system, how the apparatus of the brain provides us with our rich visual experience. How is it that we are conscious of a three-dimensional world, when the information given to us is a two-dimensional image formed on the retina by the optics of the eye? The problem of vision is to retrieve the three-dimensional information about the scene given the two-dimensional image, this is known as the inverse problem of vision and it is non-trivial. In simple terms, there are infinitely many possible three-dimensional scenes that could have given rise to any particular two-dimensional image. This has been understood at least since the philosophical considerations of George Berkeley (Berkeley, 1709). The modern approach to the problem has roots in classical constructivism (e.g. Helmholtz, 1867). This holds that, given the ill-posed nature of the problem of vision, the processes of visual perception must use assumptions, or 'unconscious inferences', to arrive successfully at an interpretation of the three-dimensional scene. Indeed, the existence of visual illusions provides compelling evidence for these inferential visual processes (e.g. Hoffman, 1998). The fundamental concept in the modern approach is to consider these inferential visual processes as information processing events, and this is best exemplified in the theoretical framework of

Marr (1982). Marr was not the first to champion a computational approach to vision, indeed there are a number of significant developments in applying information processing concepts to cognition in general that precede his work (Shannon, 1949; Turing, 1950) and also more specifically to visual processing (Rosenblatt, 1962; Minsky & Papert, 1969; Land & McCann, 1971; Horn, 1975). Nevertheless, Marr's contribution cannot be understated. His work provided a meta-theoretical framework for vision research, and in particular proposed a specific representational scheme that continues to inform our understanding of the structure and function of the visual system.

In his posthumously published book, 'Vision' (1982), Marr contrasted the development of the modern computational approach with a dominating purely physiological approach, which taken to its logical conclusion equates descriptions of physiological responses with explanations of visual behaviour (see Barlow, 1972). In contrast to the reductionist enterprise, Marr described three levels of analysis necessary to explain any information-processing problem. These levels are the computational level, the level of representation and algorithm, and the implementational level. The computational level is essentially a theory of the goal of the system. In the case of vision, the basic goal of the visual system is to solve the inverse problem and recover a three-dimensional representation of the scene, given a particular two-dimensional image. The level of representation and algorithm is concerned with specifying how to implement

this theory, in terms of the necessary transformations. Marr proposed a particular representational framework, which has shaped much of the modern approach. The final implementational level is concerned with specifying precisely how the algorithm is realised in a physical system. In vision this means understanding how the algorithm is realised in the biological system. This is not simply a case of looking at physiological data, but of using the data to constrain explanatory models of the system. To provide a complete understanding of the visual system will require an interaction between these three levels of analysis. In practice this demands a synergy between computational theory and modelling, psychophysics and physiology.

This thesis is concerned with a particular problem in vision, the recovery of surface transparency. I approach this problem in terms of the three levels of analysis described by Marr. At the computational level, I employ a particular computational theory of the task to analyse behavioural performance, the efficiency approach. In the second chapter of this thesis I explain this approach in detail. At the representational level, I consider the problem of transparency in terms of a general representational framework. In the following section 'A Sequence of Representational Stages' I describe in further detail this representational framework. At the implementational level, I interpret my results in terms of particular physiological constraints. I provide a general and brief background to the known physiology in the Section 'A Visual Hierarchy'.

The Section on 'The Problem of Transparency', describes the particular problem that I investigate in this thesis. The final Section in this chapter, 'Thesis Outline', states the aims of the present thesis, and provides a brief summary of the research reported in the following chapters.

1.2 A Sequence of Representational Stages

An important task of vision is to provide a representation of the three-dimensional scene, given a two-dimensional retinal image. Marr (1982) described how this could be achieved by a sequence of three main representational stages, an initial representation of luminance differences (specifically edges), a subsequent elaboration into a partial three-dimensional representation of the location of surfaces in the world, and the final three-dimensional 'object-centred' representation. The representations at each stage are constructed by algorithms that use specific assumptions to make explicit aspects of the scene not available in the immediately preceeding stage. While the particular algorithms Marr and colleagues (Marr & Poggio, 1976, 1979; Marr & Nishihara, 1978; Marr & Hildreth, 1980; Marr & Ullman, 1981) provided to construct these representations have not gone unchallenged, the notion that the visual system consists of a sequence of representational stages has (at least implicitly) become an integral part of the modern approach. The stages of representation are more generally referred to as low-level (image representation

stage), mid-level (surface representation stage) and high-level (object representation stage).

This thesis focuses on mid-level vision. In Marr's representational framework, the surface representation provided a crucial link between image representations and object representations. The importance of surfaces in visual perception had earlier been emphasised in the work of J. J. Gibson. Gibson's contribution was to consider the information available to an observer in the environment. In his analysis (Gibson, 1979) it was clear that the surfaces in the environment provided crucial information for visual perception:

"The main invariants of the terrestrial environment, its persisting features, are the layout of its surfaces and the reflectances of these surfaces." (p. 87)

However, Gibson denied there was any need for the visual apparatus to extract or represent the information in the environment. Rather the information was assumed to be directly available to the observer, the apparatus of the visual system was assumed simply to 'resonate' with the information given in the 'ambient optic array' (the changing visual angles of projection from surfaces to the observer as the observer explores the environment). Gibson's argument therefore goes as far to deny the retinal image and the fundamental problem of

vision. Marr (1982) acknowledged the contribution of Gibson in emphasising the information available in the environment, but restated this in contructivist terms. Surfaces indeed provide crucial information for perception, but this information is available only indirectly in the two-dimensional retinal image:

"The principal factors that determine the intensity values in an image are (1) the illumination, (2) the surface geometry, (3) the surface reflectance, (4) the vantage point." (p. 272)

This brings us back to the retinal image as the start of the problem. Marr emphasised that the visual system could exploit the structure inherent in the two-dimensional image, in the form of assumptions, to recover a representation of the surfaces in the environment. The importance of this representational stage intermediate between low-level and high-level visual processing was recently re-emphasised by Nakayama, He & Shimojo (1995), and is increasingly becoming a key issue in vision science.

1.3   The Visual Hierarchy

Throughout this thesis, I refer to physiological aspects of the visual system, particularly in the interpretation of the behavioural data. Here I provide a brief background of the key physiological research. While Marr's approach was developed largely as a reaction to a purely physiological approach, extensive

13

physiological studies have provided fascinating descriptions of the visual system. Primate visual cortex consists of a number of distinct areas, known to occupy around 60% of the entire cortical area (Van Essen & Maunsell, 1980). Furthermore, analysis of the intercellular connections between these different areas revealed a complex neural architecture beginning in the retinal layers in which information is fed up (and also back down) through a sequence of areas or visual processing modules (Maunsell & Van Essen, 1983; Maunsell & Newsome, 1987; Van Essen et al., 1992). A key development in the physiological approach was the development of techniques to record the electrical activity in single cells of the system (see Hubel & Wiesel, 1998), allowing researchers to map the 'receptive fields' of neurons i.e. to identify the types of visual stimuli that cells respond best to. A wide range of single-cell recording studies (initially with mammals such as cat and rabbit, but later focusing on the primate visual system, largely in the macaque monkey) had provided fascinating insights into the nature of the visual network following the initial transduction of light energy into neural signals (electrical impulses) by the retinal photoreceptors, the selective responses of visual neurons suggested a complex system of information processing. Studies of the early retinal architecture found cells sensitive to local luminance differences (e.g. Barlow, 1953; Kuffler, 1953), and it was suggested that these responses could be achieved by spatially pooling information from many photoreceptors selectively. These responses are combined in increasingly selective ways as the information is fed through the LGN to visual cortex.

In striate cortex (V1), the first visual cortical area largely receiving retinal input relayed from the LGN, the information from retinal cells is pooled selectively. Hubel & Wiesel's studies (1962, 1965, 1968) established the principle that the sensitivity of cells could be achieved through selective combinations of the outputs from the preceeding stage, and by this principle they established a hierarchy of responses in V1 (simple cells -> complex cells -> hypercomplex cells). The simple responses were initially thought to be extracting important features in the image, namely edges of particular orientations (Hubel & Wiesel, 1962, 1965, 1968), although these responses may be better understood not as feature detectors but as spatial frequency analysers (Campbell & Robson, 1968; Blakemore & Campbell, 1969; De Valois & De Valois, 1988). Further studies in monkey delved further and further into the system. In area V2, an area immediately following V1, the selectivity appears to integrate this early information e.g. for contours (von der Heydt et al., 1984). Higher up in the system, there is selectivity for more complex geometric stimuli in V4 (Gallant et al, 1993), and even hands (Gross et al, 1972), faces (Perret et al., 1982) and specific objects (Logothetis & Sheinberg, 1996) in area IT.

This range of evidence suggested that the visual system consists of an anatomical and functional hierarchy of areas, in which processing proceeds from the simple (feature detection or spatial frequency analysis) to the complex (specific object or

face detection). Therefore, in principle the responses in the visual hierarchy can be mapped onto Marr's representational stages. However, the exact hierarchical organization of the system is indeterminate, while anatomical constraints can limit the possibilities, no single hierarchy can be determined (Hilgetag et al, 1996; Crick & Koch, 1998). Furthermore, it has been hypothesised that the hierarchy (however it is arranged) can be subdivided into two parallel pathways, one specialising in analysis for object recognition, referred to as the 'what' pathway, and the other in the analysis for spatial location, the 'where' pathway (Ungerleider & Mishkin, 1982; Mishkin, Ungerleider & Macko, 1983), famously re-interpreted as an 'action' pathway and a 'perception' pathway (Milner & Goodale, 1995). A subdivision in visual processing can be identified as early as the LGN (Livingstone & Hubel, 1988), the distinction between selectivity for color and form and selectivity for motion and depth (see also Van Essen & Maunsell, 1983). The motion pathway proceeds from cells in layer 4b of V1 (Maunsell & Newsome, 1987) where responses are local and ambiguous (Movshon et al, 1986), directly through to area MT where cells are selective for global pattern motion (Movshon et al, 1986), through to area MST where cells are selective for global optic flow (Tanaka & Saito, 1989). Evidence for parallel processing of different visual attributes is consistent with the principle of modular organization (Fodor, 1983). The following section expands on the concept of modularity in terms of the general problem of vision, and introduces the particular problem addressed by this thesis.

1.4   The Problem of Transparency

A fundamental problem of vision is the recovery of the third dimension, depth. We have seen that surfaces are fundamental to visual perception; the distance of a surface to the observer specifies depth. A number of types of information are useful cues in the recovery of depth, and are generally divided into the monocular cues, such as texture, shading, motion, linear perspective, relative size, and binocular cues, such as stereopsis, the difference between the images in the two eyes. These different sources of information available to the system can be combined to obtain reliable estimates of depth in the scene (e.g. Landy et al., 1995), particularly useful in complex situations such as the case of transparency (Kersten, 1991). Perceptual transparency occurs when two surfaces are perceived simultaneously in the same visual direction, a far surface is perceived through a nearer transparent surface. Here the visual system is able not only to reconstruct depth from a two-dimensional image, but multiple depths in the same visual direction. Information arising from a common surface is integrated, and information arising from different surfaces is segmented. These segmentation processes are fundamental in recovering a surface representation (Marr, 1982). Therefore, probing the ability of the system to recover such a complex percept may reveal some basic principles of visual surface processing.

A number of different types of transparency that arise in day-to-day viewing of the environment can be identified: such as film transparency, which occurs when we look through a transparent sheet, or specular transparencies, as occurring when looking through a pane of glass, and other familiar occurrences such as shadows and occlusions are also cases of transparency (see Kersten, 1991, for a detailed and mathematical description of the possible cases). Nevertheless, an interesting aspect of perceptual transparency is that it can be elicited by particular two-dimensional stimuli, where there is phenomenal transparency in the absence of physical transparency. This indicates that the mechanisms of surface representation function according to particular constraints or rules. For example, particular 2D spatial arrangements of areas of different luminance elicit perceptual transparency conforming to a theory of 'color scission', that is to say that the visual system will construct transparent surfaces when provided with a stimulus consistent with the physics of transparency (Metelli, 1974). A particularly compelling impression of transparency occurs with random dot displays moving in different directions, or presented stereoscopically at different depths. Random dot displays are in themselves interesting stimuli in probing surface representation. A sparse display of randomly placed dots on a background can be perceived as a continuous surface (White, 1962), suggesting that the visual system interpolates surface information; the exact process of 'filling-in' has received much research attention recently (Shimojo et al., 2001; Komatsu et al., 2000; De Weerd et al., 1998). When two random dot displays are

presented simultaneously in different directions, or presented stereoscopically at different depths, two surfaces are perceived; an opaque surface is perceived beyond a near transparent surface. Indeed, stereopsis and motion are known to be particularly strong cues to depth (e.g. Marr, 1982). This thesis focuses on the ability of the visual system to recover surface transparency from motion and stereo cues.

1.5   Thesis Outline

I use a particular computational approach, the efficiency approach, to examine visual transparency, specifically for transparency defined by motion and stereo cues in random dot displays. The aim of the thesis is to elucidate the mechanisms underlying the surface representation, how does the visual system selectively combine local information to recover a description of the surfaces in the environment? In Chapter Two I describe the efficiency approach in greater detail. It will be shown that there are two main benefits in the approach: 1) efficiency normalises performance to the available information, therefore patterns in performance can be directly interpreted in terms of the underlying mechanisms, 2) efficiency is an absolute measure of visual performance, thus we can compare performance directly across different tasks. In Chapter Three, I report an experimental study on the efficiency of visual transparency defined by motion cues. The main results are: 1) there is a cost for processing motion transparency compared to non-transparent motions, and 2) performance is

impaired by increasing the density of the random dot displays. In Chapter Four, I report an experimental study on the efficiency of visual transparency defined by stereoscopic information. The main results are: 1) there is no cost for processing stereoscopic transparency compared to a non-transparent surfaces, and 2) performance is impaired by increasing the density of the random dot displays. In Chapter Five, I report an experimental study that extends the analysis of transparent motion efficiency to the motor domain, specifically to the analysis of smooth pursuit eye movements to transparent motion. This study has two aims: 1) to permit a comparison of motor and perceptual efficiencies for transparent motion, and 2) to probe the temporal dynamics of motion integration and segmentation. Finally, in Chapter Six I review the main findings of these experimental studies, and consider the future directions of the research.

Chapter Two: The Efficiency Approach

2.1 Introduction

The behavioural study of visual perception involves the measurement of observer's responses, their observable behavioural performance, for well-designed visual tasks in order to make inferences about the unobservable sensory systems that underlie that behavioural performance. Traditionally the discipline is known as 'psychophysics', emphasising the empirical goal of relating physical stimulus properties to their psychological effects. The fundamental measurement of psychophysics is a threshold, measured in units of the manipulated variable. A number of methods to estimate thresholds were developed in the nineteenth century by Gustav Fechner (1860), the 'father of psychophysics'. These were the method of adjustment, the method of limits, and the method of constant stimuli. As a simple example, we could measure the 'absolute' threshold for detecting a spot of light. We would present an observer with a range of light intensities (method of constant stimuli) and ask the observer to state whether or not they detected the spot of light, and we could then construct a psychometric function from this data, which are typically well fit by sigmoid functions. We might then arbitrarily define their sensory threshold as the light intensity the observer requires to detect the spot of light 50% of the time. This would give us an indication of the sensitivity of the visual system to light.

Thresholds can be computed across a range of stimulus parameters, and the pattern of thresholds as a function of the stimulus parameters, the psychophysical function, may be used to infer properties of the underlying visual mechanisms. Indeed, in the late nineteenth century this approach led to the discovery of basic psychophysical laws, such as Weber's law which states that the change in light intensity required to detect a change increases as the baseline light intensity increases. Since then the basic method of computing psychophysical functions from psychometric data has been applied to a broad range of visual tasks e.g. the detection or discrimination of motion signals, where the threshold might be expressed in terms of speed, or the detection of stereoscopically defined depth, where the threshold might be expressed in terms of disparity. This approach is used to make inferences about the nature of the underlying mechanisms within a domain e.g. the range of speeds the visual system is sensitive to for a given task.

However, a number of drawbacks can be identified that limit the power of basic threshold measurements. First, simple threshold measurements (derived from yes/no tasks) confound the measurement of the observer's sensitivity (what psychophysicists are interested in) with the observer's criterion, the willingness of the observer to respond in one way or another based on the sensory information available to them. This problem was a key concern driving the development of signal detection theory, which provided a method to measure

the observer's sensitivity independent of their criterion (Tanner & Swets, 1954; Swets, Tanner & Birdsall, 1961). By this approach the observer does not respond in an all or none fashion when a given stimulus attribute exceeds an estimated threshold value, rather the observer uses the available sensory information to perform a statistical decision (see Swets, 1964). Second, while it would be useful to compare visual performance across different domains to build a complete description of the system, it is meaningless to *directly* compare thresholds measured in different units e.g. to compare a disparity threshold directly with a speed threshold. To compare visual performance directly across different visual domains such as motion and stereo we require a unit-free measure of performance. Third, threshold measurements do not depend upon a theory of the task, that is to say that while our aim is to make inferences about the underlying visual mechanisms, without specifying what information is actually available to perform the task we cannot be sure of the underlying cause of any performance limitations we may find. In particular, without a theory of the task *we cannot determine whether performance is limited by the information we provided to perform the task, or the observer's use of that information.*

All three of these issues can be addressed by the efficiency approach. While the problem of the criterion can be circumvented by the use of 2AFC tasks, and unit-free measures of performance are available, such as the d-prime measure of signal detection theory (which also controls for the observer's criterion),

efficiency has a distinct advantage over these methods in that it is a unit-free measure of *information* processed or required. A level of performance given by any other measure will reflect the ability of the visual mechanisms to use the visual information, *but also will reflect the information available for the task at hand.* This is a subtle but crucial distinction. When these two possible causes of performance levels are confounded, this puts serious limitations on the inferences we can draw about the underlying mechanisms, when looking at performance across a range of parameters within a domain or comparing performance for different stimuli (within or across domains). The advantage of efficiency over other traditional information processing measures is that *it compares two observers,* human and ideal. Essentially, the ideal observer makes optimal use of the information available to satisfy a performance criterion for a given task, providing a theoretical upper bound to compare with human visual processing. The level of discrepancy between actual and optimal performance (measured sensitivities) can then be computed simply by taking the ratio of the two. Computing this ratio normalises human performance *relative to the available information.* Therefore any performance limits can be interpreted purely in terms of the ability of the underlying mechanisms to use the available information. This absolute measure can then be compared both across parameters within a task and also between different tasks.

The efficiency measure can inform us about the nature of the mechanisms of the visual system, free from confounding effects of the available information. Furthermore, the measure can be used to tease apart the internal factors contributing to performance; the effectiveness of the underlying mechanisms, the variability in the decision process, and the internal noise of the system (see section 2.4). There are a number of ways in which the approach has been applied. We can search a space of stimulus dimensions and identify the 'best' parameters, e.g. the spatial & temporal frequencies giving the highest efficiencies for detecting a Gabor patch (Watson et. al, 1983). These 'optimal' stimuli are excellent candidates for visual mechanisms, i.e. we would expect to find mechanisms 'tuned' to such stimuli at some stage of the visual system. As the efficiency measure is absolute, we may also compare efficiencies between different tasks. Simply stated, we can assess how good (or bad) the visual system is at performing different tasks. Then, by comparing efficiencies across diverse tasks the method could provide us with a more complete description of the system. That is to say that, not only can we identify the optimal parameters for a given class of stimuli (e.g. Gabor patches), but by comparing efficiencies across different classes of stimuli we could also identify the tasks the visual system has been designed to perform. We may also use the measure to test our models of the system. Having established the absolute efficiency for a task, we may introduce constraints to the ideal observer. We may permit the ideal observer the use of only a subset of the available information (e.g. Liu et al, 1995;

Knill, 1998a, 1998b, 1998c), or introduce physical limitations known to affect human performance e.g. optical blurring of the image and photoreceptor spacing (Andrews et al., 1973; Watt & Andrews, 1982) or place the ideal observer at any stage of the known physiological mechanisms (Geisler, 1989). If these modifications led to improved efficiencies we could infer that we have succeeded in identifying a key limitation of the system. Similarly, we would compute efficiencies relative to a model of the system, high efficiencies would confirm the validity of our models (Liu & Kersten, 1998). The present thesis uses the efficiency measure to control for the information provided in speed and depth discrimination tasks. The basic logic is to use the measure to make direct comparisons between different experimental conditions, specifically for discriminations of transparent stimuli and comparable non-transparent stimuli, and also between different tasks, specifically between speed discriminations and depth discriminations.

In the following sections I provide a more detailed discussion of the efficiency measure. First I trace the development of the efficiency approach, specifically the application of the approach to vision. I provide a series of definitions to emphasise the development and applicability of the approach. Second, I consider in further detail what an ideal observer is. I will provide a definition of the ideal observer, that uses the available information to implement the optimal

decision strategy. Third, I consider the causes of inefficiency. Finally, I review a broad range of studies that have employed the efficiency approach.

2.2   What is Efficiency?

The concept of efficiency can be traced to Fisher's (1925) discussion of statistics. Fisher defined an efficient statistic as that with the least possible variance, specifically it minimises the variance of the error distribution. The error is the discrepancy between statistical estimates of parameters based on a given sample with the actual parameters of the population. For those statistics that tend to the actual parameters of the population as the sample size is increased, the error distribution is normal. In Fisher's terminology, an efficient statistic is then one that makes use of all the relevant information available in the sample, while an inefficient statistic makes use of less than 100% of the relevant information available. The (relative) efficiency of a given statistic can then be assessed by comparing the variance of the error distribution for that statistic with that of the efficient statistic. A statistic with twice the variance of the efficient statistic can be said to have an efficiency of 50%. Therefore, efficiency measures the use of available information by comparison with the optimal use of information. This is the fundamental logic of the efficiency approach applied to perception.

Rose (1948) first applied Fisher's logic to the visual domain. Rose aimed to provide an absolute scale to measure performance of the 'human eye' to assess

how many quanta could be absorbed by the human eye. To achieve this, an optimal eye or 'ideal picture pickup device' was first defined, limited only by the inherent random fluctuations in the distribution of quanta (a statistical limit). Given a task of detecting a signal on a uniform background, the (theoretical) performance of the ideal device was then matched to that of a human observer by limiting the number of quanta available to be absorbed by the ideal device. Thus it was possible to estimate the *quantum efficiency*, the proportion of quanta used by the human eye. By this method Rose estimated that the human 'eye' has a quantum efficiency of 5%, in other words we are able to absorb 5% of the available photons. This approach was an interesting development, but it was limited to assessing the efficiency of the initial information pick-up. Barlow (1962a, 1962b) extended the quantum efficiency approach to address not only the initial pickup of visual information, but also the further processing and translation into performance of the visual information, aspects of visual processing of interest to modern day vision science. *Overall quantum efficiency* is defined as:

$$F = \frac{\text{Least quantity of light theoretically required for performing a task}}{\text{Least quantity required in practice for performing that same task}} \qquad (1)$$

But again, this definition was limited. By definition the overall quantum efficiency measure applies to tasks only where the discrimination of light intensities limits performance. Generally, we desire a measure of visual

performance that can assess a broad range of performance limitations, beyond the initial absorption of quanta.

Signal detection theory provides us with the method to compute human and ideal performance, and therefore the efficiency, for visual tasks more complex than the detection of light. Efficiency is defined as (Tanner & Birdsall, 1958):

$$\eta = \frac{E_i}{E_h} \qquad (2)$$

where E_h is the experimentally determined signal energy required by an observer to reach a specific performance level and E_i is the signal energy required by an ideal device to *match* the performance of the observer under study. Later, Barlow (1978) re-introduced the efficiency approach as *statistical efficiency*:

$$F = \frac{\text{Sample size required by ideal device}}{\text{Sample required by subject doing the same task}} \qquad (3)$$

This is computed experimentally as the ratio of human sensitivity to that of the Ideal Observer:

$$F = \left(d'_h/d'_i\right)^2 \qquad (4)$$

Here d'_i is the sensitivity of an ideal detector for the task, and d'_h is the experimentally determined sensitivity of the human observer for the same task. Barlow was particularly concerned with the detection of signals in random dot displays, however this definition applies to any task for which the human and ideal sensitivity can be computed, allowing us in principle to compute efficiencies for a broad range of complex tasks. In the following section I provide the general definition of an Ideal Observer.

2.3 What is an Ideal Observer?

As we have seen above, efficiencies can be computed for a visual task by comparing human performance with that of the ideal observer. As we have also seen, the ideal observer is the theoretical observer that makes use of all the available information to perform the given task in the optimal way. But how do we define optimal performance? The optimal strategy can be expressed given a statistical description of the task (Kersten, 1990). For example, a general task faced by an observer is to decide what the three-dimensional scene in the world is (the distal stimulus) given a particular two-dimensional image (the proximal stimulus). However, because many scenes could have given rise to any particular image, there is an uncertainty about the state of the world, the observer cannot be sure if a particular scene gave rise to the given image. Due to this uncertainty, we can say the task of the observer is to decide how likely a particular scene is, given a particular image. Bayes' theorem (e.g. Berger, 1985)

provides a way to break this probability down into parts, and lends itself well to an analysis of visual perception (for example see Mamassian et al., 2001). Essentially, the theorem allows us to re-express a conditional probability, given some knowledge about how the states of the world relate to the image, and some additional prior knowledge about plausible states of the world. Applied to present example, the theorem states that the probability of a particular scene given a particular image, the posterior (*a posteriori*) information, is given as:

$$p(s|i) = \frac{p(i|s)\,p(s)}{p(i)}$$

Here $p(s|i)$ is a distribution of the probabilities of possible scenes given a particular image. The right hand side of the equation expresses this in terms of a likelihood, a prior, and a normalising factor. The likelihood term, $p(i|s)$, represents the likelihood that a given image would arise from a particular scene. Thus it embodies knowledge about the possible scenes. The *a priori* term, $p(s)$, represents a distribution of the probabilities of the possible real world states i.e. it embodies knowledge about the plausibility of states of the world. The final ingredient in the equation, $p(i)$ is simply a normalising factor that ensures the posterior is a true probability distribution (i.e. that the integral over all scene values sums to 1). This formulation of the visual problem is important, as it permits us to specify an optimal decision strategy. If the task of the observer is to respond incorrectly as few times as possible, the optimal decision rule that

minimises the probability of error is to *maximise the posterior probability* (MAP). However, when the prior information is uniform (i.e. the prior probabilities are constant) the MAP strategy reduces to maximising the likelihood, or MLE (Kersten, 1990; Liu et al, 1995).

Similarly, the likelihood (of a particular stimulus) is also a fundamental concept in signal detection theory, but applies to a more specific experimental situation of deciding between two alternatives given a particular observation (e.g. Swets et al., 1964). In a basic detection task the observer must decide whether the sensory information arose from the signal (e.g. a spot of light) or alternatively was simply a result of the background noise (which may simply be the noise inherent in the sensory process). At the core, detection theory assumes that the amount of sensory information that arises from signal and noise together, and from noise alone, can be described by normal distributions. That is to say that, if we were to repeatedly present the same signal and measure the amount of sensory information it transmits, the frequencies of sensory information would follow a normal distribution. These distributions represent the likelihood of the stimulus (signal and noise, or noise alone) occurring given a particular amount of sensory information. The distributions are assumed to have equal variance, and the distance between these two distributions specifies the observer's sensitivity, such that the observer's sensitivity is linearly related to the strength of the signal. Therefore, each sensory response will have a likelihood that it resulted from

signal and noise, *and also* a likelihood that it arose from noise alone, so each

sensory stimulus will correspond to a ratio of the two likelihoods. To decide if a

signal was presented a decision criterion can be applied corresponding to a

particular value of likelihood ratio, e.g.

If $x > 1$ say 'yes'

If $x < 1$ say 'no'

where x is the likelihood ratio for a given trial. This particular criterion (a

likelihood ratio of 1) corresponds to the case where the two alternatives (signal

and noise, or noise alone) are equally likely to occur, but if these a priori

probabilities were to change (or any other aspect of the task were to change the

area of overlap between the distributions) the optimal criterion would be

different. Thus detection theory specifies the optimal decision criterion for a

given task. The ideal observer for a given task of distinguishing between two

alternatives will therefore compute the likelihood ratio for a given trial and apply

the optimal decision criterion to this value. The question remains, given the

optimal decision rule, how to define the ideal observer for a given task.

Detection theory specifies the ideal observer for a detection task, it computes the

likelihood by taking the cross-correlation of the observed stimulus with the

expected signal (p. 163 Green & Swets, 1966), this is known as *template matching,*

and applies the optimal decision rule to this computed likelihood. The template

matching procedure is consistent with the previous considerations of ideal performance, it makes use of all the available information and knowledge of the task, a perfect representation of the possible alternatives. There are mathematical proofs that this is optimal for the detection of known signals on backgrounds (Green & Swets, 1966) and has also been shown to apply to more complex object recognition tasks (Tjan et al, 1995; Liu et al, 1995). The present thesis applies the template matching ideal observer to speed and depth discrimination tasks (and a formal proof is provided in the Appendix).

2.4 Why are we Inefficient?

Efficiency normalises human performance to the available information, therefore it measures the ability of the observer to use the available information. What factors limit the ability to use all the available information? Causes of efficiency loss may be either internal noise in the system, arising from variability in the decision process (Burgess, 1990; Pelli, 1990) or in any neural firing (Tolhurst et al, 1983), or to a 'faulty memory' for the signal (Burgess et. al, 1981; Burgess & Barlow, 1983). Of course, what we are interested in is the effectiveness of the visual mechanisms, beyond the effects of any internal noise. Statistical efficiencies will reflect the contribution of both. A number of methods have been developed to factor out the contributions of internal neural noise to efficiency loss for simple tasks - *sampling efficiency* (Burgess et. al, 1981), *calculation efficiency* (Pelli, 1990; Bennett et al, 1999) and *high-noise efficiency* (Pelli & Farell, 1999) –

while decision noise can be assessed by a double-pass paradigm, in which the consistency with which observers perform a task on identical noisy stimuli across two sessions is assessed (Burgess, 1990).

Beyond the effects of internal noise, what form does inefficiency take? In a simple detection task e.g. of a Gabor patch on a background of Gaussian luminance noise, the ideal observer compares the stimulus with a template and this template is an exact match of the stimulus i.e. *the ideal observer knows the stimulus exactly*. In contrast, the human observer may not be able to use an exact representation of the stimulus i.e. the visual mechanisms do not match the stimulus exactly and so are not able to use all the available information. Instead, the visual mechanisms may provide a partial but incomplete representation, thereby introducing uncertainty into the stimulus. For example, Burgess et al (1981) found extremely high efficiencies for the detection of simple Gabor signals, the sampling efficiency was 83% on average. This indicates that human observers are able to use visual mechanisms that closely (but *not exactly*) match the stimuli, and based on the known physiology these are likely to be the 'simple' cells of V1. Such high efficiencies are rare. In the following section, I review the broad range of efficiency studies to assess the type of stimuli human observers can use effectively.

2.5 Summary of Efficiencies

In summary, by comparing human performance with an optimal baseline, the ideal observer, we can assess how efficient human observers are at extracting visual information to perform a given task. In essence, we normalise performance relative to the available information. Indeed, this fact can be exploited to test whether threshold differences in similar tasks are indeed due to differences in the underlying mechanisms, or differences in the available information. Moreover, we can indeed compare efficiencies directly between diverse tasks. Efficiencies have already been computed for a range of tasks over the years by various authors, and I summarise these in Table 1. The table includes a description of the basic task, the stimulus, and the maximum efficiency reported. Where efficiencies were reported for observers individually, I have computed the average efficiency across observers (mean number of observers across studies = 3). The table is ordered by maximum efficiency in descending order.

There are a number of interesting aspects of the efficiency literature. From examination of Table 1, it is clear that there is a broad range of efficiencies, from the close to optimal 83% of Burgess et al. (1981) for the discrimination of Gabors, through efficiency of 50% for the density discrimination of random dots (Barlow, 1978) down to the extremely small efficiencies of 0.05% for the direction

Authors	Year	Efficiency	Task	Stimulus
Burgess et al.	1981	83.00%	Signal Discrimination	Gabors
Legge, Gu & Luebker	1989	80.00%	Discrimination of Means	Numbers, scatterplots or luminance bars
Burgess et al.	1981	70.00%	Signal Discrimination	Gabors
Legge, Gu & Luebker	1989	70.00%	Discrimination of Variances	Numbers, scatterplots or luminance bars
Kersten	1987	59.50%	Signal Detection	Gabors
Burgess	1999	58.00%	Signal Detection	"Nodules", Simulated Tumours (Gaussian Noise)
Burgess	1999	58.00%	Signal Detection	"Nodules", Simulated Tumours (Power-Law Noise)
Burgess & Barlow	1983	54.00%	Numerosity Discrimination	Randomly Positioned Dots
Burgess	1999	51.00%	Signal Detection	"Nodules", Simulated Tumours
Barlow	1978	50.00%	Density Discrimination	Randomly Positioned Dots
van Meeteren & Barlow	1981	50.00%	Signal Detection	Sinusoidally Modulated Random Dots
Burgess & Ghandeharian	1984	50.00%	Signal Identification	Disks
Burgess, Li & Abbey	1997	50.00%	Signal Detection	Disks on Uniform & Lumpy Backgrounds
Burgess, Li & Abbey	1997	50.00%	Signal Detection	Gaussian on Uniform & Lumpy Backgrounds
Burgess & Ghandeharian	1984	50.00%	Signal Identification	Disks
Burgess & Ghandeharian	1984	50.00%	Signal Identification	Squares
Liu & Kersten	1998	50.00%	Object Recognition	Balls, Irregular, Symmetric, V-Shaped
Burgess	1999	45.00%	Signal Detection	Gaussian
Parish & Sperling	1991	42.00%	Letter Identification	Gaussian Filtered
Burgess	1985	40.00%	Signal Detection - 10AFC	Hadamard Signals
Eckstein et al.	2001	38.00%	Search (+ Saccades) - 10AFC	Gaussian-Blurred Disk
Kingdom et al.	1987	37.00%	Line detection - 500AFC	Sinusoidally Modulated Lines
Harris & Parker	1992	35.00%	Depth Discrimination	Random Dots
Watamaniuk	1993	35.00%	Direction Discrimination	Random Dots - Gaussian Direction Distribution
Eckstein et al.	2001	33.00%	Search (+ Fixation) - 10AFC	Gaussian-Blurred Disk
Burgess	1985	33.00%	Signal Detection - 2AFC	Hadamard Signals
Eckstein et al.	1997	31.20%	Signal Detection - 4AFC	Gaussian-Blurred Disk in Uniform Backgrounds
Kersten	1984	30.00%	Signal Detection	Gabors
Banks, Geisler & Bennett	1987	30.00%	Signal Detection - 2IFC	Sinusoidal Gratings
Barlow & Reeves	1979	25.00%	Symmetry Detection	Random Dot Patterns
Liu, Knill & Kersten	1995	25.00%	Object Recognition	Balls, Irregular, Symmetric, V-Shaped
Banks, Geisler & Bennett	1987	24.10%	Signal Detection - 2IFC	Sinusoidal Gratings
Barlow & Tripathy	1997	22.50%	Direction Discrimination	Random Dots
Eckstein et al.	2001	20.00%	Search (Saccades) - 10AFC	Gaussian-blurred Disk
Eckstein et al.	1997	18.70%	Signal Detection - 4AFC	Gaussian-blurred Disk in Different Backgrounds
Eckstein et al.	1997	16.60%	Signal Detection - 4AFC	Gaussian-blurred Disk in Repeated Backgrounds
Tjan et al.	1995	16.30%	Letter Recognition	Geneva Font
Liu, Kersten & Knill	1999	15.00%	Object Recognition	Balls, Tinker Toys and Wire objects
Burgess	1999	13.00%	Signal Detection	Simulated Microcalcification Clusters
Legge et al.	1987	12.00%	Contrast Discrimination	Disk
Solomon & Pelli	1994	11.50%	Letter Identification	Geneva Font
Knill, Field & Kersten	1990	10.00%	Texture Discrimination	Fractal Noise Textures
Hallett	1987	9.00%	Signal Detection	Blue-green stimulus - 5 flashes per intensity
Braje, Tjan & Legge	1995	8.40%	Object Detection	Wedges, Cones, Cylinders & Pyramids (Silhouettes)
Tjan et al.	1995	7.84%	Object Recognition	Wedges, Cones, Cylinders & Pyramids (Small Silhouettes)

Authors	Year	Efficiency	Task	Stimulus
Legge et al.	1987	7.50%	Contrast Discrimination	Gabor
Solomon & Pelli	1994	7.00%	Signal Identification	Gratings
Hallett	1987	6.40%	Signal Detection	Blue-green stimulus - 50 flashes per intensity
Simpson et al.	1999	6.00%	Speed Discrimination - 2AFC	Random Dots - Two Step Motion
Gold et al.	1999a	5.50%	Face Identification - 10AFC	Band-Pass Filtered and Unfiltered Faces
Tjan et al.	1995	4.74%	Object Detection	Wedges, Cones, Cylinders & Pyramids (Silhouettes)
Tjan et al.	1995	4.62%	Object Detection	Wedges, Cones, Cylinders & Pyramids (Line Drawings)
Barlow	1962	4.60%	Intensity discrimination	Flashes of Light
Tjan et al.	1995	4.51%	Object Recognition	Wedges, Cones, Cylinders & Pyramids (Large Silhouettes)
Braje, Tjan & Legge	1995	4.50%	Object Detection	Wedges, Cones, Cylinders & Pyramids (Line Drawings)
Braje, Tjan & Legge	1995	3.75%	Object Recognition	Wedges, Cones, Cylinders & Pyramids (Line Drawings)
Tjan et al.	1995	3.28%	Object Recognition	Wedges, Cones, Cylinders & Pyramids (Shaded)
Braje, Tjan & Legge	1995	3.23%	Object Recognition	Wedges, Cones, Cylinders & Pyramids (Silhouettes)
Tjan et al.	1995	2.69%	Object Recognition	Wedges, Cones, Cylinders & Pyramids (Line Drawings)
Gold et al.	1999b	2.00%	Face Identification - 10AFC	Unfiltered Faces
Tjan et al.	1995	1.53%	Object Detection	Wedges, Cones, Cylinders & Pyramids (Large Silhouettes)
Gold et al.	1999a	0.90%	Letter Identification - 10AFC	Band-Pass Filtered and Unfiltered Letters, Geneva Font
Gold et al.	1999b	0.60%	Texture Identification - 10AFC	Band-Pass Filtered Gaussian Noise
Watson et al.	1983	0.05%	Signal Detection	Gratings
Watson & Turano	1995	0.05%	Direction Discrimination	Gabors

Table 1. Summary of Reported Efficiencies. The table includes a description of the basic task, the stimulus, and the maximum efficiency reported for a broad range of studies. The table is ordered by maximum efficiency in descending order.

discrimination of translating Gabors (Watson & Turano, 1995). Thus, it appears that the visual system is more efficient in processing some visual stimuli than others. An important question to ask then is, is there a general trend in variation of efficiencies as a function of the task/stimulus? Although there is large variation in the values of efficiencies for similar stimuli and tasks, the higher efficiencies reported tend to be for simple detection or discrimination of simple stimuli, such as Gabors, disks and sinusoidal gratings. On the other hand, the lower efficiencies reported tend to be for the recognition or identification of more

complex stimuli, such as letters, objects and faces. This is an intriguing pattern, as it suggests that low-level image based stages of processing represent more of the available information than representations at high-level object based stages of processing. It is premature to make any strong conclusions here, without a detailed analysis of these patterns in efficiency, but it is interesting to note that this pattern of efficiencies may be consistent with an early "efficient coding" of natural images (e.g. Simoncelli & Olshausen, 2001).

It is clear that very little 'mid-level' work has been done using the efficiency approach. Notably, Harris & Parker (1992) calculated the efficiency for depth discrimination in random dot stereograms, and Watamaniuk (1993) calculated the efficiency for motion integration in random dot stimuli. Certainly nothing has been done on perceptual transparency. The resolution of transparency (in the natural world) is likely to require the integration of different kinds of visual information (Kersten, 1991). By the efficiency approach, we can compare performance directly for different types of information; therefore *we can begin to assess the type of visual information that is useful for the resolution of transparency, and the surface representation in general.* Crucially, because efficiency normalises performance to the available information, the measure permits stronger inferences to be made about the nature of the underlying mechanisms than other available measures of visual performance. This point cannot be overemphasised. A performance level for a particular task given by any other measure could be a

function of either a) the observers ability to use the available information (what we are interested in), or b) how useful the available information is for a given task. Without the thorough analysis of the information available to perform a given task, these two alternatives cannot be teased apart. In the two experimental studies that follow this chapter, I compute the efficiency of transparency defined purely by motion (Chapter Three) and stereopsis (Chapter Four). In Chapter Five, I extend the study of motion efficiency to the motor domain.

Chapter Three: The Efficiency of Motion Transparency

3.1 Introduction

Local motion signals, known to occur early (area V1) in the primate visual system, must be integrated and segmented. In general, the motion system must integrate local motion signals that arise from the same surface into a global, coherent motion, and segment motion signals that arise from different surfaces (Braddick, 1993). Transparent motion, in which two or more surfaces are perceived segregated in depth, is a particularly good stimulus to study the limitations of these motion mechanisms as it involves the simultaneous integration and segmentation of local motion signals. In this chapter I use the efficiency approach to assess the visual mechanisms underlying transparent motion. First I briefly introduce a fundamental problem in motion vision that has implications for the recovery of transparency, the motion correspondence problem, and then provide a brief summary of the previous research on transparent motion.

3.1.1 The Motion Correspondence Problem

The motion correspondence problem (Marr, 1982) occurs when a number of local elements at a time t_1 are displaced at a time t_2. There are a number of possible

matches between the elements, some correspond to the true displacements but some to spurious correspondences. Therefore, a fundamental task for the visual motion system is to identify those correspondences arising from true displacements in the world. The local motion signals of the early visual system cannot distinguish between a true and spurious correspondence. Therefore an interaction of local information is required to overcome the correspondence noise in the stimulus. Random-dot stimuli (e.g. Braddick, 1974) have provided a useful tool to investigate the correspondence problem, as they permit direct manipulations of the local elements that contribute to the correspondence problem. A number of studies that have employed these stimuli have suggested that the quantity of false correspondences that actually occur in a stimulus may be a crucial limiting factor for motion perception; direction discrimination performance (Williams & Sekuler, 1984; Barlow & Tripathy, 1997) and the maximum detectable displacement (Eagle & Rogers, 1996) are limited by dot density, thus the number of possible correspondences in the stimulus. The correspondence problem will be particularly severe in the case of transparent motion, as here the system must identify the correspondences belonging to two different surfaces simultaneously.

3.1.2 Previous Research

Transparency can be perceived in random dot stimuli purely from differences in motion (such as direction and speed). However, there is a performance cost

associated with this stimulus. While transparency can be perceived when two random dot stimuli are presented simultaneously in opposite directions (Mulligan, 1993; Murakami, 1997) motion detection thresholds are higher for such transparent motion stimuli than for each motion stimulus presented alone (Mather & Moulden, 1983) and for transparent motions in orthogonal directions (Lindsey & Todd, 1998). The maximum detectable displacement, D_{max}, is less for transparent motions in orthogonal directions than for single coherent motions (Snowden, 1989). Similarly, direction discriminations are impaired for superimposed transparent motions relative to segmented motions (Smith, Curran & Braddick, 1999). This cost in processing transparent motion has been interpreted in terms of inhibitory interactions between different directionally tuned detectors (Snowden, 1989). This account is consistent with the 'direction repulsion' effect, in which the perceived directions of transparent random dot displays are exaggerated when the angle between the different directions is within a critical value (Marshak & Sekuler, 1979; Mather & Moulden, 1980; Hiris & Blake, 1996; Chen, Matthews & Qian, 2001). Inhibitory mechanisms of this kind have been identified in area MT (Snowden, Treue, Erickson & Andersen, 1991). In contrast, V1 responses are not suppressed for transparent motion (Snowden et al., 1991). Moreover, the suppression of MT cell responses varies depending upon the spatial proximity of opposing dots, in a manner that parallels perceptual behavior (Qian & Andersen, 1994; Qian, Andersen &

Adelson, 1994a). This suggests that MT processing limits the perception of transparent motion.

Recent psychophysical evidence questions the directional inhibition account of the cost for transparency. Firstly, De Bruyn & Orban (1999) suggested that the suppressed MT responses for opposite direction transparent stimuli reflect sub-optimal responses to transparent stimuli. This was based on a psychophysical speed enhancement effect, in which observers overestimated the speeds of opposite direction transparent motions. Secondly, Masson, Mestre & Stone (1999) found a cost for transparent motions moving in the same direction compared to unidirectional coherent motions. This suggests that the cost for transparent motion cannot be entirely due to directional inhibition. An alternative account suggested by Masson et al. was that the cost for transparency reflects a cost for segmenting different motions, and involves different neural substrates for transparent and coherent motion. In support of this they found that speed tuning for transparent motion was low-pass, similar to V1 speed tuning functions, and the speed tuning for coherent motion was high-pass, similar to MT speed tuning functions. This account contrasts with physiological evidence suggesting MT limits transparent motion perception (Stoner & Albright, 1992; Movshon, Adelson, Gizzi & Newsome, 1986). However, previous psychophysical data supports this account of a local signal for segregation and a global signal for discrimination (Bravo & Watamaniuk, 1995).

3.1.3 Present Study

The experiments that follow test whether a difference in the available information in transparent and coherent motion stimuli may be contributing to the cost in processing transparent motion regardless of the particular directional combinations.   One difficulty in comparing performance for coherent and transparent motion is a difference in controlling for the dot density of the random dot stimuli. The total density of a unidirectional coherent motion can be equated to that of two transparent motions, moving in different directions (Mather & Moulden, 1983; Lindsey & Todd. 1988) or in the same direction (Masson et al., 1999). Here, the overall density of the coherent and transparent stimuli is the same. However, there are less dots moving in the same direction in the transparent interval. In other stimuli, the density of a single coherent motion can be equated to the density of one of two transparent motions (Mather & Moulden, 1983; Snowden, 1989, 1990; Smith et al., 1999). Here the number of dots moving in the same direction in each condition is the same, but the overall density differs between the two conditions. Despite these stimulus differences all these studies find a cost for transparency. The question I ask here is whether this cost is due to the difference in the available information in coherent and transparent motion conditions, or due to a difference in the way the stimulus is processed. To address this question I used the efficiency approach (see Chapter Two for a general review of the efficiency approach). The efficiency approach

has recently been applied to a range of motion tasks (Watamaniuk, 1993; Watson & Turano, 1995; Barlow & Tripathy, 1997; Simpson, Manahilov & Mair, 1999; Simpson & Manahilov, 2001), but has yet to be applied in an analysis of transparent motion perception.

I computed the efficiency for speed discrimination of coherent and transparent motion in two experiments. The use of a speed discrimination task is of course an indirect method to target the mechanisms underlying perceptual transparency. However, there are two serious limitations in using a more direct task to probe transparency: 1) if observers are asked simply to respond whether or not they perceive transparency or not (e.g. Adelson & Movshon, 1982; Stoner et al., 1990; Qian et al., 1994), the measurements will be confounded by criterion effects, and 2) even if the observer is forced to respond to one of two stimulus alternatives (either spatial locations or temporal intervals), this does not necessarily require the observer to represent both motions (Braddick, 1997). Moreover, the application of the efficiency approach to an analysis of perceptual transparency demands a well defined task. The two-alternative speed discrimination task used in these experiments avoids criterion effects, demands the representation of the speeds of both motions in a transparent stimulus, and lends itself well to an ideal observer analysis. The main goal of the two main experiments of this chapter was to make a general comparison between coherent and transparent motion efficiencies across a range of relevant parameters, to

assess whether there is indeed a processing limitation for transparent motion in opposite directions. In Experiment 1 I fixed the speeds of the stimuli and varied their dot density. In Experiment 2, I fixed the dot density and varied the speed. In both experiments I found a consistent cost in efficiency for transparent motion. In the additional Experiment 3, I again fixed the speeds of the stimuli and varied their dot density, but asked observers to respond if they perceived two surfaces. This experiment is included simply to provide a comparison between the results of the speed discrimination experiments and a more direct task.

3.2   General Methods

The methods common to all three experiments are described below. The manipulations unique to each experiment are described in those sections.

3.2.1 *Human Observers*

Three experienced psychophysical observers participated, 1 experimenter (JW), 1 postdoctoral researcher (EG) and 1 paid graduate student (RG). All observers had normal or corrected-to-normal visual acuity.

3.2.2 Apparatus

Stimuli were presented on a 17" Sony Trinitron monitor via a G4 Power Macintosh running MATLAB with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The maximum luminance of the display was 80.6 cd/m². The monitor refresh rate was set to 75Hz at a resolution of 832 by 624 pixels. The stimuli were viewed monocularly (right eye) in a dimly lit room at a distance of 573 mm. Each pixel subtended a visual angle of 0.035° by 0. 035°. Observers used a chin rest to stabilize head position throughout the experiment and fixated on a central white fixation point, a square of side 0.14°.

3.2.3 Stimuli

The stimuli consisted of randomly positioned signal and noise dots. Each signal dot was displaced by a fixed increment on each frame continuously, the exact increment depending on whether the signal was the *standard* or the *target*. The standard speed was fixed for all experiments, while the faster target speed was fixed for Experiment 1 but was varied in Experiment 2. Noise dots were randomly displaced on each frame, such that they reappeared with a uniform probability anywhere on the screen. All dots were white squares of side 2 pixels, subtending 0.07 by 0.07 degrees of visual angle, and were presented on a square black background, 7 by 7 degrees of visual angle. The remainder of the screen was set to the mean luminance of the stimulus (which varied with the dot

density), to maintain a uniform mean luminance across the entire display. In both experiments, the dot density was controlled as described in the Appendix.

3.2.4 Procedure

I presented opposite motion random dot displays in two conditions. In the coherent motion condition, each trial consisted of two random dot signals, presented sequentially in temporal intervals of 267ms duration. This stimulus duration lies beyond the temporal integration asymptote for coherent motion, estimated by Masson et al. (1999) to be approximately 65ms. In one interval the signal moved to the left, in the other the signal moved to the right. The direction of the *standard* and *target* motions was randomised across trials. Each trial was preceded for 1000ms by a fixation point, centred in the presentation window. The fixation point was present throughout each trial. There was an interval of 500ms between intervals, in which only the fixation point was present. The observer's task was to indicate the direction of motion of the faster stimulus, 'left' or 'right'. In the transparent motion condition, again each trial consisted of two motion signals, but now superimposed in the same interval of 267ms duration. This stimulus duration lies beyond the temporal integration asymptote for transparent motion, estimated by Masson et al. (1999) to be approximately 200ms. One signal moved to the left, the other moved to the right. Again, the observer's task was to indicate the direction of motion of the faster stimulus, 'left' or 'right'.

3.3 Experiment 1

My aim was to assess whether there is a processing limitation for transparent motions of opposite directions by comparing efficiencies for speed discriminations of transparent motions with efficiencies for speed discriminations of coherent motions. In this experiment I made this comparison over a range of dot densities, for a constant speed difference.

3.3.1 Methods

The basic methods were as described in the General Methods section.

3.3.1.1   Stimuli

For each trial two sets of dots were generated, one for the standard speed and another for the target speed. For each signal, a 'strip' of randomly placed dots was generated (a binary matrix), the width of which was the size of the image plus the total speed increments over the 10 frames. Sampling this strip at successive increments generated the subsequent frames of the movie. The increment corresponded to the standard or target speed. In the transparent condition, corresponding frames of the target and standard speeds were

superimposed. Before presentation of each frame of the stimulus, a proportion of noise dots were randomly placed in the image.

3.3.1.2 *Procedure*

Here I presented transparent and coherent random dot stimuli at a range of dot densities. I used dot proportions of 0.01, 0.02, 0.04, 0.08, 0.16 and 0.32, corresponding to 2.04, 4.08, 8.16, 16.3, 32.6 and 65.3 dots per squared degree of visual angle. The dot density refers to the total dot density of the stimulus. Therefore each interval of the coherent condition had a density of half the total value. The *standard* signal dots were displaced 0.07° (2 pixels) horizontally left/right on every frame, giving a speed of 2.63° s^{-1}. The *target* signal dots were displaced 0.14° (4 pixels) horizontally right/left on every frame, giving a speed of 5.26° s^{-1}. To limit performance, I presented the signals in a number of noise levels using the method of constant stimuli. I tested five high noise levels per condition and measured d' (Tanner & Birdsall, 1958) for each noise level tested. In both the coherent and transparent motion conditions each observer completed 20 practice trials with 0% noise to become familiar with the stimulus before beginning a session for a new condition. There were equal numbers of left faster and right faster trials. Each condition was blocked, with 40 trials per each noise condition (20 left faster, 20 right faster) for observers EG & RG, and 80 trials per condition (40 left faster, 40 right faster) for observer JW. Within each condition,

trials for different noise levels were randomly interleaved. Each observer participated in over 10 hours data collection.

3.3.1.3   Ideal Observer

The ideal observer for a given task makes use of all the relevant information in a given stimulus to perform that task optimally i.e. maximising the number of correct responses by performing a maximum likelihood estimate (Green & Swets, 1966). I provide a formal derivation of the Ideal Observer in the Appendix and here describe its implementation. For the experiments in this study, the ideal observer is facing the same speed discrimination task as any human observer. The ideal observer needs to represent the speeds displayed in the stimulus, compare these speeds to the speeds of the possible templates, and choose the appropriate template that best matches the speeds in the stimulus (Figure 3.1). The speeds of each stimulus are given by the cross-correlation across successive frames of the stimulus (see also van Doorn & Koenderink, 1982a). The cross-correlation function simply describes the quantity of matches at each speed with no loss of information. It is a representation of the stimulus, and does not implement any particular model of speed perception. Any other model, e.g. motion energy filtering, would reduce the information content. However, because the task is to discriminate only leftward or rightward displacements, the ideal observer needs only to consider horizontal displacements, information given by a simple one-dimensional cross-correlation. For the coherent stimulus

the speed correlation is performed separately for each interval, and then summed. For the transparent stimulus a single speed correlation is performed. At low external noise levels, the peaks of this speed correlation correspond to the standard and target signal speeds. This can be seen in Figure 3.1A for a transparent stimulus with 0.70 noise dots (0.30 signal dots), in which the target speed is moving to the right. The ideal algorithm computes the likelihood of each possible outcome by comparing the incoming stimulus with a number of 'templates'. Each template is a representation of the possible stimulus alternatives, correlations that peak at the expected speeds (Figure 3.1B). The exact speeds will correspond to the speeds presented within a given block of trials. In Figure 3.1B the possible alternatives are given for a disparity ratio of 2. To compute the likelihood of each possible outcome, the ideal algorithm cross-correlates each template with the stimulus. The ideal decision rule is then to choose the template that returns the largest cross-correlation value with the stimulus (Figure 3.1C), a maximum likelihood decision rule (Green & Swets, 1966). In the case of low external noise, the template with the highest value will correspond to the actual signal presented, and in Figure 3.1C the ideal observer indeed selects the correct template. However, at much lower signal levels the value of the incorrect template can be higher than that of the correct template. Only these occurrences limit the ideal observer performance.
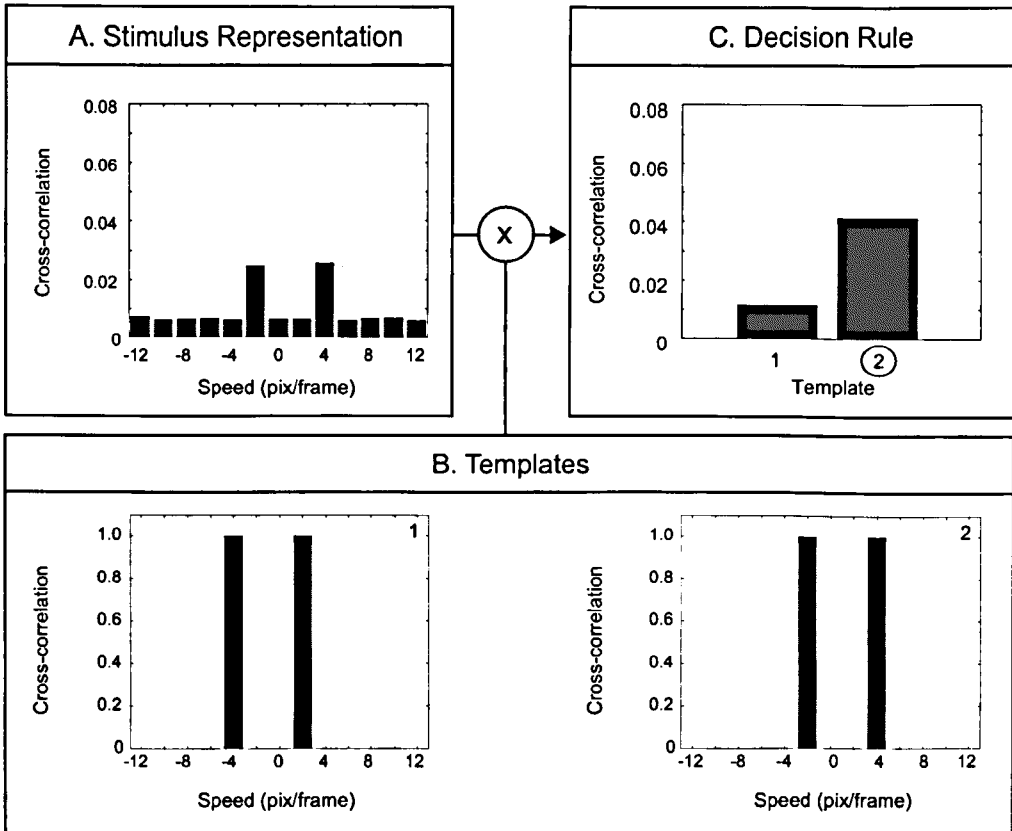
Figure 3.1. (A) *Stimulus Representation: The cross-correlation for a random dot display of 8% density, 50% noise, and a speed ratio of 2 (target moving to the right). The correlation peaks at a lag equivalent to a rightward displacement of four pixels per frame, and at a lag equivalent to a leftward displacement of two pixels per frame. (B) Templates: There are two memory templates for a speed ratio of 2. The template on the left represents a stimulus in which the leftward motion is faster. The template on the right represents a stimulus in which the rightward motion is faster. (C) Decision Rule: The ideal observer computes the correlation for each template of Figure B with the random dot display of Figure A. The ideal observer selects the template with the largest value, a maximum likelihood decision rule.*

The effects of varying the signal level and the dot density on the stimulus correlation, and therefore the predicted effects on ideal performance, can be seen in Figure 3.2. The left columns are correlations for stimuli of 16% density (d = 0.16), and the right columns are correlations for stimuli of 32% density (d = 0.32). The correlations represented by filled bars are for the coherent condition, and the correlations represented by open bars are for the transparent condition (the open bars are presented upside-down for better comparison with the filled ones). Each row contains correlations for a particular level of signal, the top row is for 100% signal dots (where the proportion of noise dots is zero, n' = 0), the middle row is for 50% signal dots (n' = 0.50), and the bottom is for 0.5% signal dots (n' = 0.995). First consider the effects of decreasing the proportion of signal dots (thereby increasing the proportion of noise dots). In the top row two peaks are clearly distinguishable; these correspond to the displacements of the signal dots. However, even with 0% noise dots, there are spurious matches at other displacements, *due to matching different signal dots.* I refer to this as the baseline level of the correlation, and the reader should be careful to distinguish this from the proportion of noise dots in the stimulus. The ideal observer selects the correct template because the amplitude of the baseline correlation is much lower than the peak amplitudes, which do correspond to the correct signal speeds. In the middle row the proportion of signal dots has dropped and the corresponding
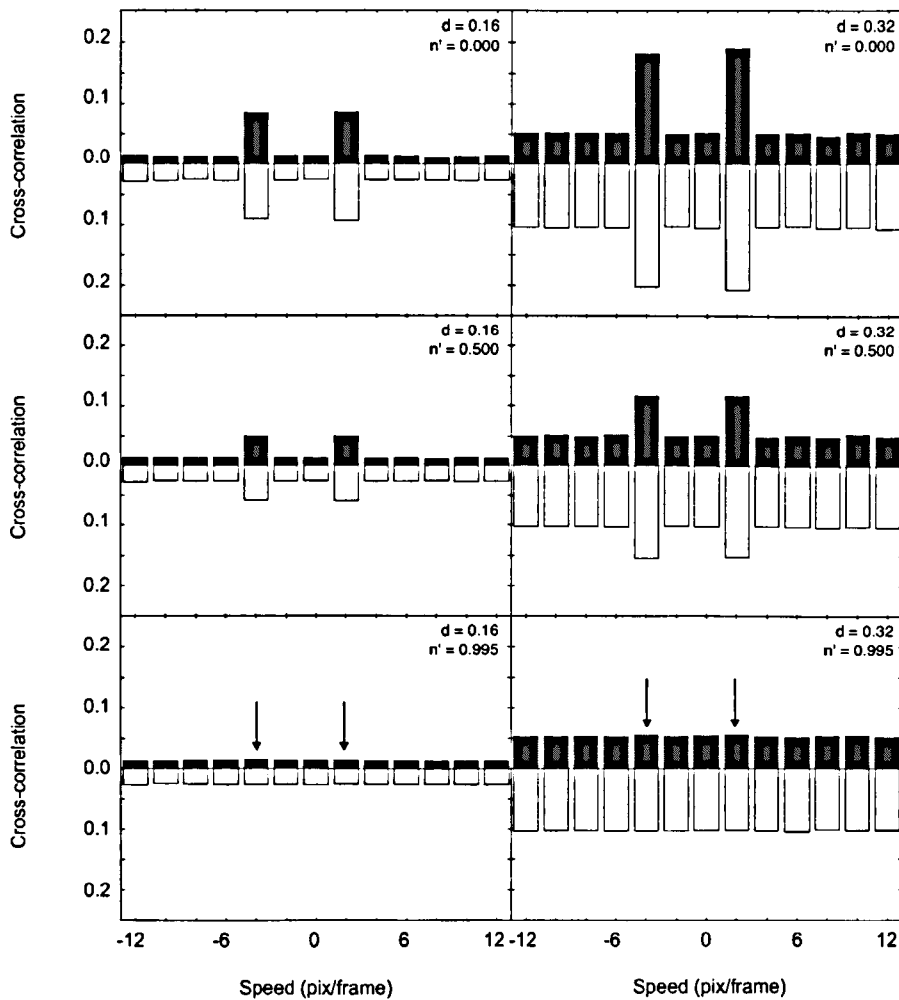
Figure 3.2. Cross-correlations for a number of stimuli, for each correlation 'd' indicates the dot density and 'n*' the proportion of noise (so $1 - n'$ is the proportion of signal dots). All the correlations are for a speed ratio of 2, in which the leftward motion is faster. Dark bars are for the coherent condition, and light bars are for the transparent condition. Increasing the noise level decreases amplitude of the peaks, whereas increasing the dot density increases the amplitude of both the peaks and baseline correlations. Note that the baseline correlations are stronger in the transparent condition than in the corresponding coherent condition. Arrows indicate the theoretical location of the peaks when the background noise is large.

56

peaks have also dropped, however the value of the baseline correlation has not changed. In the bottom row the proportion of signal dots has been decreased further still. Here dropped, however the value of the baseline correlation has not changed. In the bottom row the proportion of signal dots has been decreased further still. Here the peaks are no longer present in the transparent condition, but are still present in the coherent condition (this is not easily apparent in the 0.16 density correlation, but is clear for the 0.32 density condition). Now the ideal observer is just as likely to select the incorrect template as the correct template in the transparent condition, as the values for the incorrect speeds may be larger than the correct speeds by chance matches. However, in the coherent condition the correct template will be selected. This predicts that the ideal observer thresholds will be higher in the transparent condition. The second aspect of the correlations to consider is the effect of density. As density is increased two fold from the left column to the right column, it is clear that the amplitude of the baseline correlation increases. However, the peak amplitude also increases. Therefore, dot density will affect ideal performance if the increase in peak and baseline amplitudes differs e.g. if the peak amplitude increases proportionally more than the increase in the baseline amplitude then ideal performance should improve. I return to these aspects when considering the actual simulated data, but for now it should be noted that the peak and base amplitudes of the cross-correlation are not equivalent to signal and noise. While the baseline amplitude of the cross-correlation is simply a function of the number

of dots in the stimulus, the peak amplitudes depend on the relative proportions

of the signal and noise dots (and therefore also the number of dots).

To compute ideal sensitivity, I ran simulations of the ideal observer for both the

transparent and coherent motion tasks in the same conditions as the human

observers. The simulations were performed at five noise levels for each

condition, with 400 trials (200 left faster, 200 right faster) per noise level.

Efficiency is the ratio of human sensitivity to that of the ideal observer (Barlow,

1978):

$$F = \left(\frac{d'_h}{d'_i}\right)^2 \qquad (1)$$

The problem in using this definition is that the ideal observer easily reaches

ceiling performance for a suitable range of signal values for the human observer.

Thankfully, as we will see in the results section below, d' is a linear function of

the proportion of signal dots presented. I can therefore compute efficiency as the

squared ratio of the signal thresholds:

$$F = \left(\frac{\theta_i}{\theta_h}\right)^2 \qquad (2)$$

Causes of human efficiency loss may be either internal noise or inefficient sampling. The internal noise for the motion detection system is known to be low, equivalent to an external noise level of between 5% & 10% (Burns & Zanker, 2000). Therefore, any loss in efficiency can be attributed mainly to incomplete use of the available information.

3.3.2 Results

An example of the data obtained is shown in Figure 3.3, of a human observer and a set of simulation of the ideal observer. These data are for the transparent condition, with a dot density of 8%, and a speed ratio of 2. It can be seen that d' increases linearly as the proportion of signal dots is increased (and therefore as the proportion of noise dots is decreased), for both the human and ideal observers. A linear fit constrained to pass through the origin gave an excellent fit (r^2 = 0.89 for the human data, r^2 = 0.98 for the ideal data). I define the signal threshold (θ_h & θ_i) as the proportion of signal dots required for d' of 1. Note the much higher levels of noise required to limit performance of the ideal observer.

The ideal and human thresholds for both the coherent and transparent motion condition are shown in Figure 3.4. First I consider the performance of the ideal observer, shown in Figure 3.4A. There are two features to ideal performance.
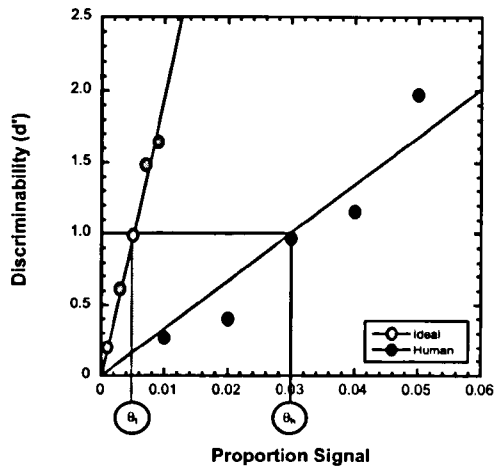
Figure 3.3. *Sensitivities for a human observer (black circles) and the simulated ideal observer (grey circles). A linear function gave very good fits to the data (ideal $r^2 = 0.98$, human $r^2 = 0.89$). It is clear that the slope of the fitted line for the ideal observer data is much steeper ($\alpha = 195$) than that of the human data ($\alpha = 33.6$). Thresholds (θ_i & θ_h) are taken at $d' = 1$.*

The first is that there is a performance cost for the ideal observer in the transparent condition. Transparent thresholds are consistently higher than that of the coherent condition, a greater number of dots are required for each signal in the transparent condition to attain an equivalent level of performance as the coherent condition. This confirms that the baseline correlation is indeed higher in the transparent condition than the coherent condition (shown in Figure 3.2). The second feature to these data is that ideal observer thresholds improve with increasing dot density in both the coherent and transparent motion tasks. This is somewhat counter–intuitive, as increasing the dot density increases the number of possible correspondences, which will raise the value of the baseline

60
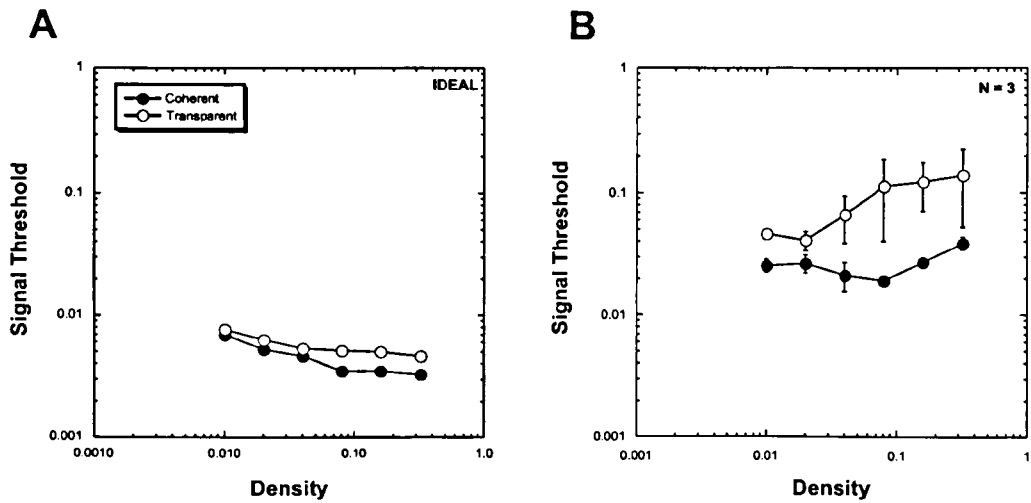
Figure 3.4. *(A) Signal thresholds for the ideal observer as a function of dot density. (B) Average signal thresholds for the human observers as a function of dot density. Error bars indicate the standard error across the three observers. Filled circles indicate coherent thresholds, open circles indicate transparent thresholds.*

correlation. However, increasing dot density will also increase the peak amplitude corresponding to the signal displacements. The improvement in performance suggests that increasing the dot density affects the peak and base amplitudes differently. To assess this I analysed the effect of dot density on the peak and baseline amplitude. I computed the average amplitude (across 400 trials) for transparent stimuli with a signal proportion equal to 1, and a speed ratio of 2 in which the rightward motion was faster. I then took the average of the peak amplitudes (that correspond to the two signal speeds that the ideal observer isolates with the correct template), and compared this to the average baseline amplitude (that correspond to the two signal speeds that the ideal
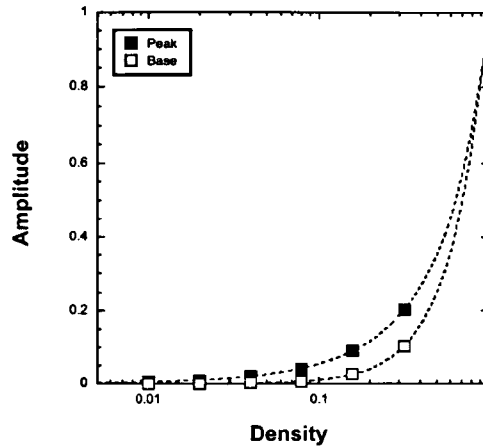
Figure 3.5. *The average peak amplitude (filled squares) and base amplitude (open squares) of the cross-correlation, as a function of dot density (speed ratio: 2). The dashed lines are taken from the closed form solution (see Appendix), and follow the simulations well. Note that the peak and base amplitudes do not increase at the same rate, thereby accounting for the effect of dot density on ideal performance (Figure 3.4A).*

observer isolates with the incorrect template). The average amplitudes for the peak and baseline correlations are plotted in Figure 3.5.

Figure 3.5 shows that the simulated data (filled and open symbols) follows the closed form solution provided in the Appendix (dotted lines). The different effect of increasing dot density on the peak and baseline amplitudes determines ideal performance. Recall that the effect of adding noise lowers the peak amplitudes corresponding to the signal displacements but has a negligible effect on the baseline amplitudes. Within the range of densities tested, at low densities a smaller proportion of noise will be required to bring the peak amplitude back
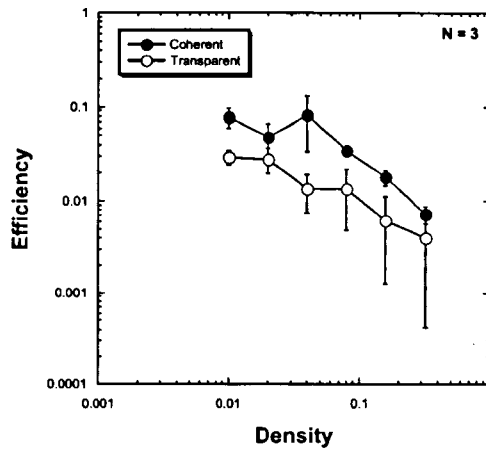
Figure 3.6. *The average efficiencies as a function of dot density (speed ratio: 2), for both the coherent (filled circles) and transparent (open circles) conditions. Error bars indicate the standard error across the three observers.*

to the baseline level, while at larger densities a larger proportion of noise will be required to return the peak to the baseline level. The peak and baseline amplitudes behave in the same way for coherent stimuli, with the exception that the baseline amplitudes are generally lower than the transparent baseline (accounting for the lower coherent thresholds).

Overall results for three human observers are shown in Figure 3.4B. There is an overall cost in human performance in the transparent condition, similar to ideal performance. However, the effects of density are not comparable to ideal performance. Overall, performance declines (the signal thresholds increase) as density is increased in both conditions, and the effect is greater in the transparent condition. For completeness, the results for the individual observers are presented in Figure 3.7B – C.

Figure 3.7. *(A) – (C) Coherent (filled circles) and transparent (open circles) signal thresholds for each of the human observers as a function of dot density (speed ratio: 2). (D) – (F) Coherent (filled circles) and transparent (open circles) efficiencies for each of the human observers as a function of dot density (speed ratio: 2).*

I assessed the cause of the performance loss in the transparent motion condition by computing the efficiency. The overall efficiencies for the three observers are shown in Figure 3.6. First, there is a residual cost in efficiency for transparent motion, indicating that the difference in stimulus information cannot account for the cost in human performance. Furthermore, efficiencies decrease as dot density is increased in both the coherent and transparent motion conditions. For completeness, the results for the individual observers are shown in Figure 3.7D – F.

3.3.3 Discussion

The main aim of this experiment was to compare performance between the coherent and transparent motion tasks. I found that signal thresholds were consistently higher for the transparent conditions. This finding is consistent with previous findings covered in the introduction (e.g. Mather & Moulden, 1983). However, the results extend these findings. I found that ideal observer thresholds were also higher for transparent motion compared to coherent motion, confirming that there is indeed a difference in the available information in the different conditions. Therefore we should be cautious about interpreting the previous findings where performance measures were not normalized relative to the available information. Nonetheless, I found that this difference in the stimulus information did not account entirely for the psychophysical cost for transparent motion. Transparent efficiencies are higher on average than the coherent efficiencies. This cost in efficiency for transparent motion indicates that constraints imposed by the visual system

limit performance for transparent motion. I consider possible mechanisms underlying this constraint in the General Discussion.

An interesting outcome of this experiment was that the efficiencies decreased as dot density increased in both conditions. We saw from the ideal observer analysis that the effect of increasing the dot density increases the level of spurious correlations in the stimulus. I think that the further decline in efficiency with increasing dot density suggests that the mechanisms underlying both coherent and transparent motion are increasingly impaired by these false correspondences. Indeed, this sensitivity to false correspondences may account for the low maximum efficiencies. The effect of density in both the coherent and transparent conditions can be considered in terms of the effect of density on the signal and noise amplitudes of Figure 3.5. We saw that the ideal thresholds initially improve because the peak and baseline amplitudes diverge with increasing dot density (within the range we tested). Clearly, the human observers cannot be taking advantage of the increase in the peak amplitudes with increasing density. Instead, observers appear to quickly reach a limit on the information that they are able to use effectively, their subsequent performance determined by the increase in false correspondences. This is demonstrated by the decay in efficiency with increasing density. There is a similar finding for the efficiency of stereopsis (Harris & Parker, 1992). I consider an account for this effect on performance in the General Discussion.

3.4 Experiment 2

In the second experiment I compared efficiencies for coherent and transparent motions across a range of speed differences, for a constant dot density.

3.4.1 Methods

3.4.1.1 Stimuli

The stimuli were random dot movies, as described in the General Methods section, constructed as described in Experiment 1.

3.4.1.2 Procedure

I presented transparent and coherent stimuli as described in the General Methods section. Here I used a constant density of 0.05 for all the conditions, equivalent to 10.23 dots/deg^2. This gives a density of 0.025 for each interval of the coherent condition. The standard speed was set to 2.63°s^{-1}. The target speeds were 5.26°s^{-1}, 7.89°s^{-1}, 10.5°s^{-1}, 13.2°s^{-1} and 15.8°s^{-1}. These correspond to speed ratios of 2, 3, 4, 5 & 6. I tested five high noise levels per condition and measured d' for each noise level I tested. In both the coherent and transparent motion conditions each observer completed 20 practice trials with 0% noise to become familiar with the stimulus before beginning a session for a new condition. There were equal numbers of left faster and right faster trials. Each condition was blocked, with 40 trials per each noise condition (20 left faster,

20 right faster) for observers EG & RG, and 80 trials per condition (40 left faster, 40 right faster) for observer JW. Within each condition, trials for different noise levels were randomly interleaved. Again, observers were required to indicate whether they perceived the leftward or rightward motion as faster.

3.4.1.3   Ideal Observer

The ideal observer for this task was identical to that described in Experiment 1 in detail. The quantity of matches of a given speed is given by the cross-correlation of successive frames of the stimulus. This is then compared with templates, by cross-correlation. The templates used by the ideal observer described the two possible speed combinations (the location of the peaks in the templates) for a given condition of speed ratio. The ideal observer then selects the template with the highest correlation, a maximum likelihood decision rule. Note that, for the ideal observer, the effect of changing the speeds will simply change the locations of the peaks in the correlations. Therefore, because the ideal observer will apply templates matched exactly to these speeds, there should be no effect of speed on ideal performance.
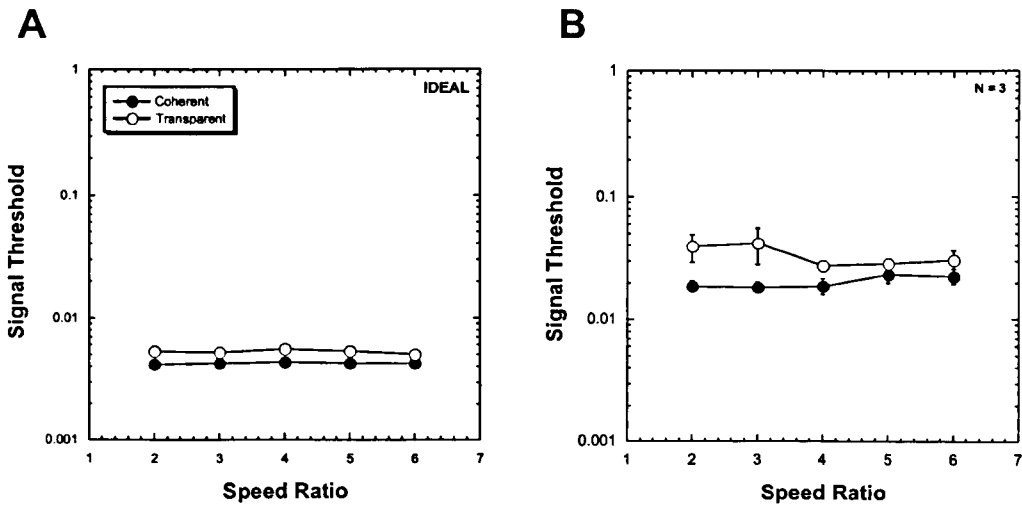
Figure 3.8. (A) *Signal thresholds for the ideal observer as a function of speed ratio (dot density: 0.05). (B) Average signal thresholds for the human observers as a function of speed ratio (dot density: 0.05). Error bars indicate the standard error across the three observers. Filled circles indicate coherent thresholds, open circles indicate transparent thresholds.*

3.4.2 Results

Ideal observer performance is constant across the speed ratios, but again displays a cost for transparent motion (Figure 3.8A). For human observers, thresholds are generally higher for transparent motions across the range of speed ratios tested (Figure 3.8B). For completeness, the individual data is shown in Figure 3.10A – C. It can be seen that RG is an exception to the trend, performance is impaired in transparency for this observer only at the two lower speed ratios tested.
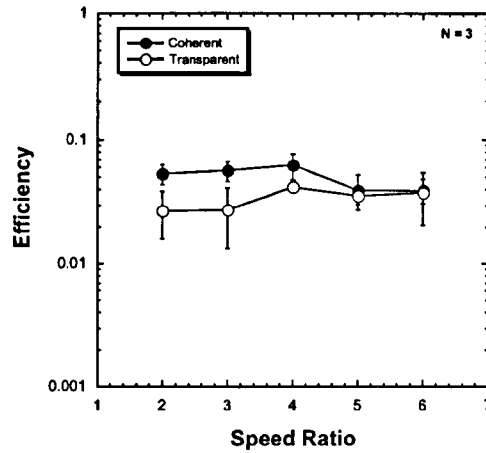
Figure 3.9. Average coherent (filled circles) and transparent (open circles) efficiencies as a function of speed ratio (dot density: 0.05). Error bars indicate the standard error across the three observers.

The average efficiencies for three observers are shown in Figure 3.9. Efficiencies in the transparent motion condition are consistently lower than for the coherent motion condition. For completeness, the individual data is shown in Figure 10D – F. It can be seen that observer RG has a cost in efficiency for transparency only for the smaller speed ratios tested (pulling the average efficiency up at the higher speed ratios). This may reflect the strategy reported by RG to try and attend only to the slowest speed, although it is hard to see how this strategy would work at threshold where the detection of the two speeds will be difficult.
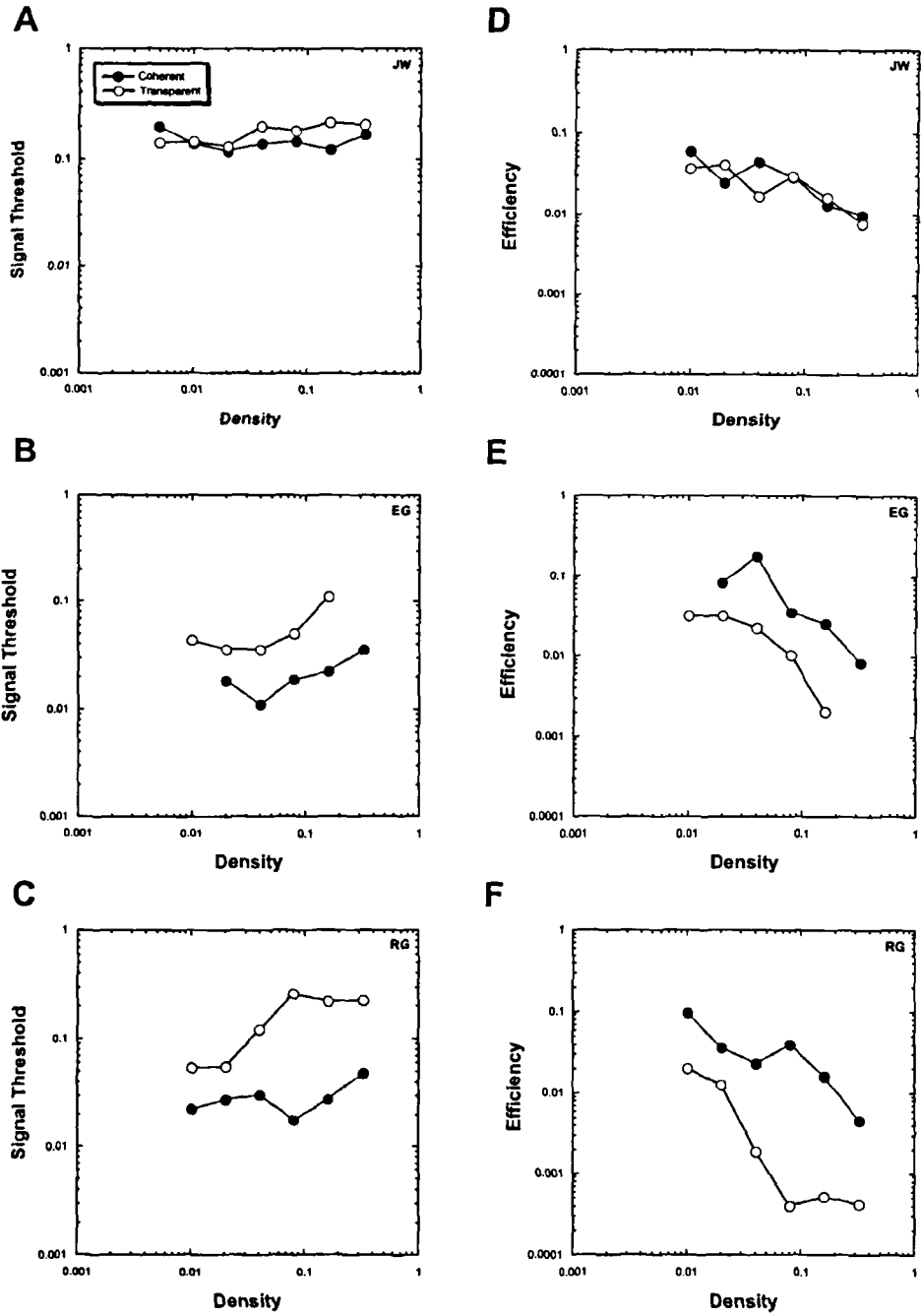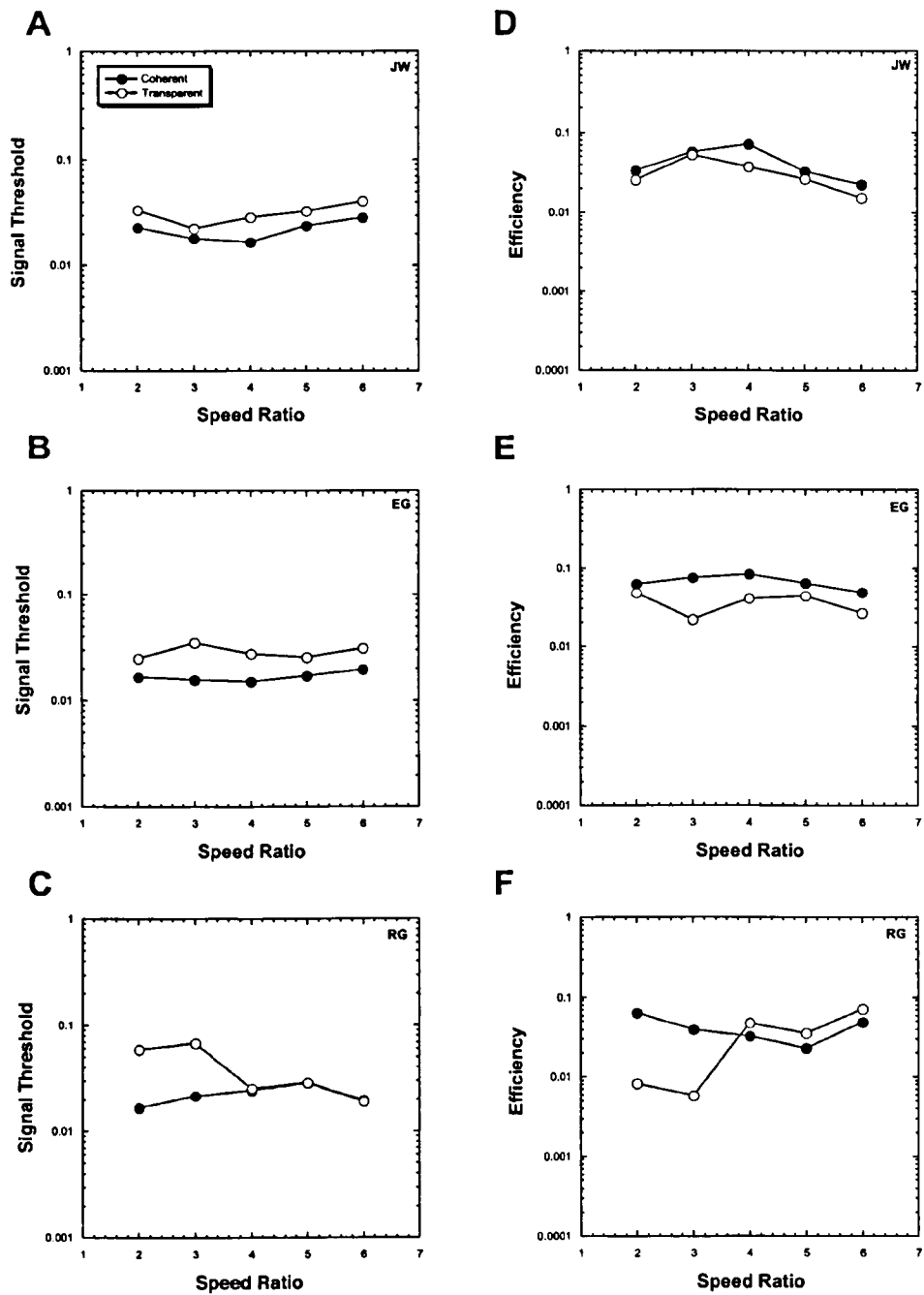
Figure 3.10. (A) – (C) Coherent (filled circles) and transparent (open circles) signal thresholds for each of the human observers as a function of speed ratio (dot density: 0.05). (D) – (F) Coherent (filled circles) and transparent (open circles) efficiencies for each of the human observers as a function of speed ratio (dot density: 0.05).

3.4.3 Discussion

In this experiment I computed thresholds for human and ideal observers across a range of speed differences. Again, I found that signal thresholds were consistently higher in the transparent condition. By computing efficiencies, I normalised human observer performance to the available information and found that transparent efficiencies were consistently lower than coherent efficiencies. This confirms the results of Experiment 1, demonstrating that a visual mechanism limits performance for transparent motions over a range of speed differences. A further aspect of the results is the effect of speed ratio (Figure 10a). Previous psychophysical (McKee et al., 1986; Masson et al., 1999) and fMRI (Chawla et al., 1999) results have suggested an optimal speed sensitivity of around $10°s^{-1}$. However, this behaviour is only hinted at by the present results, with a very slight peak at a speed ratio of 4 (corresponding to a target speed of $10.5°s^{-1}$). In fact, there is very little effect of speed. The similarity between this results and the performance of the ideal observer suggests that the human observers have access to a fine representation of the different speeds in the stimulus.

3.5   Experiment 3

Observers' reported that it was difficult to perceive surfaces at the lowest densities used in Experiment 1, but a clear surface perception accompanied the higher densities. To quantify this subjective change, I ran a short experiment in which observers were required to indicate whether they did

perceive two surfaces, a similar subjective task to that used in a number of influential studies of motion transparency (e.g. Adelson & Movshon, 1982; Stoner et al., 1990; Qian et al., 1994).

3.5.1 Methods

3.5.1.1   Stimuli & Procedure

The apparatus, stimuli and observers were identical to Experiment 1. Here, I used the adaptive QUEST procedure (Watson & Pelli, 1983) to find the signal thresholds for surface perception over a range of density conditions, from 1% to 32% of the available dot positions. I modified the QUEST procedure to use a cumulative Gaussian psychometric function and the mean estimate of King-Smith et al. (1994), and a threshold at 75% correct.   Each session for a particular density was terminated after a fixed number of trials (100). The density of the stimulus was constant throughout each session. The direction of the faster motion was randomly determined from trial to trial.   The proportion of signal dots varied from trial to trial depending on the current threshold estimate (mean of the posterior probability distribution function). I used the same standard (2.63°s^{-1}) and target (5.26°s^{-1}) speeds as Experiment 1, and the same presentation methods for both the coherent motion task and the transparent motion task. On each trial observer's indicated with a key press whether they perceived a coherent surface (in the coherent motion task) or two surfaces, one sliding over the other (in the transparent motion task).

3.5.2 Results

For two of the three observers (JW & EG, Figure 3.11A & B), the thresholds for surface perception are very similar in both the coherent and transparent conditions. Observer RG (Figure 3.11C) requires much more signal to
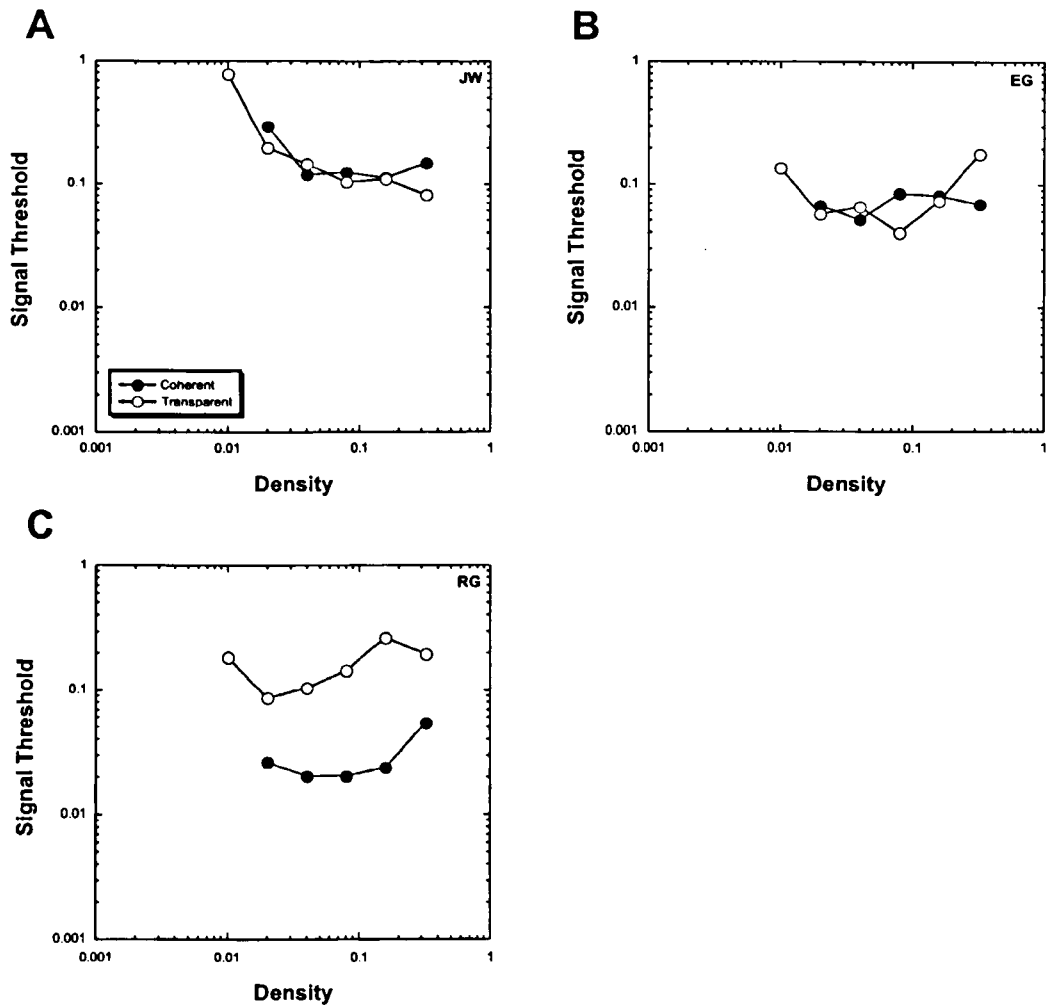
Figure 3.11. *Coherent (filled circles) and transparent (open circles) signal thresholds as a function of dot density for each of the human observers (A – C).*

perceive surfaces in the transparent case. Generally, across all the observers the surface thresholds are lowest in the range of densities from 1% to 10% in both conditions.

3.5.3 Discussion

This experiment set out to quantify the change in the perception of surfaces reported in Experiment 1. The results demonstrate that there is an effect dot density on surface perception, but this is not consistent across observers. Thus, while it was interesting to attempt to confirm that the stimuli of Experiments 1 & 2 were providing a surface perception, the problem with this approach is that the subjective task will be contaminated by a criterion for surface perception. Therefore the results do not necessarily reflect the mechanisms underlying the surface perception. It would be possible to devise a criterion free subjective task, by forcing the observer to choose between two stimulus alternatives, but even in this case it is not possible to establish whether the observer is really relying on a full representation of transparency (Braddick,1997). In this particular case, the surface perception thresholds do not capture the performance cost identified by the indirect method in Experiments 1 and 2. Indeed, the thresholds for surface perception are much higher than those for speed discrimination. This inconsistency between the results of this experiment and the previous two implies that the two methods are tapping into different processes, and while there are serious confounds in the present experiment, much stronger conclusions can be drawn from the indirect speed discrimination task.

3.6 General Discussion

3.6.1 *Summary*

I measured performance in terms of signal thresholds for speed discrimination of both coherent and transparent motion. From these data, I also computed efficiencies for these tasks by comparing human with ideal observer performance, thus normalizing performance relative to the available information. In both Experiment 1 and Experiment 2 I found that there is an overall cost in raw performance for transparent motion, consistent with previous findings (Mather & Moulden, 1983; Snowden, 1989; Lindsey & Todd, 1998; Smith, Curran & Braddick, 1999). Here I extended these findings through an ideal observer analysis and demonstrated that part of the loss of performance I found can be attributed to a difference in the available information in the transparent condition. Nonetheless, I found a consistent residual loss of efficiency in the transparent conditions. This indicates that constraints imposed by the visual system limit performance for transparent motion. However, the difference is small, generally less than 5%. Generally, I found that efficiencies for both coherent and transparent motion were less than 10%. Therefore, observers were using only a small sample of the available information. In Experiment 1 I found that speed discrimination efficiencies for both coherent and transparent motion depended upon dot density. This demonstrated that the mechanisms underlying both coherent and transparent motion are sensitive to the level of false correspondences in the stimulus, observers were less able to use all the available information the

76

greater the number of possible correspondences in the stimulus. Therefore we should be cautious when comparing performance for random dot stimuli with different overall densities, and also for random dot stimuli with the same overall density but with different densities contributing to different signals.

3.6.2 *Comparisons with other Efficiency Measures*

Generally the efficiencies were approaching 10%. The highest efficiency reported in the literature has been 83% for the discrimination of a gabor patch (Burgess et al., 1981). Other representative efficiencies are 50% for density discrimination of random dot displays (Barlow, 1978) and 50% (Liu, Knill & Kersten, 1995) to 2.69% (Tjan et al., 1995) for object recognition, and as low as 0.05% for grating detection (Watson et al., 1983). The range of efficiencies found for the motion tasks in the present study compares well with efficiencies of less than 10% reported by Simpson et al. (1999) for various motion tasks, using two-frame horizontal random dot jumps. This range is also similar to efficiencies reported for direction discrimination of random dot stimuli with a small direction distribution (Watamaniuk, 1993), suggesting that these studies are isolating similar visual mechanisms, however these efficiencies are also somewhat lower than found for direction discrimination of coherent motions (Barlow & Tripathy, 1997). It is worth noting that no absolute efficiencies reported in the literature approach 100%. However, this should not be taken as an indication that the human visual system is inherently sub-optimal. Rather, the visual system is not optimally configured

to perform any single psychophysical experiment, but rather multiple ecological tasks. A fruitful avenue for future research could be to devise stimuli (and tasks) with greater ecological validity in an attempt to maximize visual efficiency.

3.6.3 Correspondence Noise

I found that dot density, and therefore the level of false correspondences in the stimulus, limits performance in speed discrimination of both coherent and transparent motions. This effect of 'correspondence noise' has been explored using random dot stimuli by a number of authors (Braddick, 1974; Williams & Sekuler, 1984; Todd & Norman, 1995; Eagle & Rogers, 1996, 1997; Barlow & Tripathy, 1997). In particular, Barlow & Tripathy (1997) found that direction discrimination efficiencies improved as the ideal observer pooled information over increasing areas. This indicates that, for coherent motion stimuli, the visual system pools information over quite a large area, up to about 4 degrees of visual angle. This pooling is functionally significant, as it would serve to average out the effects of correspondence noise. This pooling operation could be limiting performance in both the coherent and transparent motion tasks. The spatial pooling of motion information would effectively reduce the available information, accounting for the low efficiencies found. However, this mechanism will cease to take advantage of increasing information when the available information exceeds the amount that can actually be pooled, performance would then increasingly be driven by the correspondence noise.

This would account for the decay of efficiency with increasing density. Therefore, spatial pooling of motion information provides a parsimonious account for the low efficiencies and the effects of dot density. There is a further account for the difference in performance level between the human and the ideal observer. The ideal observer considers only the horizontal displacements, given by the one-dimensional cross-correlation. However, while the ideal observer knows to look only for horizontal displacements, the human observers may be unable to isolate only the horizontal displacements from the full range of potential mismatches, the two-dimensional correspondence noise.

3.6.4 Visual Mechanisms underlying Transparent Motion

The processing limitation I found for transparent motion is supported by previous evidence for detrimental interactions between simultaneously presented motions of different directions (Snowden, 1989; Lindsey & Todd, 1998; Mather & Moulden, 1983). These psychophysical results have a physiological parallel, the responses of motion selective cells in area MT are reduced, or inhibited, in response to simultaneously presented motions of different (preferred and anti-preferred) directions compared to single (preferred) directions of motion (Snowden et al., 1991; Qian & Andersen, 1994). This effect has also recently been identified in human MT+ (Heeger, Boynton, Demb, Seidemann & Newsome, 1999). These results are generally consistent with models of motion detection that involve a stage of subtractive

inhibition between motions of opposite directions, these are the Reichardt detector models (e.g. van Santen & Sperling, 1985; Zanker et al., 1999; Zanker, 2001) and the alternative energy models (Adelson & Bergen, 1985; Qian, Andersen & Adelson, 1994b). A theoretical basis for this interaction is that a motion sensor that considers only the displacements (or the spatiotemporal energy) in one direction cannot unambiguously signal a particular direction of motion, indeed it cannot disambiguate motion from a static stimulus! Moreover, and of particular relevance to the present study, subtractive inhibition may serve to reduce correspondence noise (Qian & Andersen, 1994; Snowden et al., 1991).

However, the results do not rule out other possibilities. What I have shown is that, by normalizing performance to the information content of the stimuli by comparison with the ideal observer, a visual mechanism does indeed constrain performance for transparent motions. Because I used opponent transparent motions the finding is entirely consistent with directional inhibition. Two alternatives are also consistent with the findings. First, the coherent motions are presented sequentially, while the transparent stimuli are by their nature presented simultaneously. Perhaps the system is not effective at representing two global motions (surfaces) at the same time. In support of this idea, Braddick et al. (2002) found that observers were impaired in a global directional judgment for two motions compared to one, for both transparent motions and two coherent motions side by side. It would therefore be interesting to test whether performance, normalized to the stimulus

information, for transparent stimuli of opposite directions would be comparable to that of segmented coherent motions of opposite directions. Second, it remains a possibility that the cost for transparency reflects a cost for segmenting motions (Masson et al., 1999). To further explore this hypothesis, I suggest that comparisons should be made for unidirectional transparent and coherent stimuli, in which performance is normalized relative to the informational content in these different stimuli.

3.6.5 Conclusions

I found an overall cost in efficiency for speed discriminations of transparent motions compared to coherent motions. This demonstrates that constraints imposed by the visual system limit the processing of opponent transparent motions, consistent with a range of psychophysical and physiological evidence for directional inhibition. Efficiencies for speed discrimination of both coherent and transparent motions are less than 10% and decay with increasing dot density. This may be the result of a spatial pooling of motion signals.

Chapter Four: The Efficiency of Stereoscopic Transparency

4.1 Introduction

Perceptual transparency occurs when one surface is viewed behind another, such as when looking through transparent, reflective surfaces like glass or water. Here a scene is perceived through the transparent surface, yet the surface through which the scene is viewed is also perceived, due to reflections or specularities. The visual system successfully groups information arising from each surface and recovers two surfaces segregated in depth. In this chapter I use the efficiency approach to assess the stereoscopic mechanisms, those relying on the difference in information from the two eyes, underlying visual transparency. First I introduce a fundamental problem in stereoscopic vision, the stereo correspondence problem, and then briefly summarise the previous research on stereoscopic transparency.

4.1.1 The Stereo Correspondence Problem

One of the basic facts of human vision is that we receive visual information from two eyes. Because the eyes occupy different positions in space, they receive different views of the visual scene. A fundamental problem faced in understanding human vision is how the visual system integrates the different two-dimensional images transmitted by the two retinae to achieve a unified

'cyclopean' percept of the three-dimensional scene. From a geometrical perspective it is clear that binocular disparity, the difference in retinal positions between corresponding points in the two eyes, provides a cue to depth that the visual system can exploit to recover a veridical estimate of the three-dimensional scene. However, there is no *a priori* way that the visual system can know the corresponding points in the two images. Indeed, each point in one image could conceivably be matched with any point in the other image. The problem of this ambiguity has been termed the *stereo correspondence problem*. Indeed, this computational problem is of exactly the same form as the motion correspondence problem discussed in the previous chapter. While in the motion case there was an ambiguity in matching points at successive intervals of time, in the stereo case there is an ambiguity in matching points in the left and right images. In one case the problem is to identify the true displacements (differences in position between a monocular image at successive intervals of time), and in the other the problem is to identify the true disparities (differences in position between the left and right images at the same moment in time). The similarity between these computational problems invites a comparison between stereo and motion performance. To facilitate such a comparison, the design of the stereo experiments in this chapter are directly analogous to the previous motion experiments.
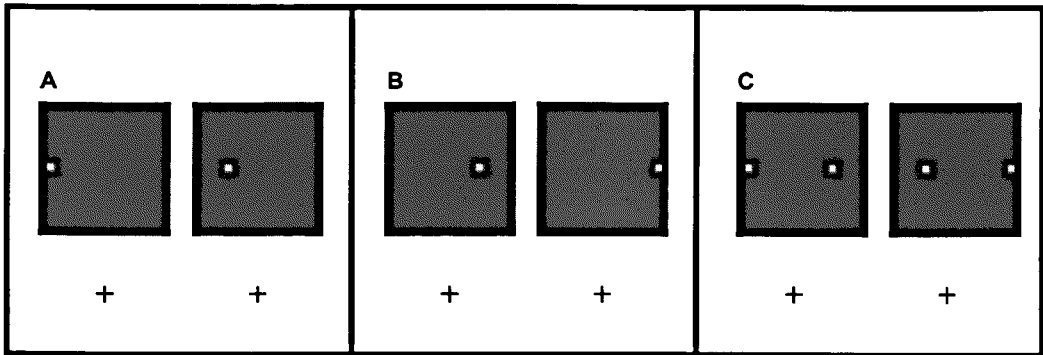
Figure 4.1. The correspondence problem. (A) & (C) a dot in a left window has only one partner in the right window. The dots are displaced relative to each other, and when fused this creates a sensation of depth. (C) the dot pairs in (A) and (B) are superimposed, such that each window contains two dots. Now there is a matching ambiguity, as each dot in one window could be matched to either two dots in the other window. This ambiguity will be more severe for random dot stereograms with greater dot densities, and multiple disparities.

Random dot stereograms (Julesz, 1964) have provided a powerful tool to investigate the stereo correspondence problem. These stimuli are constructed by randomly placing dots on a background. The resulting pattern is presented to each eye separately (for example, by way of a Wheatstone stereoscope), but with a relative displacement, or disparity, between corresponding dots in each eye. Despite the multitude of possible matches for each dot, and the absence of monocular cues to depth, such stimuli produce a vivid sensation of depth. This is illustrated in Figure 4.1. The stereograms of Figure 4.1A and Figure 4.1B are unambiguous, there is only one dot in each stereo-half; therefore there is only one possible match between the stereo-halfs. When fused, in each of these stereograms are fused one dot is perceived in depth. However, increasing the number of dots in each opaque stereogram would create a matching ambiguity. The stereogram of Figure

4.1C consists of the two previous stereograms superimposed. Now there is an ambiguity in the matching, each dot in a stereo-half can possibly be matched to two dots in the other stereo-half. How does the visual system resolve this ambiguity?

4.1.2 Previous Research

A particularly acute case of the correspondence problem occurs when two disparities are present simultaneously in the same visual location. The visual system successfully groups similar disparities and segments dissimilar disparities to recover two surfaces segregated in depth, despite the fact that points from each surface will project to only one point on the retina. Stereo algorithms that employ the uniqueness and continuity constraints of Marr & Poggio (1976, 1979) will be unable to recover such scenes, as they do not permit the occurrence of more than one disparity at a given visual location. Indeed, these constraints apply only to smooth opaque surfaces. Psychophysical studies using variations of random-dot stereograms have demonstrated that these constraints can be violated. Specifically, random dot versions of Panum's limiting case (Kaufman et al., 1973) and the double-nail illusion (Weinshall, 1989, 1991) can be perceived as one or more transparent surfaces in depth against an opaque background, although it has been argued that these percepts do not necessarily depend upon non-unique matches (Pollard et al., 1985, 1990). The PMF stereo algorithm (Pollard et al., 1985, 1990) implements the uniqueness constraint by restricting matching to a

disparity gradient limit (Burt & Julesz, 1980). This algorithm can recover isolated patches of the different disparities in transparent random dot stereograms, but does not interpolate these patches of disparity to recover two surfaces at different depths.

Despite the computational significance of stereoscopic transparency, the psychophysical research is surprisingly sparse. A few studies have assessed the limits of stereoscopic transparency with random dot stereograms that contain two disparities. For such stimuli, there is a continuum of percepts as the difference in disparity is increased (Tyler, 1991; Parker & Yang, 1989), from a single plane, through a thickened plane ('pyknostereopsis'), to transparency ('diastereopsis'). Observers will tolerate a disparity difference of only 3.6 arcsec to perceive a single plane, and up to 38 arcsec difference for a thickened plane, but for transparent surfaces the observers required up to at least 351 arcsec (5.85 arcmin) of disparity difference to detect two surfaces (Stevenson et al., 1989). Parker & Yang (1989) claimed that a low-pass spatial filtering of the image, followed by a cross-correlation between the two images, could account for this transition. A similar model that additionally includes an edge extraction mechanism can account for disparity attraction and repulsion effects (without the need for inhibitory interactions), contrast effects and effects of stimulus correlation (Stevenson et al., 1991; Cormack et al, 1991). With regard to a transparent stimulus, cross-correlating the left and right images of a stereogram would result in a function that peaks at locations corresponding to the disparities in the stimulus. When the images are first

convolved with a Gaussian filter, the peaks of the cross-correlation are Gaussian. Parker & Yang (1989) proposed that due to these combined operations, at larger disparity differences the two peaks are separate, and the two surface disparities can thus be easily extracted. However, when the difference in disparity is small, the two peaks would merge into a single peak at the average disparity of the two, accounting for the 'disparity averaging' of two disparities into a single plane. In fact, this account does not necessarily depend upon Gaussian filtering, but simply that there be some continuous representation of disparity and that the representation of a given disparity be distributed in some way. This account is attractive as it is not restricted to the disparity domain, it has been argued that the perception of transparent motion is also limited by the detection of separable peaks in a response distribution (Smith, Curran & Braddick, 1999), although some argue that the perception of transparency is not limited by the detection of separable peaks in a response distribution but rather by the width of the distribution (Treue, Hol & Rauber, 2000).

Akerstrom & Todd (1988) found that observers were less likely to perceive segregated transparent planes as the overall disparity, and the disparity difference, of the two planes was increased (the disparity differences were above the lower limits previously reported). In contrast, increasing the disparity did not impair the segregation of the opaque surfaces. Moreover, increasing the dot density of the transparent stereograms impaired the segregation of the two surfaces (but to a lesser extent when the two surfaces

were defined chromatically). Akerstrom & Todd (1988) argued that these results demonstrated both facilitatory and inhibitory interactions between different disparity detectors. In the transparent condition, disparity varies sharply across the image, and inhibitory interactions between different disparities would limit any facilitatory interactions. They argued that increasing the dot density would increase the strength of the inhibition, leading to the degraded perception of transparency they found. More recently, Gepshtein & Cooperman (1998) also argued for inhibitory interactions between differently tuned disparity detectors. They presented a random dot stereogram of a cylinder behind a transparent plane. Observers were required to report the orientation of the cylinder, horizontal or vertical. They found that, to perform at a particular level, observers required the dot density of the transparent plane to be lowered as the depth separation between the surfaces was increased. They argued that this behaviour could be accounted for by inhibitory interactions between disparity detectors. Indeed, adaptation experiments (Stevenson et al., 1992) and subthreshold summation techniques (Cormack et al., 1993) have provided evidence for inhibitory effects in disparity tuning, which could arise from a centre-surround receptive field structure or lateral interactions between disparity-tuned channels. Inhibitory (and excitatory) interactions are exactly the type of interactions proposed by Marr & Poggio (1976, 1979) to implement the uniqueness and continuity constraints (see also Grimson, 1985). This is in contrast, for example, to the stereo algorithm of Prazdny (1985) that includes only excitatory interactions to permit the resolution of transparency.

4.1.3 Present Study

In the present study I use the efficiency measure to quantify the limitations on stereoscopic transparency. Efficiency is an absolute measure of performance computed by comparing human performance with that of the ideal observer that utilises all of the information in a given stimulus to perform a given task optimally (Green & Swets, 1966; Barlow, 1978). Therefore, it is a measure of the amount of visual information actually used by a human observer to perform a task. Harris & Parker (1992) computed efficiencies for depth discrimination of random dot stereograms of two side-by-side surfaces at different depths. Human and ideal performance was limited by randomly perturbing the disparity of the dot pairs that constituted the surfaces to be discriminated. These authors found that the efficiency of depth discrimination for these opaque surfaces fell as the number of dots in the stimulus was increased, from about 20% efficiency at 4 dots to about 1% at 350 dots. Thus observers were less and less able to utilise all the available dots as the number of dots, and consequently as the number of potential matches, increased. Similarly, Cormack et al. (1997) found that the efficiency for detecting correlations in dynamic random dot stereograms also decreases with increasing dot density. This demonstrates that the effect is not task-dependent, but is indeed a property of the stereoscopic system. Harris & Parker (1992) suggested that the effect of dot density reflected a difficulty in solving the correspondence problem. Indeed, there is a parallel effect of density on the motion correspondence problem, as demonstrated in the

measurement of d_{max} (Eagle & Rogers, 1996, 1997) and the efficiency of motion discriminations (see Chapter Three).

Here I compute the efficiency for depth discrimination of transparent random dot stereograms and comparable opaque stereograms. In the transparent condition I presented two populations of dots at different disparities simultaneously, while in the opaque condition I presented each disparity sequentially. The key difference between these conditions is that there is a greater correspondence problem in the transparent case. Two different stereograms of different disparities are superimposed. Here there is a matching problem across the entire image, as by chance dots from one surface can be matched to dots in the other surface (or to other dots of the same surface). By comparing performance with an ideal observer that is only limited by correspondence noise (i.e. false dot matches) I could assess whether correspondence noise accounts for the impairment in performance for stereoscopic transparency.

4.2   General Methods

Here I describe the basic methods for the experiments. More specific details will be provided for each experiment.

4.2.1. Human Observers

Three experienced psychophysical observers participated, one experimenter (JW), & two paid graduate students (RG & VL). All observers had normal or corrected-to-normal visual acuity.

4.2.2. Apparatus

Stimuli were presented on a 21" Sony Trinitron Flatscreen monitor via a G4 Power Macintosh running MATLAB with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The monitor refresh rate was set to 75Hz at a resolution of 1152 by 870. The stimuli were viewed binocularly via a Wheatstone mirror stereoscope in a dimly lit room at a distance of 800 mm. Observers used a chin rest to stabilize head position throughout the experiment and fixated on a central white fixation cross, of length 0.30 degrees of visual angle.

4.2.3 Stimuli

The stimuli were random dot stereograms constructed by randomly placing dots on the left and right images and presenting these images separately to each eye via the Wheatstone stereoscope. Each image consisted of white squares ('dots') of side 0.075 degrees of visual angle on a black background, 7.5 by 7.5 degrees of visual angle. The remainder of the screen was set to the mean luminance of the stimulus (which varied with the dot density), to

maintain a uniform mean luminance across the entire display. A proportion of the dots were referred to as 'signal dots'; these dots corresponded to the projection of dots on surfaces located either near or far relative to the fixation plane. The remaining dots were referred to as 'noise dots'; these dots were randomly placed independently in each image to introduce disparity noise. Two examples of stimulus are illustrated in Figure 4.2A (without noise) & Figure 4.2B (with noise). In Figure 4.2A the stereogram contains only two disparities, corresponding to a 'near' transparent surface and a 'far' opaque surface. Figure 4.2B contains the same signal disparities, but now a proportion of dots are 'noise dots'. The effect of these added dots is to create more ambiguity in the matching, and results in the perception of dots at many depths. Indeed, some of these matches will be false correspondences resulting from incorrectly matching a signal dot with a non-corresponding noise dot. At the level of noise shown in Figure 4.2B it is still possible to perceive the two surfaces, but they are noticeably less clear. In both experiments, the dot density was controlled following the Appendix (Equations (1) – (4)).
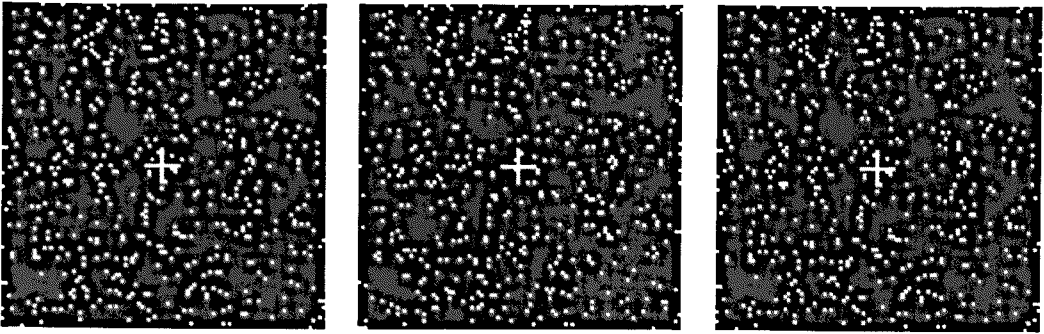
Figure 4.2A. This stereogram contains two populations of dots, one at crossed disparity and the other at uncrossed disparity. When the stereogram is fused (the two leftward panels are arranged for crossed convergence, and the two rightward panels for uncrossed convergence), a 'near' transparent surface is perceived in front of a 'far' opaque surface. Here the 'far' surface is further from the fixation cross.
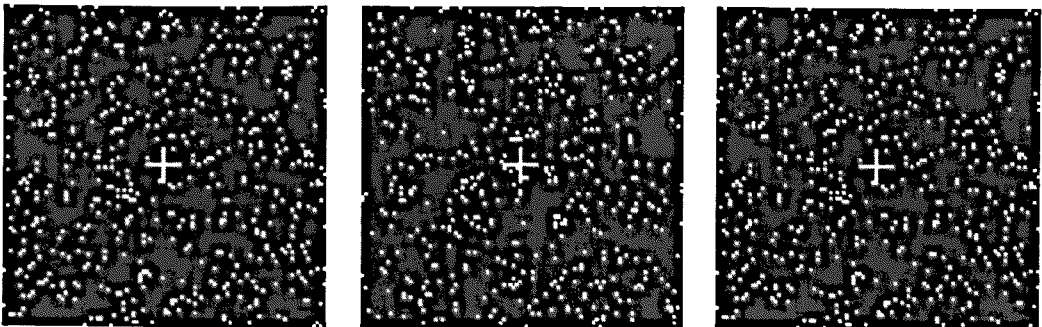
Figure 4.2B. This stereogram contains two populations of dots, one at crossed disparity and the other at uncrossed disparity. Here, a proportion of dots are 'noise', randomly placed in the left and right window. When the stereogram is fused (the two leftward panels are arranged for crossed convergence, and the two rightward panels for uncrossed convergence), it is still possible to perceive a 'near' transparent surface and a 'far' opaque surface, but they are now embedded in a cloud of dots and are harder to see than before.

4.2.4 Procedure

In all experiments, I presented random dot stereograms in two conditions. In the *transparent* condition, each trial consisted of two disparity signals superimposed in the same interval of 2000ms duration. Two observers (RG & VL) initially had difficulty in perceiving the transparent stereograms at this duration, but with little training on stereograms with a longer duration the observer could perceive the transparent stereogram at the shorter duration without difficulty. One signal (standard or target) was at uncrossed disparity (for a 'far' depth), while the other (target or standard) was at crossed disparity (for a 'near' depth). The depth (near or far) of the target stimulus was randomised across trials. To ensure fusion of the stereograms, each trial was preceded for 500ms by a fixation cross with nonius lines, centred in the presentation window. The fixation cross was present throughout each trial. In the *opaque* condition, again each trial consisted of two random dot signals, but now presented sequentially in temporal intervals of 2000ms duration each. In one interval the signal disparity was crossed for a near depth, in the other the signal disparity was uncrossed for a far depth. There was an interval of 500ms between intervals, in which only the fixation cross was present. The observer's task is illustrated in Figure 4.3. On each trial, the observer indicates which surface is perceived as being farther from the fixation plane.
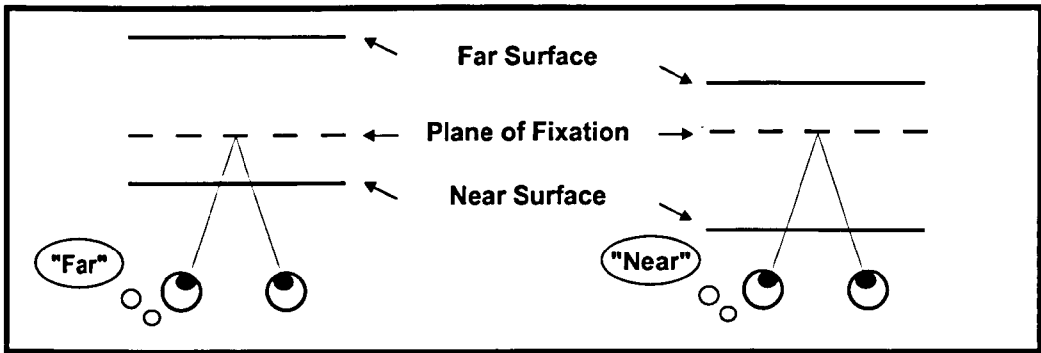
Figure 4.3. A cartoon illustration of the two stimulus alternatives, viewed from an overhead. The plane of fixation is defined by a fixation cross, and the near and far surfaces by dots at the appropriate disparities. Observers decide which surface is further from this reference plane. On the left of the illustration the far surface is further from the fixation plane, and the example observer makes the correct response. On the right of the illustration the near surface is further from the fixation plane, and the example observer makes the correct response.

The dot density of the transparent and opaque conditions was equated, such that each interval of the opaque condition had a density of half the total value. Therefore, a density of 4% corresponds to a 4% dot density for the transparent condition, but a 2% dot density for *each* interval of the opaque condition. This convention follows Akerstrom & Todd (1988) who also matched the dot densities of their transparent and opaque stimuli. However, they presented each opaque surface side by side, within the same stimulus area as the transparent stimulus such that each opaque surface was half the size of the corresponding transparent surface. Here the opaque surfaces are presented sequentially, therefore the average spacing of dots belonging to each surface is equivalent in the two conditions.

4.3 Experiment 4

The general aim was to assess whether there is a processing limitation for stereotransparency by comparing efficiencies of depth discriminations for both the opaque and transparent conditions. Specifically, in this experiment I compare the efficiency of depth discrimination in the opaque and transparent conditions across a range of dot densities. As dot density increases, the number of dots and therefore the number of possible correspondences between dots increases. If the mechanisms of stereopsis underlying performance in both the opaque and transparent conditions are sensitive to false correspondences, performance will be similarly impaired as dot density is increased.

4.3.1 Methods

4.3.1.1   Stimuli

For *each trial* two sets of signal dots were generated, one for the 'near' surface and one for the 'far' surface. One of these surfaces could be further from a zero-disparity fixation plane, while the other could be nearer. The surface further from fixation was defined by a *target* disparity, and the surface nearer to fixation by a *standard* disparity. For each surface, a 'strip' of randomly placed dots was generated (a binary matrix), the width of which was the size of the image plus the disparity. Sampling the strip at a horizontal increment generated the left and right images. This increment corresponded to either

the standard or target disparity. For example, for a disparity of 6 pixels (far depth) the strip would be sampled at +3 pixels for the left image and –3 pixels for the right image. This sampling increment results in corresponding dots to be uniformly displaced in each image at the appropriate disparity (because each image contained a displaced sample of the strip, a small proportion of 'signal' dots in each image had no corresponding points). In the transparent condition, the left images of each signal were superimposed, and similarly for the right images of each signal. Before presentation of the stimulus, a proportion of noise dots were randomly placed on the images, independently for the left and right images.

4.3.1.2   Procedure

The purpose of the first stereo experiment was to study the effect of dot density on depth discrimination for transparent and opaque random dot stimuli. I used a range of dot densities; 0.005, 0.01, 0.02, 0.04, 0.08, 0.16 and 0.32. These densities correspond to total dot numbers of 50, 100, 200, 400, 800, 1600 and 3200 dots, and to 1.78, 3.56, 7.12, 14.2, 28.5 and 57.0 dots per squared degree of visual angle. Recall that the dot density refers to the total dot density of the stimulus, such that each interval of the opaque condition had a density of half the total value. The observer's task was to decide whether the 'near' or 'far' surface was further from the fixation plane, a 2-AFC depth discrimination. The two possible alternatives ('far' is further from fixation, and 'near' is further) are illustrated in Figure 4.3 for the case of a transparent

stimulus. The standard disparity was fixed at 9 arcmin for all three experiments, while the larger target disparity was fixed at 18 arcmin, giving a *disparity ratio* of 2 (18/9). To limit performance, I presented the signals in a number of noise levels by the method of constant stimuli. I tested five noise levels per condition and measured d' for each noise level tested. In both the transparent and opaque conditions each observer completed 20 practice trials with 0% noise to become familiar with the stimulus before beginning a session for a new condition. There were equal numbers of near-further and far-further trials. Each condition was blocked, with 40 trials per each noise condition (20 near-further, 20 far-further). Within each condition, trials for different noise levels were randomly interleaved.

4.3.1.3   *Ideal Observer*

The ideal observer for a given task makes use of all the available information in a given stimulus to perform that task optimally i.e. maximising the number of correct responses by performing a maximum likelihood estimate (Green & Swets, 1966). For the experiments in this study, the ideal observer is facing the same depth discrimination task as any human observer. The ideal observer needs to represent the disparities displayed in the stimulus, compare these disparities to the disparities of the possible templates, and choose the appropriate template that best matches the disparities in the stimulus (Figure 4.4). The disparities of each stimulus are computed by cross-correlating the left and right images of the stimulus (see also Harris & Parker, 1994). These

images are simply binary matrices, in which '1' signals the presence of a dot and '0' is the background. The cross-correlation function describes the quantity of matches at each disparity, with no loss of information. It is not a model of the human stereoscopic system (although the cross-correlation function has been used as the basis of a model of human stereoscopic vision; e.g. Cormack et al, 1991). For the transparent stimulus a single disparity correlation is performed on the left and right images. For the opaque stimulus two disparity correlations are performed, one for each interval. The correlations for both intervals are then summed. At low external noise levels, the peaks of this disparity correlation correspond to the standard and target signals. This can be seen in Figure 4.4A for a transparent stimulus with 0.70 noise dots (0.30 signal dots), in which the far surface is further. The ideal algorithm computes the likelihood of each possible outcome by comparing the incoming stimulus with a number of 'templates'. Each template is a representation of the possible alternatives that were illustrated in Figure 4.3 ('far' is further or 'near' is further). These templates are correlations that peak at the expected disparities (Figure 4.4B). The exact disparities will correspond to the disparities presented within a given block of trials. In Figure 4.4B the possible alternatives are given for a disparity ratio of 2. To compute the likelihood of each possible outcome, the ideal algorithm cross-correlates the stimulus correlation with each template. The ideal decision rule is then to choose the template that returns the largest cross-correlation value with the stimulus (Figure 4.4C), a maximum likelihood decision rule (Green & Swets, 1966). In the case of low external noise, the template with the highest value

A. Stimulus Representation

C. Decision Rule

B. Templates
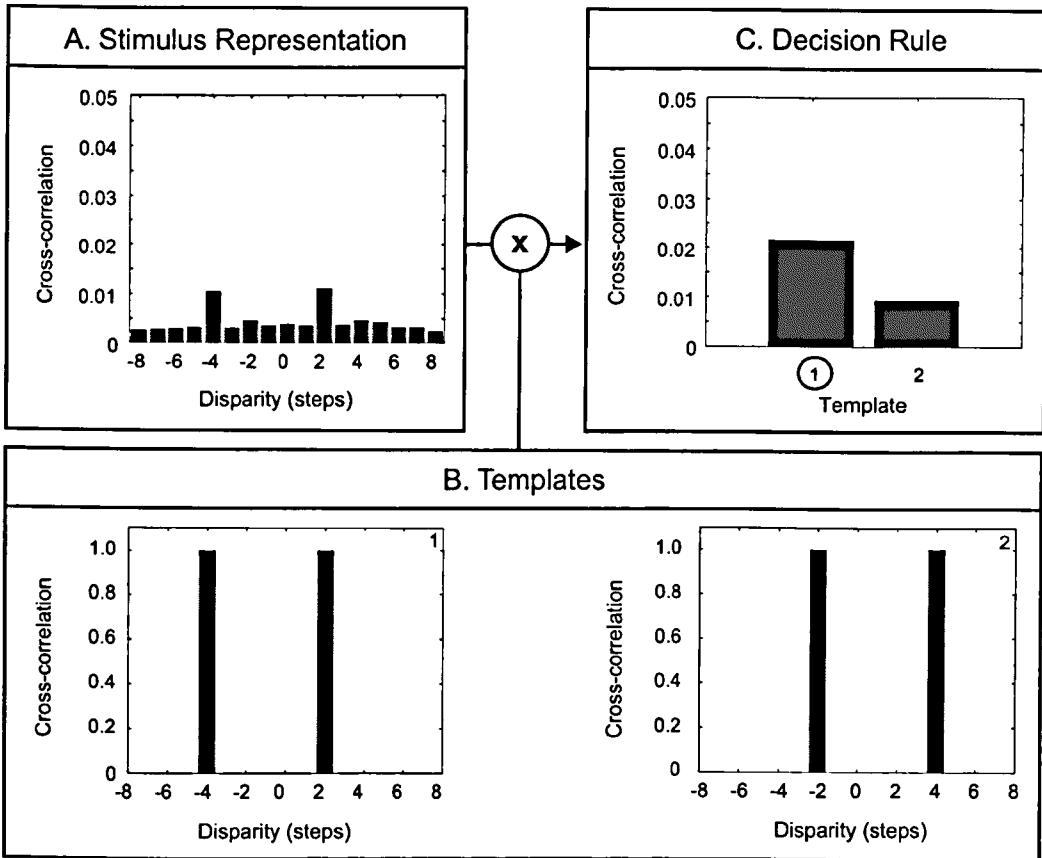
Figure 4.4. A schematic illustration of the ideal observer for the depth discrimination task of this study. (A) Stimulus Representation: This is the cross-correlation of the left and right images for a transparent stimulus, in which the 'far' surface is further from fixation (disparity ratio: 2; dot density: 0.05; proportion of signal dots: 0.30). The correlation peaks at a lag of –4 (a total uncrossed disparity of four dot steps), and +2. (B) Templates: These are the templates for a disparity ratio of 2. Template '1' on the left represents the stimulus in which the 'far' surface is further from fixation, and template '2' on the right represents the stimulus in which the 'near' surface is further from fixation. (C) Decision Rule: The computed correlations for template '1' and template '2' with the stimulus correlation. The correlation is largest for template 1, the correct stimulus, and is selected by the ideal observer.

will correspond to the actual signal presented, and in Figure 4.4 the ideal observer indeed selects the correct template. However, at much lower signal levels the value of the incorrect template can be higher than that of the correct template. Only these occurrences limit the ideal observer performance.

The ideal observer defined here contrasts with the ideal observer of the Harris & Parker (1992) study, which I summarize here for comparison. In Harris & Parker's (1992) study, observers were required to indicate which side of a stereogram stood out closer towards them in depth. Their ideal observer was given the disparity of each pair of dots in the stimulus (thereby assuming that the correspondence problem was somehow solved). It then computed the mean disparity of each side of the stimulus and chose the side with the largest mean disparity as being further towards the observer. Ideal performance was limited by randomly perturbing the disparity of the dot pairs, such that sometimes the mean disparity of one side of the stimulus could be larger than the other due to the random perturbation. This is in contrast to the present study, where performance is limited by varying the strength of the correlation at a particular disparity. In the present study the task is similar to Harris & Parker's (1992), in the sense that the ideal observer must decide which of two alternatives was presented (either 'near' surface or 'far' surface further away from fixation). Here, the ideal observer knows that the stimulus will always contain disparity signals corresponding to one of the possible alternatives in a block (these alternatives are also available to the human observer who can see the stimuli without noise before running a block). On a given trial, these

signals are presented together with noise, dots that have no partner in the other image. To decide which combination of disparity signals were presented, the ideal observer computes the quantity of matches at each disparity by cross-correlation. It then makes a maximum likelihood decision by matching templates that describe the possible combination of signals with the stimulus correlation (thereby discarding all spurious correlations that do not correspond to the expected signal disparities). Therefore, the ideal observer for the present study indeed uses all of the available information in the stimulus presented on a given trial to implement the optimal decision rule.

The effects of varying the signal level and the dot density on the stimulus correlation, and therefore the predicted effects on ideal performance, can be seen in Figure 5. The left columns are correlations for stimuli of 16% density (d = 0.16), and the right columns are correlations for stimuli of 32% density (d = 0.32). The correlations represented by filled bars are for the opaque condition, and the correlations represented by open bars are for the transparent condition (the open bars are presented upside-down for better comparison with the filled ones).
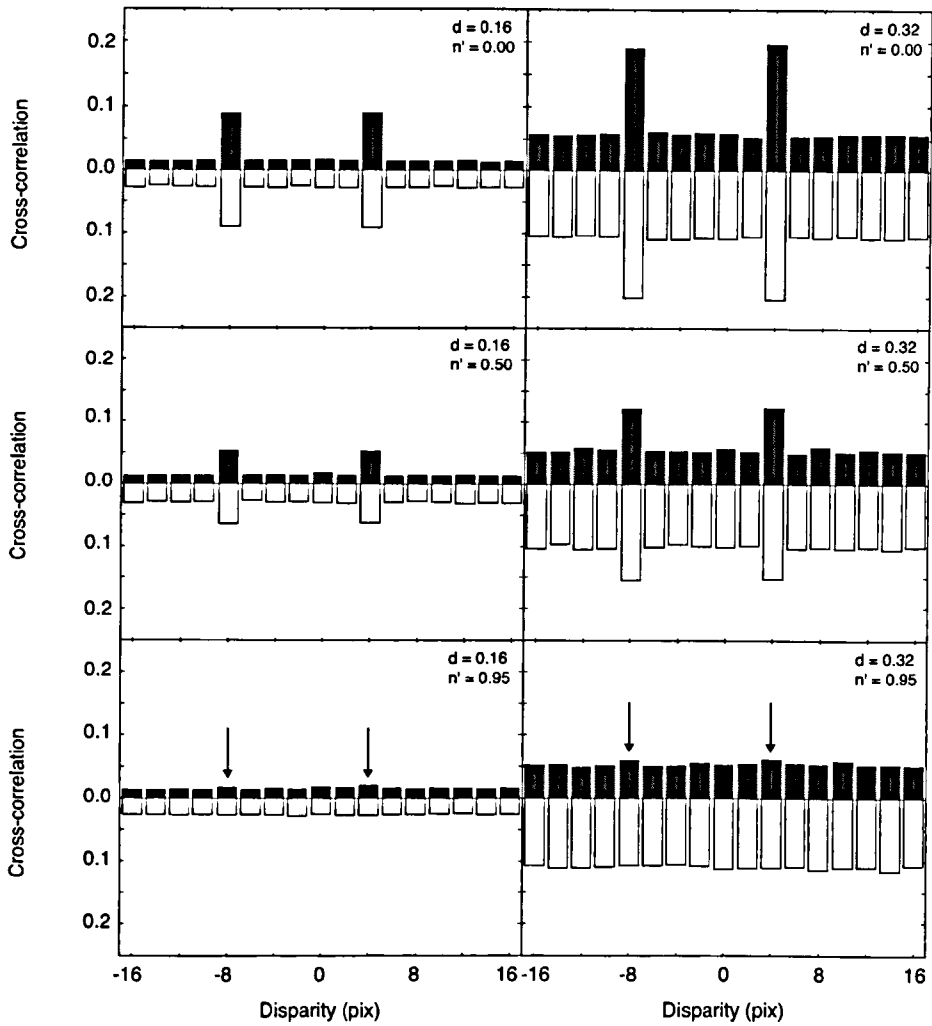
Figure 4.5. Cross-correlations for a number of stimuli, for each correlation 'd' indicates the dot density and 'n' the proportion of noise (so 1 − n' is the proportion of signal dots). All the correlations are for a disparity ratio of 2, in which the 'far' surface is further. Dark bars are for the opaque condition, and light bars are for the transparent condition. It can be seen that increasing the noise level decreases the strength of the signal, the peaks in the correlation. Increasing the dot density increases both peak strength and the baseline correlations. Note that the values of the baseline correlations are larger in the transparent condition than the corresponding opaque condition. The arrows indicate the theoretical location of the peaks in high levels of noise.

103

Each row contains correlations for a particular level of signal, the top row is for 100% signal dots (where the proportion of noise dots is zero, $n' = 0$), the middle row is for 50% signal dots ($n' = 0.50$), and the bottom is for 5% signal dots ($n' = 0.95$). First consider the effects of decreasing the proportion of signal dots (thereby increasing the proportion of noise dots). In the top row two peaks are clearly distinguishable, corresponding to the signal disparities. However, even with 0% noise dots, there are spurious matches at the non-signal disparities, due to matching different signal dots. I refer to this as the baseline level of the correlation. The ideal observer selects the correct template because the baseline values are much lower than the peaks. In the middle row the proportion of signal dots has dropped and the corresponding peaks have also dropped, and the baseline value has not noticeably changed. In the bottom row the proportion of signal dots has been decreased further still. Here the peaks corresponding to the two signal disparities are no longer distinguishable from the baseline correlations in the transparent condition, but are still present in the opaque condition (this is not easily apparent in the 0.16 density correlation, but is clear for the 0.32 density condition). Now the ideal observer is just as likely to select the incorrect template as the correct template in the transparent condition, as the values for the incorrect disparities may be larger than the correct disparities by chance matches. However, in the opaque condition the correct template will be selected. This predicts that the ideal observer thresholds will be higher in the transparent condition. The second aspect of the correlations to consider is the effect of density. As density is increased two fold from the left column to the right

column, it is clear that the values of the noise disparities increase. However, the value of the signal disparities also increases. Therefore, dot density will affect ideal performance if the increase in signal and noise amplitudes differs e.g. if the signal amplitude increases proportionally more than the increase in the noise amplitude then ideal performance should improve. I return to these aspects when considering the actual simulated data. I ran simulations of the ideal observer for both the transparent and opaque conditions. To compute ideal sensitivity, the simulations were performed at five noise levels for each condition, with 400 trials (200 near further, 200 far further) per noise level. Efficiency is the ratio of human sensitivity to that of the ideal observer (Tanner & Birdsall, 1958; Barlow, 1978):

$$F = \left(\frac{d'_h}{d'_i} \right)^2 \qquad (1)$$

The problem in using this definition is that the ideal observer easily reaches ceiling performance for a suitable range of signal values for the human observer. Thankfully, as we will see in the results section below, d' is a linear function of the proportion of signal dots presented. I can therefore compute efficiency as the squared ratio of the signal thresholds:

$$F = \left(\frac{\theta_i}{\theta_h} \right)^2 \qquad (2)$$

105

4.3.2 Results

An example of the data obtained is shown in Figure 4.6 for both a human observer and a set of simulation of the ideal observer. These data are for the transparent condition, with a dot density of 1%, and a disparity ratio of 2 (standard disparity 0.15 degrees of visual angle, target disparity 0.30 degrees of visual angle). It can be seen that d' increases linearly as the proportion of signal dots is increased (and therefore as the proportion of noise dots is decreased), for both the human and ideal observers. A linear fit constrained
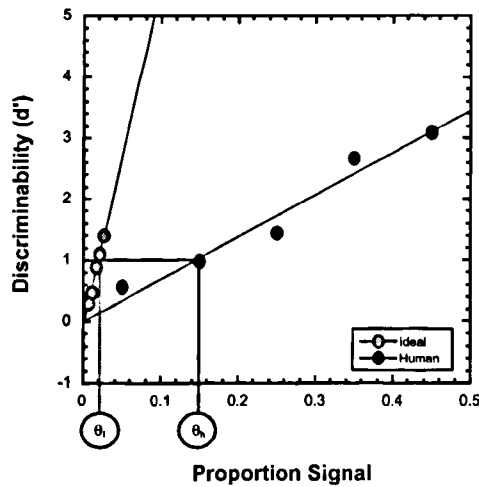
Figure 4.6. Sensitivities for a human observer (black circles) and the simulated ideal observer (grey circles). A linear function gave very good fits to the data (human r^2 = 0.96, ideal r^2 = 0.98). It is clear that the slope of the line to the ideal observer data is much more steep (α = 55.9) than that of the human data (α = 6.93). Thresholds (θ_i & θ_h) are taken at d' = 1.

to pass through the origin gave an excellent fit ($r^2 = 0.96$ for the human data, $r^2 = 0.98$ for the ideal data). I define the signal threshold (θ_h & θ_i) as the proportion of signal dots required for d' of 1. Note the much higher levels of noise required to limit performance of the ideal observer. Figure 4.7A plots the ideal signal thresholds as a function of the total dot density for both the opaque and transparent conditions. There are two features to these data. The first is that the ideal signal thresholds are consistently higher in the transparent condition than in the opaque condition, across the range of dot densities. This indicates that there is indeed a higher quantity of false matches in the transparent condition (shown in Figure 4.5). The second feature to these data is that ideal performance initially *improves* rapidly as dot density is increased, but levels off at around 5% density. This is somewhat counterintuitive, as increasing dot density increases the number of possible correspondences, which will raise the value of the correlation for the noise disparities. However, increasing dot density will also increase the strength of the signal (see Figure 4.5). The improvement in performance indicates that the signal strength initially improves faster than the strength of the correspondence noise (the heights of all the other peaks of the stimulus correlation), but these rates increase similarly from a dot density of around 5%.

Figure 4.7B plots the average signal thresholds for three human observers in the opaque and transparent conditions as a function of the total dot density.

A

B
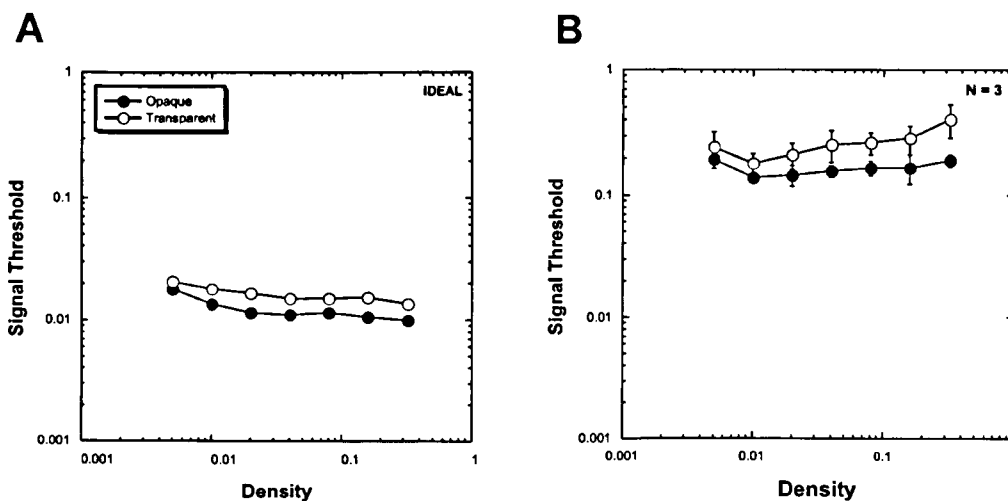
Figure 4.7. *(A) Opaque (filled circles) and transparent (open circles) signal thresholds for the ideal observer as a function of dot density. (B) Average opaque (filled circles) and transparent (open circles) signal thresholds for the human observers as a function of dot density. Error bars are standard errors across the three observers.*

The error bars are standard errors of the mean across observers. By comparison with Figure 4.7A, it is clear the performance is much worse than ideal performance in both opaque and transparent conditions. However, similarly to the ideal data, the thresholds for the transparent depth discrimination are consistently higher than those in the opaque condition. Thus, more signal dots are required to perform depth discrimination in the transparency case at an equivalent level of performance as the opaque case. The effect of dot density on human observer performance contrasts with the ideal observer, performance declines as dot
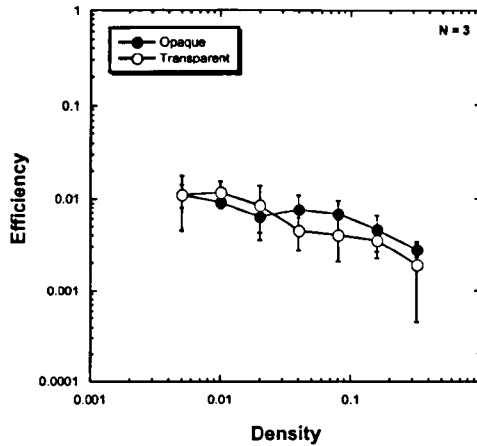
Figure 4.8. Average opaque (filled circles) and transparent (open circles) efficiencies as a function of dot density. Error bars are standard errors across the three observers.

density is increased. Thresholds for the individual observers are provided in Figures 4.9A – C.

Figure 4.8 plots the overall efficiencies for the three observers as a function of dot density. Error bars are standard errors of the mean across observers. The efficiencies for the opaque and transparent condition are approximately equal. The cost in performance (higher signal thresholds) for depth discrimination of transparent surfaces does not translate into a lower efficiency, but is in fact removed by comparing human performance with that of the ideal observer. A second aspect of this data is that efficiency decreases similarly for both the opaque and transparent conditions as dot density increases. Efficiencies for the individual observers are provided in Figures 4.9D – F.
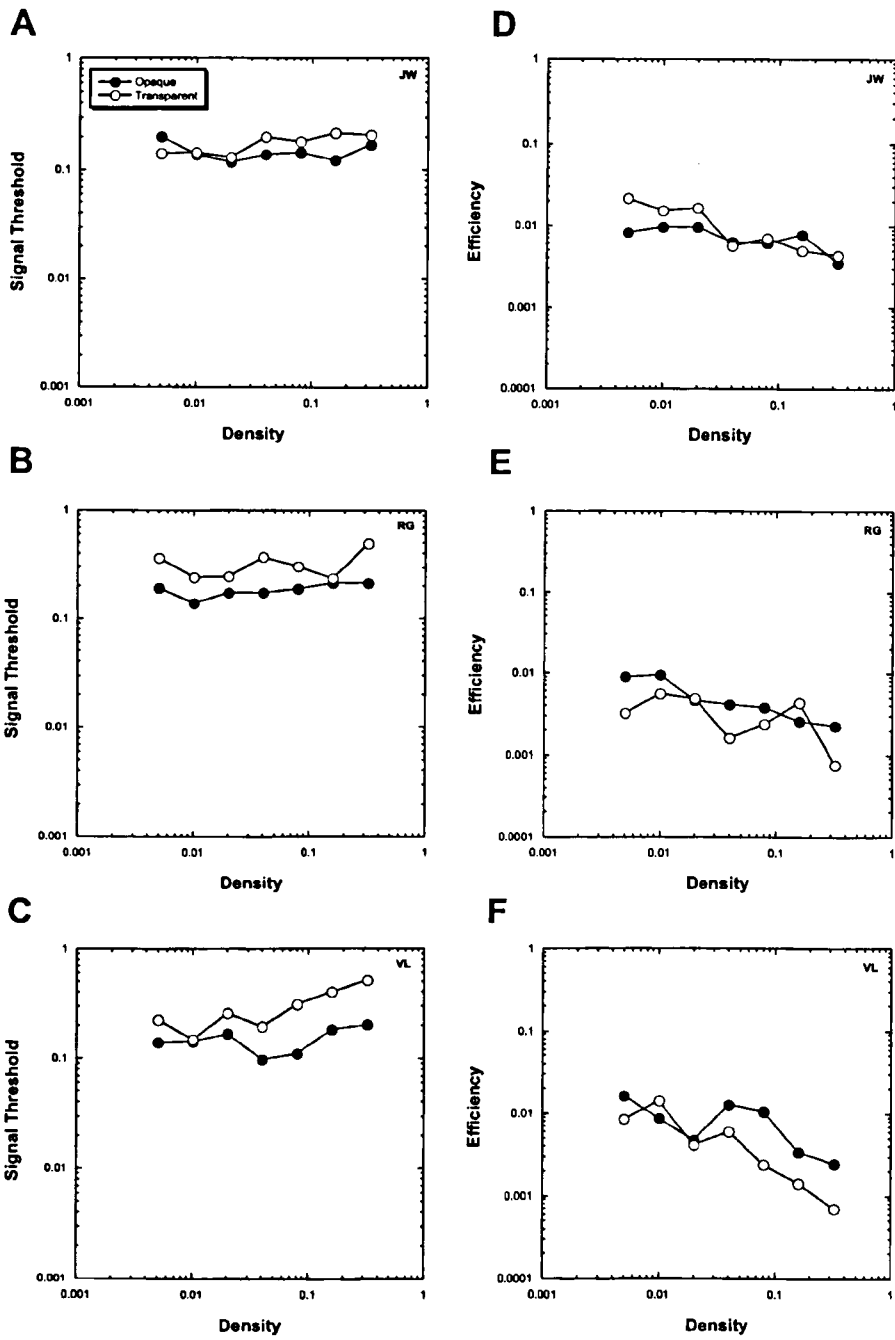
Figure 4.9. (A) – (C) Opaque (filled circles) and transparent (open circles) signal thresholds as a function of dot density for each of the human observers. (D) – (F) Opaque (filled circles) and transparent (open circles) efficiencies as a function of dot density for each of the human observers.

110

4.3.3 Discussion

In this experiment I compared human performance for depth discrimination of transparent and opaque surfaces as a function of dot density. The first issue to address was whether discrimination performance was impaired for transparent stereograms. I found that the human observers' signal thresholds were higher in the transparent condition than the opaque condition. This could imply that there is an additional limitation on performance in the transparent case, such as inhibitory interactions between different disparities. However, I found that ideal thresholds were also higher in the transparent condition, which indicates an informational limit on performance. By computing the efficiency, I could assess the relative cost between human and ideal observer performance. Indeed, I found that efficiencies were approximately equal in the two conditions. Therefore, the limitations on ideal performance account for the limitations on human performance. In other words, false matching accounts for the higher thresholds in the transparent condition, for both the human observers and the ideal observer. The similarity in the opaque and transparent efficiencies also indicates that there is no additional processing limitation in recovering depth from transparent stereograms (at least at the depths tested here), suggesting that the same mechanism underlies performance in both conditions.

The effect of dot density confirms that this mechanism is limited by correspondence noise. Efficiency decreases similarly with increasing dot density in both opaque and transparent conditions, indicating that the human

observers are increasingly impaired as the number of potential matches increases. Indeed, the maximum efficiencies are around 1%, indicating that human observers use far less information than is available to perform the task (e.g. human observers use only a proportion of the available disparity samples). We can think of this limitation in terms of the stimulus cross-correlations in Figure 4.6. From Figure 4.6 it was shown that decreasing the level of signal decreased the height of the peaks in the correlation. If human observers cannot use all of the available disparity information, these peaks will be lower than the ideal case (and so will indeed require more signal dots than the ideal observer to raise the peaks above the background correspondence noise). This result confirms the finding of Harris & Parker (1992) in which the efficiency of detecting a step-change in depth declined as the number of dots in their stereograms was increased, and also the findings of Cormack et al. (1997) in which the efficiency for detecting correlation in dynamic random dot stereograms decreased with increasing dot density. The present experiment extends these findings, demonstrating this effect of density is true also for depth discrimination of transparent stereograms. The similarity in the findings across these different studies is striking given the differences in stimuli, task, and the consequent ideal observer. This encourages the view that all the studies are tapping into the same correspondence noise limited mechanism, and furthermore demonstrates that absolute measures of efficiency can indeed be meaningfully compared across studies.

4.4   Experiment 5

In Experiment 4 I fixed the disparities of the standard and target surfaces, resulting in a constant disparity ratio, while varying the dot density. Here I aimed to see if equal efficiencies are found in the transparent and opaque conditions across a range of disparity ratios, keeping the dot density constant. The results of both Akerstrom & Todd (1988) and Gepshtein & Cooperman (1998) would predict that performance in the transparent condition should be increasingly impaired as the disparity between the two surfaces is increased.

4.4.1 Methods

4.4.1.1   Stimuli

The stimuli were random dot stereograms as described in the General Methods section, constructed as described in Experiment 4.

4.4.1.2   Procedure

I presented transparent and opaque stereograms as described in the General Methods section. Here I presented transparent and opaque random dot stimuli at a range of disparity ratios. I fixed the standard disparity to 0.15 degrees of visual angle, and used five target disparities of 0.30, 0.45, 0.60, 0.75 & 0.90 degrees of visual angle, giving disparity ratios of 2, 3, 4, 5 & 6. I used a fixed dot density of 0.05. To limit performance, I

presented the signals in a number of noise levels by the method of constant stimuli. I tested five noise levels per condition and measured d' for each noise level I tested. In both the transparent and opaque conditions each observer completed 20 practice trials with 0% noise to become familiar with the stimulus before beginning a session for a new condition. There were equal numbers of near further and far further trials. Each condition was blocked, with 40 trials per each noise condition (20 near further, 20 far further). Within each condition, trials for different noise levels were randomly interleaved.

4.4.1.3   Ideal Observer

The ideal observer for this task was identical to that described in Experiment 4 in detail. The quantity of matches of a given disparity is given by the cross-correlation of the left and right images. This is then compared with templates, by correlation. The templates used by the ideal observer described the two possible disparity combinations (the location of the peaks in the templates) for a given condition of disparity ratio. The ideal observer then selects the template with the highest correlation, a maximum likelihood decision rule.

4.4.2 Results

Figure 4.10A plots the ideal signal thresholds as a function of disparity ratio for both the transparent (open symbols) and opaque (filled symbols) conditions. There are two features to these data. The first is that the ideal signal thresholds are consistently higher in the transparent condition than in
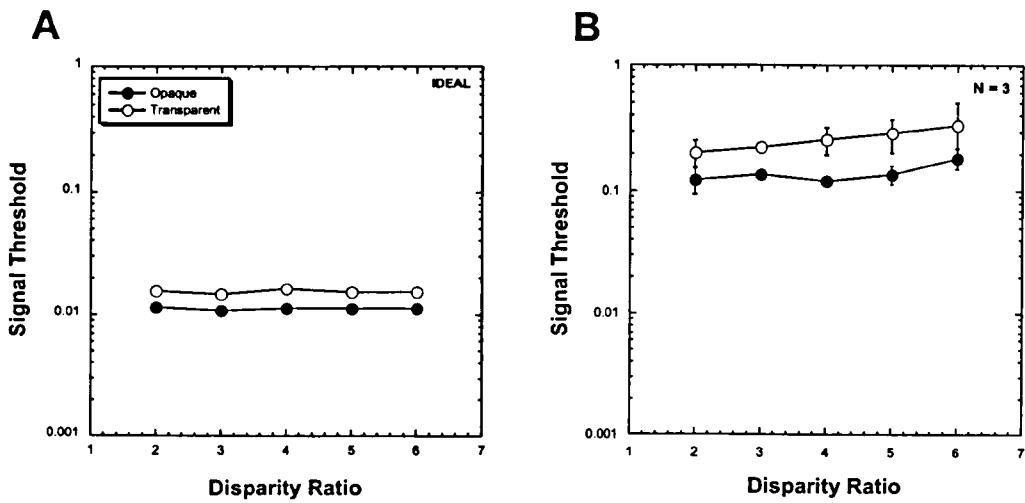
Figure 4.10. (A) Opaque (filled circles) and transparent (open circles) signal thresholds for the ideal observer as a function of disparity ratio. (B) Average opaque (filled circles) and transparent (open circles) signal thresholds for the human observers as a function of disparity ratio. Error bars indicate standard errors across observers.

the opaque condition, across the range of disparity ratios. The second feature to this data is that ideal performance is constant across the disparity ratios. Indeed, there is no reason to expect an effect of increasing the difference in disparity between the standard and target surfaces. This simply changes the location of the peaks of the disparity correlations. The only limitation on ideal performance is the disparity noise.

Figure 4.10B plots the average signal thresholds for three observers as a function of disparity ratio for both the transparent (open circles) and opaque (filled circles) conditions. As in Experiment 4, error bars are standard errors

115
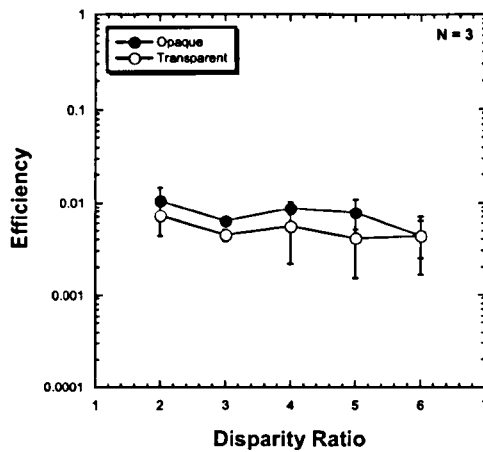
Figure 4.11. Average opaque (filled circles) and transparent (open circles) efficiencies as a function of disparity ratio. Error bars indicate the standard error across the three observers.

of the mean across observers. Again, there are two features to these data. The first is that transparent thresholds are consistently higher than opaque thresholds. The second feature is that there is little effect of disparity ratio, thresholds are more or less constant across the range of disparities tested. Thresholds for the individual observers are plotted in Figure 4.12A – C.

Figure 4.11 plots the average efficiencies for the three observers as a function of disparity ratio. Error bars are standard errors of the mean across observers. The efficiencies are similar for the opaque and transparent conditions across the range of disparity ratios. The cost in performance (higher signal thresholds) for depth discrimination of transparent surfaces does not translate into a lower efficiency. Efficiencies are constant across the disparity ratios for both the opaque and transparent conditions, and similar in amplitude across
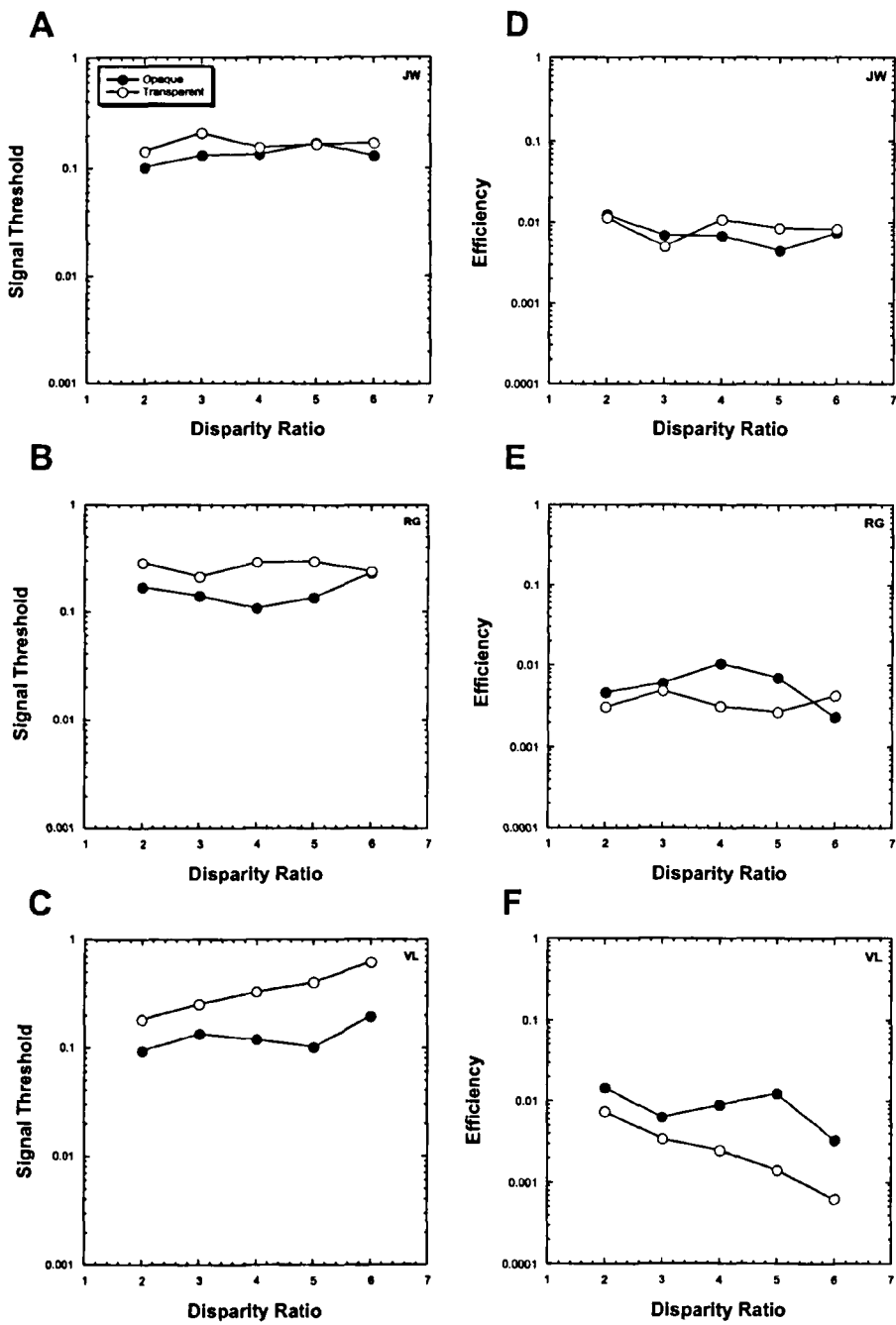
Figure 4.12. (A) – (C) Opaque (filled circles) and transparent (open circles) signal thresholds as a function of disparity ratio for each of the human observers. (D) – (F) Opaque (filled circles) and transparent (open circles) efficiencies as a function of disparity ratio for each of the human observers.

117

conditions. Efficiencies for the individual observers are plotted in Figure 4.12D – F.

4.4.3 Discussion

In this experiment I compared human performance for depth discrimination of transparent and opaque surfaces as a function of disparity ratio. This addressed the question of whether discrimination performance is impaired for transparent stereograms across a range of disparity ratios, and whether the disparity ratio has an effect on depth discrimination. I found that signal thresholds are consistently higher in the transparent condition than the opaque condition, for both the human observers and the ideal observer, across a three-fold range of disparity ratios. I also found that efficiencies were approximately equal in the two conditions across the range of disparity ratios. This confirms the finding of Experiment 4, false matching accounts for the cost in the transparent condition. However, I found that there is no effect of disparity ratio on depth discrimination of transparent or single opaque surfaces. This is in contrast to the findings of Akerstrom & Todd (1988) who found that increasing the disparity difference between transparent surfaces impaired perceived transparency. The results are also in contrast to the findings of Gepshtein & Cooperman (1998) who found that the limiting density to discriminate an oriented cylinder behind a transparent plane decreased as the depth between the surfaces was increased, which they termed the 'farther worse' effect. The difference between the present study and the Akerstrom & Todd (1988) study may be due to the disparities used.

Here I fixed a standard disparity at ±9 arcmin and increased a target disparity in steps up to ±54 arcmin. In contrast, Akerstrom and Todd (1988) used a minimum difference of ±7 arcmin and ±21 arcmin up to a maximum of ±49 arcmin and ±63 arcmin (although the exact disparities varied across observers). Therefore the largest absolute disparity in their study was 112 arcmin, while here it is 63 arcmin. It is possible that the effect of disparity on stereo-transparency found by Akerstrom and Todd (1988) is due to a problem in fusing the two-planes simultaneously. Indeed, it was noted by Akerstrom & Todd (1988) that their observers found they had to make a considerable effort to see the two surfaces in their stereograms, even over long presentation times (up to 35s), suggesting the need for vergence eye movements. In contrast, here observers were instructed to fixate on a zero disparity cross and could perceive transparency at a relatively short duration.

Effects of disparity on surface perception have been attributed to inhibitory interactions at the level of surface representations. This was suggested by the Gepshtein & Cooperman (1998) study, in which the 'farther worse' effect persisted when the two surfaces were defined by opposite polarities, although the overall magnitude of the effect was less than the same polarity condition. This parallels Akerstrom & Todd's (1988) finding that perceived transparency was impaired by increasing the disparity difference between chromatically defined surfaces, but to a lesser extent than a single colour condition. There is evidence for inhibitory interactions in disparity tuning, though not specifically at the level of a surface representation. Specifically, Stevenson et

al. found (1992) that adapting to a particular disparity resulted in a threshold elevation in the disparity sensitivity function, and Cormack et al. (1993) found that correlation thresholds for a given disparity were raised by the presence of a different disparity. The disparity tuning functions derived from these studies were very similar (see Cormack et al., 1993), with clear inhibitory regions. Stevenson et al. (1992) demonstrated that their tuning functions could be modelled by a number of narrowly tuned disparity channels with inhibitory lobes (a centre-surround receptive field), but did not rule out a mutual inhibition between disparity tuned channels. The lack of an effect of disparity in the present study suggests that the range of disparities I used were beyond the range of any inhibitory interactions, and so favours an account of disparity domain inhibition in terms of narrowly tuned disparity channels with inhibitory lobes, rather than a mutual inhibition between disparity channels, or disparity defined surfaces.

4.5   Experiment 6

Observers reported that the perception of a surface was absent at the lowest densities used here, but was more likely to occur as the dot density was further increased. I ran an experiment in which observers were required to indicate whether they did perceive two surfaces, a similar task to that of Akerstrom & Todd (1988).

4.5.1 Methods

4.5.1.1   Stimuli & Procedure

I presented transparent and opaque stereograms as described in the General

Methods section. Two of the observers were asked to indicate whether they

perceived two surfaces, 'yes' or 'no'. The standard disparity was fixed at 0.15

degrees of visual angle and the target disparity was 0.30 degrees of visual

angle, giving a *disparity ratio* of 2 (0.30/0.15). I added a proportion of noise

dots by the adaptive QUEST procedure (Watson & Pelli, 1983) to find the

threshold (75% correct) level of signal dots.

4.5.2 Results

Figures 4.13a & 4.13b plot the surface thresholds for two observers (RG & JW)

for the opaque and transparent conditions. There is no consistent pattern

across observers. For RG the transparent thresholds are consistently higher

than the opaque thresholds, and there is little effect of dot density. In

contrast, observer JW is at a ceiling for surface perception at the lower

densities in both conditions, but at higher densities the thresholds drop and

are generally higher in the transparent condition. Generally, the thresholds

for surface perception are very high and within the same range across

observers.
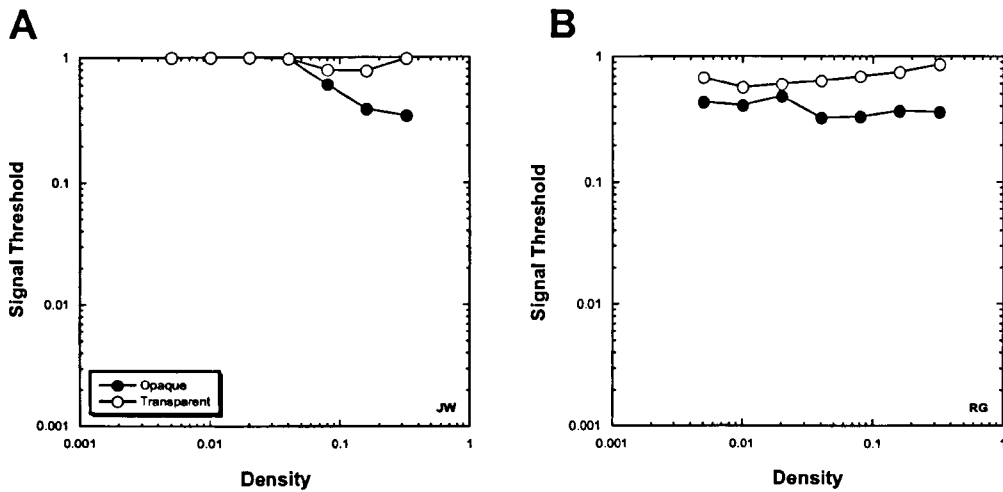
Figure 4.13. *Opaque (filled circles) and transparent (open circles) signal thresholds for surface perception as a function of dot density for two observers.*

4.5.3 Discussion

There is a trend for surface perception thresholds in the transparent condition to be higher than the opaque condition. This pattern of results is not entirely consistent with those of Experiment 4, the surface thresholds are generally higher and generally the pattern of thresholds does not parallel the depth discrimination thresholds. Similar to the previous chapter, the inconsistency between the subjective results and the depth discrimination results argues against the use of subjective measures to probe transparency. The subjective results will be contaminated by a criterion for surface perception, whereas the indirect depth discrimination task is a criterion free method to probe the mechanisms underlying stereoscopic transparency.

4.6   General Discussion

4.6.1 Summary of Results

In this study I have computed the efficiency for depth discrimination of transparent and similar opaque random dot stereograms. The advantage of the approach was twofold. The objective method not only gives a more reliable estimate of perceptual performance free of subjective criteria, but the efficiency measure allows the experimenter to normalize that performance to the information available in the stimulus. An efficiency experiment thereby allows us to compare performance across observers (because it is objective) and across task (because performance is normalized to absolute performance). In Experiment 4 I found that the efficiencies were approximately equal for the transparent and opaque conditions. This demonstrated that the higher thresholds in transparency are accounted for by a greater incidence of false matching in that condition. This suggests that the findings of Akerstrom & Todd (1988), who found thresholds for the perception of transparency to be higher than for similar opaque surfaces, may also be due to a higher rate of false matching in that condition and do not necessarily imply inhibitory interactions. The very low efficiencies I found, of around 1% or less, in both the opaque and transparent conditions suggest there is a problem in using all the available signal information. In support of this I also found that increasing the dot density, thus increasing the number of possible correspondences, decreased the efficiency in both conditions. In Experiment 5, I found that the efficiencies in the opaque and transparent conditions were

approximately equal across a range of disparity differences, supporting the finding of Experiment 4. In addition, I found that there was no effect of disparity ratio on stereoscopic transparency. This contrasts with other studies that have found an effect of disparity on transparency, attributed to mutual inhibition between disparity detectors or disparity defined surface representations (Akerstrom & Todd, 1988; Gepshtein & Cooperman, 1998). The present findings call into question these inhibitory accounts. I consider further implications of these results in the following sections.

4.6.2 *Comparison with Motion Efficiencies*

These findings are comparable to those of motion efficiency (Chapter Three). The stimuli and task used in the motion study are comparable to those used here. The ideal observer for the speed discrimination task is identical to the ideal observer for the depth discrimination task, cross-correlating subsequent frames of the motion stimulus and performing a maximum likelihood decision rule by template matching. The maximum efficiencies for the motion study were considerably higher, around 10% compared to 1% here. This suggests that motion mechanisms maintain an improved signal-to-noise ratio compared to stereo mechanisms. The motion stimulus was inherently dynamic, consisting of 10 frames of multiple motion steps, compared to the single presentation of the disparities here and so there were more available motion samples in the motion stimulus. The improved efficiency indicates that the motion system was indeed able to take advantage of this additional

information. In contrast to the present stereo results, I found there was a residual cost for processing transparent motion (indicated by higher efficiencies in that condition than the coherent condition). This residual cost was present across the range of dot densities and speed ratios I tested. The similarities and the differences between the motion and stereo results are further considered in Chapter Six.

4.6.3 *Correspondence Noise Limitations*

The low efficiencies I find suggest that human observers are unable to use all of the available disparity information to perform depth judgments. This supports previous findings of low efficiencies for other stereo tasks (Harris & Parker, 1992; Cormack et al, 1994; 1997). Both Harris & Parker (1992) and Cormack et al. (1997) found efficiencies of 20% or less and, as I found here, their efficiencies declined as the dot density of their stereograms was increased. Increasing the dot density increased the level of false matches in the stimulus (see Figure 4.5), thus creating a greater correspondence problem. Therefore, these results suggest that the mechanisms of stereopsis are limited by correspondence noise i.e. the greater the correspondence problem the less effective the system is at solving it. How do disparity selective mechanisms combat correspondence noise, if not by a mechanism of mutual inhibition? One possibility is by spatial pooling. In the same way that MT spatial pooling can serve to combat the motion correspondence problem (Barlow & Tripathy, 1997), the pooling of disparity information over a large area may also serve to

combat the stereo correspondence problem. Furthermore, such a pooling operation may account for the low efficiencies, as it would effectively reduce the quantity of disparity samples used to perform the task. Indeed, the similarity of the effects of correspondence noise here and in Chapter Three suggest that, while there is a difference in the effects of transparency, a similar mechanism underlies performance in both domains. I further consider this comparison in Chapter Six.

4.6.4 Visual Mechanisms underlying Stereoscopic Transparency

Transparency has provided a crucial 'test case' for models of stereoscopic vision. As described in the introduction, stereo algorithms that employ the uniqueness and continuity constraints of Marr & Poggio (1976, 1979) will be unable to recover stereoscopic transparency, as they do not permit the occurrence of more than one disparity at a given visual location. More recently a range of computational models have been proposed that pass the test of transparency to varying degrees of success (Prazdny, 1985; Pollard et al., 1985; Gray et al., 1998; Read, 2002; Tsai & Victor, 2003). The later of these models incorporate physiological constraints of the underlying mechanisms (DeAngelis et al., 1991; Freeman & Ohzawa, 1990; Ohzawa et al., 1990, 1996, 1997; Anzai et al., 1999a, 1999b, 1999c), understood to compute a 'disparity-energy' (Qian, 1994; Ohzawa, 1998). By controlling for the available information the present findings argue against inhibitory interactions between disparities over large regions in the same visual direction (although

the possibility remains that inhibitory interactions occur at a finer level), and are thus consistent with these models. The present results do not rule out inhibitory interactions between adjacent regions of space (across different visual directions). This form of inhibition could be useful to accentuate surface boundaries, and indeed this kind of inhibitory interaction has been identified in the center-surround disparity tuning of MT cells (Bradley & Andersen, 1998).

4.6.5 Conclusions

This study has confirmed that stereo mechanisms do not use all the available disparity information, and are significantly impaired by the correspondence problem. This may be the result of a spatial pooling of disparity information. Furthermore, I extended these findings and have shown that the same limitations underlie the recovery of transparency from disparity. There was no cost in efficiency for stereoscopic transparency. This result contrasts with the results of Chapter 3, where there was a cost in efficiency for transparent motion, and suggests an important difference in the way information is processed by motion and stereo mechanisms.

Chapter Five: The Efficiency of Smooth Pursuit

5.1   Introduction

In this chapter I extend the efficiency approach from perception to visually guided action.   The action system I examine is the oculomotor system, specifically I assess performance for the smooth pursuit of transparent and corrugated random dot patterns.   There are two main aspects to this analysis. The first is to make a direct comparison between motor and psychophysical performance for similar stimuli.   The second is to use the intrinsically dynamic pursuit response to probe the temporal dynamics of the underlying visual mechanisms.   In the following introduction I provide a brief review of the oculomotor research related to motion integration and segmentation, the attempts made to relate oculomotor and psychophysical performance, and the use of the oculomotor response to assess the temporal dynamics of the underlying mechanisms.

5.1.1 Previous Research

In the natural environment, smooth pursuit is a conjugate eye movement that tracks a (slow-moving) object of interest, to maintain the object on the fovea. Usually this is punctuated by saccades, fast ballistic eye movements that facilitate accurate tracking.   Rashbass (1961) studied this behaviour experimentally with 'step-ramp' stimuli, in which a spot was presented before
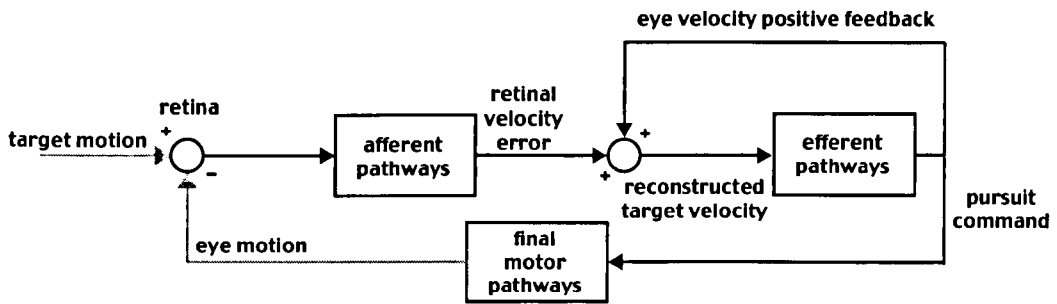
Figure 5.1. The pursuit system can be modelled as a control loop. The negative feedback loop aims to minimize the retinal error, the difference between target motion and eye motion. The positive feedback loop sends a copy of the motor command to reconstruct the target velocity, providing a stable signal to the motor pathways. This 'black-box' model captures some important aspects of pursuit (this figure is adapted from Lisberger et al., 1987).

changing position and then set into motion in the opposite direction to the change in position. It was found that after an initial latency (around 100ms) the eyes first began tracking in the direction of the target motion, then at a further delay made a saccadic movement towards the target (in the direction of the positional change) and continued to track in the direction of the target motion. This behaviour identified image motion as the stimulus for smooth pursuit, distinguished from saccadic eye-movements that respond to stimulus position.

Pursuit has been modelled as a negative feedback control loop, in which the motor system aims to reduce the retinal error, the difference between retinal motion and eye motion (Figure 5.1). This negative-feedback control loop also requires positive feedback of eye velocity, to provide a measurement of target velocity and prevent the eyes from stopping when the error signal reaches

zero (see Lisberger et al., 1987). This characterization of the system has been criticized (Steinman, 1986) and more recently there has been an interest in developing a more physiologically plausible description of the mechanism and relating this to the visual mechanisms that underlie visual perception (e.g. Krauzlis & Stone, 1999). Nevertheless, the control system description of pursuit captures some important aspects of the behaviour. Mainly, pursuit behaviour will reflect the negative-feedback loop, the interplay between visual input and motor output. However, if the loop could be 'opened' it would be possible to use the pursuit response as a probe of the underlying visual mechanisms. One method to do this is to take advantage of the delay between the onset of target motion and the initiation in pursuit: this dictates that the first 100ms or so (depending on the stimulus parameters) of the pursuit response will reflect purely the visual inputs (Lisberger & Westbrook, 1985; Lisberger et al., 1987). The pursuit system can then be used to probe the mechanisms that underlie visual motion processing. The use of eye movements to probe the underlying mechanisms is a well-established tradition, saccades are used as a probe for cognitive processing in reading (Rayner, 1977, 1998) and attention (Posner, 1980; Rizzolatti et al., 1997).

As described in Chapter Three, the human visual motion system must integrate similar motion signals and segment different motion signals to recover an accurate representation of the three-dimensional visual scene (see Braddick, 1993). Previous oculomotor research has tended to use small single-spot motion targets (e.g. Lisberger & Westbrook, 1985). These simple

local motion stimuli by their nature do not demand the integration or segmentation of local motions, and therefore do not probe the systems response to complex stimuli of the kind that occur in the natural environment e.g. tracking a butterfly in flight against the changing, textured background of foliage. Larger field motion displays have been used, generally to elicit passive eye movements, optokinetic nystagmus (OKN) (Cohen et al., 1977) or 'ocular following' responses (see Miles, 1998). These automatic 'reflex' responses reduce retinal image motion in response to movement of the entire visual field and are therefore behaviourally quite different to smooth pursuit, an intentional eye movement that tracks a visual object of interest moving against the background of the environment (Kowler et al., 1984; Collejiwn & Tamminga, 1984), although it is understood that pursuit and OKN rely on common neural mechanisms (Pola & Wyatt, 1985; Kawano, 1999). Recently, it has been demonstrated that voluntary pursuit eye movements are sensitive to large stimuli, in particular pursuit accuracy is improved, acceleration is increased and latency is decreased as the size of a global motion stimulus is increased (Heinen & Watamaniuk, 1998). This demonstrates that active, smooth pursuit does depend on the output from integrative motion mechanisms. Area MT has been identified as the neural site for these integrative motion mechanisms (Newsome & Pare, 1988; Stoner & Albright, 1992), and provides the visual information used by smooth pursuit eye movements (Komatsu & Wurtz, 1988, 1989; Groh et al., 1997; Born et al., 2000).

A few studies have examined the oculomotor response, both active and passive, to more complex global motion stimuli, particularly motion transparency. Motion transparency occurs when two global motions are perceived in the same visual direction at different depths (e.g. Adelson & Movshon, 1982; Snowden, 1989; Stoner et al., 1990; Murakami, 1997). This is a particularly useful stimulus to study the mechanisms of motion integration and segmentation, as it involves both simultaneously. It has been shown that the velocity of pursuit is reduced to transparent stimuli of opposite global motions, compared to single global motion stimuli (Niemann, Ilg & Hoffman, 1994). This oculomotor behaviour parallels the physiological evidence for suppressed responses in MT to transparent stimuli (Snowden et al., 1991). There have also been attempts recently to relate oculomotor responses to transparent stimuli to perceptual behaviour. Transparency is a depth percept, although when it is elicited by motion the depth ordering is ambiguous. Watanabe (1999) demonstrated perceptual reversals of the direction of motion for transparent stimuli of opposite directions even when observers were instructed to attend to the motion at a particular perceived depth. However, the direction of the slow-phase of the OKN elicited by these stimuli was correlated to the perceptual reports, demonstrating a coupling between the perceptual system underlying motion transparency and the oculomotor response. Mestre & Masson (1997) demonstrated that the velocity of OKN in response to same-direction transparent stimuli followed the average of the motion vectors in the stimulus, and (for stimuli with three velocities or less)

after observers reported perceiving the transparent stimuli the eye velocity follows the slowest velocity in the stimulus.

Recent studies have attempted to directly compare perceptual and pursuit performance for the same tasks. Beutter & Stone (1998) computed 'oculometric' functions, the equivalent of psychometric functions for the oculomotor system, for the direction discrimination of plaid stimuli. This approach permitted a comparison between psychophysical thresholds and oculomotor thresholds. They found that the directional biases in pursuit were very similar (see also Stone, Beutter & Lorenceau, 2000). Watamaniuk & Heinen (1999) applied the oculometric analysis to the direction discrimination of random dot motions. In Watamaniuk & Heinen's analysis thresholds were computed as a function of external noise to estimate the internal noise of the oculomotor (specifically smooth pursuit system) and perceptual mechanisms, following Pelli (1990). They found that while psychophysical performance was better than oculomotor direction discrimination, the estimated internal noise was approximately equal. It was argued that the oculomotor system was limited by the same noise as the perceptual mechanisms, i.e. both systems rely on the same underlying visual mechanisms. While identifying a common noise source for perception and action, this novel approach to comparing oculomotor and psychophysical performance does not isolate the visual information driving perceptual and oculomotor performance. The efficiency approach permits this crucial comparison to be made. Notably, Eckstein et al. (2001) demonstrated that the efficiency approach can be extended to the

motor domain, and assessed the visual information driving saccadic behaviour in visual search. Here I further extend the approach to the analysis of the visual information driving smooth pursuit for random dot motions.

5.1.2 Present Study

In the following experiment I compute the efficiency of smooth pursuit for high-density 'corrugated' and 'transparent' random dot motions. These stimuli are illustrated in Figure 5.2. This type of stimulus was originally used by van Doorn & Koenderink (1982a, 1982b). They demonstrated that varying the bar width of the configuration altered the perception of the stimulus, at large bar widths the individual strips of motion are perceived (as in the corrugated condition), but as the bar width is decreased observers perceive two transparent surfaces, one sliding over the other in depth (as in the transparent condition of Chapter Three). The terminology for the stimuli in the present experiment therefore refers to the different percepts elicited by similar stimuli with different bar-widths (and numbers of bars). Recently, Mestre et al. (2001) determined that this transition occurs at a bar-width of around 0.4 degrees of visual angle. Here, I present corrugated stimuli of non-overlapping motions, and transparent stimuli of (perceptually) overlapping motions to assess the sensitivity of smooth pursuit to mechanisms of motion integration and motion segmentation. There are two main purposes for this study. First, the efficiency computation will permit a direct comparison with the perceptual performance of Chapter Three. Second, it has previously been

demonstrated that the analysis of smooth pursuit eye movements in response to motion stimuli provides a technique that allows us to probe the temporal evolution of motion processing (Masson & Mestre, 1998; Born et al., 2002). Such studies have shown for example that in response to 'barber pole' stimuli the motion system initially provides one-dimensional motion estimates (corresponding to the motion direction orthogonal to the orientation of the line segments) and over time computes the two-dimensional motion of the terminators (Masson et al., 2000), or that tracking of 'line-diamond' stimuli is initiated in the direction of the vector average of the moving segments and over time corrects to the true object motion (Masson & Stone, 2002). Here, I use the pursuit response to examine the temporal dynamics underlying motion integration and segmentation in the corrugated and transparent stimuli.

5.2   Experiment 7

5.2.1 Methods

5.2.1.1   Human Observers

Two observers participated in the experiment. One observer (AK) was an experienced participant in eye-tracking experiments. The other (JW) was an experienced psychophysical observer but a novice to eye-tracking experiments. Both observers had normal or corrected-to-normal visual acuity.

5.2.1.2   Apparatus

Stimuli were back-projected on a large translucent screen, using a RGB video projector. Stimulus presentation was controlled by a PC with the REX software package (Hays, Richmond & Optican, 1982). Observers were seated and head position was stabilised by a chin support and forehead rest.
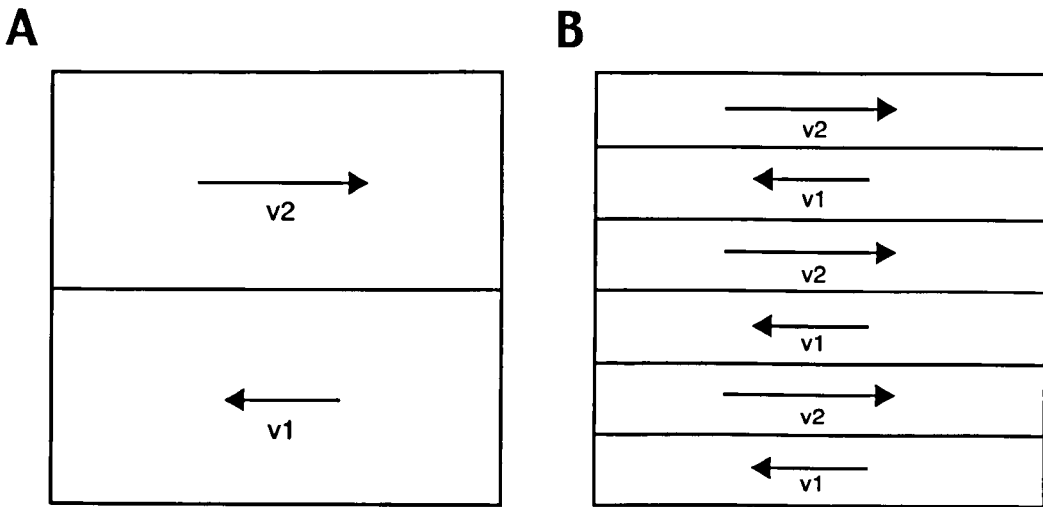
Figure 5.2. Schematic illustrations of (A) corrugated and (B) transparent stimuli.

5.2.1.3   Stimuli

The visual stimuli were 35 frame movies generated by the HIPS software (Landy et al., 1984) running on a Silicon Graphics workstation. Two types of random dot displays were used, a corrugated motion stimulus and a transparent motion stimulus. Both stimuli were constructed from a 256 by 256 pixel array, and subtended 12 by 12 degrees of visual angle when back-projected onto a large translucent screen (subtending 80 by 80 degrees of visual angle, viewed at a distance of 1m). The stimuli were back-projected by a video projector at 75 Hz. Each location in the stimulus pixel array was black or white with a probability of 0.50, corresponding to a dot density of 50%.

In the corrugated motion stimulus (Figure 5.2A), dots in the upper half of the stimulus (the first 128 rows) were displaced in one direction and dots in the lower half (the next 128 rows) of the stimulus were displaced in the opposite

direction (each half of the corrugated stimulus subtended a visual angle of 6 by 12 degrees). This stimulus is perceived as two opaque surfaces moving in opposite directions. The transparent motion stimulus (Figure 5.2B) was then constructed by alternating the direction of motion every 4 pixels (0.2 degrees). Although this stimulus consisted of a number of horizontal 'strips' of alternating directions, this stimulus is perceived as two surfaces moving in opposite directions, one sliding transparently over the other, as first described by van Doorn & Koenderink (1982a, 1982b). Note that in both of these stimuli dots moving in different direction are spatially segregated. In both stimuli dots moving in the same direction must be integrated and dots moving in the opposite direction must be segmented from each other. The key difference is that while the corrugated bars remain perceptually segregated, the transparent bars of the same direction are integrated together and are perceived as a surface extending across the entire display.

Each random dot display subtended 12 by 12 degrees of visual angle. Performance was limited by varying the amount of flicker noise in the stimulus, this type of noise is different from that used in the previous psychophysical experiments. Flicker noise refers to the probability that a given pixel in the image (which could be black or white) could change its polarity (n). Therefore, the probability that a given pixel remained the same polarity (s) was 1 minus the flicker noise probability ($s = 1 - n$). This flicker noise determines that the density of white (and black) pixels was constant at 50%. With zero probability of noise then the polarity of all dots remained

constant, and the direction of motion was clearly perceived. However, with increasing amounts of flicker noise the direction of motion was increasingly more difficult to perceive.

5.2.1.4   Procedure

In a given block of trials observers were presented with one type of motion, either the corrugated stimuli or the transparent stimuli described above. For the corrugated motion type, each trial consisted of two opposite directions of horizontal motion (left and right), the upper half of the stimulus moved in one direction and the lower half of the stimulus moved in the opposite direction. The direction that dots were displaced in the upper and lower areas was randomised across trials. For the transparent motion type, again each trial consisted of two opposite directions of motion, but now the direction of motion alternated between 'strips' of the stimulus. The direction in which dots were displaced in the 'odd' and 'even' strips was randomised across trials. For each motion type, one of the motions moved at the 'standard' speed of 4 degrees per second. The other 'target' speed was $8°$ s^{-1}, $12°$ s^{-1} or $16°$ s^{-1}, corresponding to speed ratios of 2, 3 & 4. The direction of the target (and therefore the standard) motion was randomised across trials. For each motion type and for each speed ratio, performance was limited by varying the amount of flicker noise present in the stimulus, 10 flicker noise probabilities were used ranging from 0.01 to 0.99. For each motion type, the speed ratio and noise level were randomised across trials. Observers ran 35 trials at each

noise level, for each speed ratio and condition. For both motion types, the stimulus duration for each trial was 450ms. The observer's task was to track the faster surface motion. Data was collected from each observer in sessions of one hour duration, up to 20 hours of data collection in total for each observer.

5.2.1.5   *Eye Movement Recording and Data Analysis*

Eye movements were recorded by the scleral search coil method (Collewijn, van der Mark & Jansen, 1975) supplied by Skalar Med (spatial resolution: 1 min of arc, temporal resolution: 1kHz). A coil was placed in the right eye only, following the application of a topical anaesthetic to facilitate the placing and wearing of the coil. On each session of data collection, after the coil was placed, the responses were calibrated by a fixation task (to map the display locations to the coil response). On each trial eye movements were recorded from 100ms before the stimulus onset up to the end of the stimulus presentation (450ms after stimulus onset) by the REX software package (Hays, Richmond & Optican, 1982). Responses were analysed using the IDEA software environment*, which provides a visual representation of the pursuit response. Saccades were identified by visual inspection and responses containing saccades were rejected from analysis. For each trial, a baseline response was computed as the average horizontal velocity from –30ms (30ms prior to stimulus onset) to 60ms after stimulus onset. Pursuit latency was

---

* Eye-movement data analysis software environment developed and supplied by Richard Krauzlis, Salk Institute.

determined by the linear regression technique (Carl & Gellman, 1987), and all responses were re-aligned with respect to the onset of the pursuit response. Trials were rejected if the latency was greater than 210ms. The horizontal velocity of the pursuit responses were then averaged within four bins or 'time windows', 0 – 60ms, 60 – 120ms, 120 – 180ms, and180 – 240ms, where 0ms is the onset of the eye movement. Within each time window, the criterion for deciding the eyes moved was a horizontal velocity greater than 1 standard deviation of the baseline response. To compute oculomotor sensitivity functions, the direction of the pursuit response was determined for each time window by taking the sign of the average velocity for that time window. This analysis therefore reduces the dynamic pursuit response to a simple binary decision, 'left' or 'right'. These binary responses were then used to compute the hit rate and false alarm rates over the repeated trials for each noise level within a condition, and thus the corresponding d'. Note that these hit and false alarm rates, therefore the computed sensitivities, will vary over time only if there is a change in the direction of the eyes over time.

5.2.1.6 Ideal Observer

The ideal observer for a given task makes use of all the relevant information in a given stimulus to perform that task optimally i.e. maximising the number of correct responses by performing a maximum likelihood estimate (e.g. Swets, 1964; Green & Swets, 1966). The ideal observer here is facing the same speed discrimination task as any human observer. Recall that the task posed

to the oculomotor pursuit system is to select the faster surface motion. To select the faster surface motion, the ideal observer needs to represent the speeds displayed in the stimulus, compare these speeds to the possible speed combinations, and thereby select the faster speed. These three stages are the stimulus representation, template matching and the decision stage (Figure 5.3). This is an extension of the ideal observer described in Chapter Three, for the case of multiple speed combinations. The stimulus representation is given by the one-dimensional cross-correlation across successive frames of the stimulus (see also van Doorn & Koenderink, 1982a). The cross-correlation function simply describes the quantity of matches at each speed with no loss of information. As the task is to discriminate only leftward from rightward motions, only horizontal displacements are considered. At low levels of flicker noise, the peaks of this speed correlation correspond to the standard and target signal speeds. This can be seen in Figure 5.3A for a transparent stimulus with a 0.10 probability of flicker noise, in which the target speed is moving to the left (here the correlation is computed across 20 frames). The ideal algorithm computes the likelihood of each possible outcome by comparing the stimulus representation with a number of 'templates'. Each template is a representation of the possible speed combinations, correlations that peak at the possible speeds (Figure 5.3B). As all the speed combinations can be presented within a given block of trials, the stimulus representation is compared with the six possible speed combinations. To compute the likelihood of each possible outcome, the ideal algorithm cross-correlates the stimulus correlation with each template (e.g. Green & Swets, 1966).
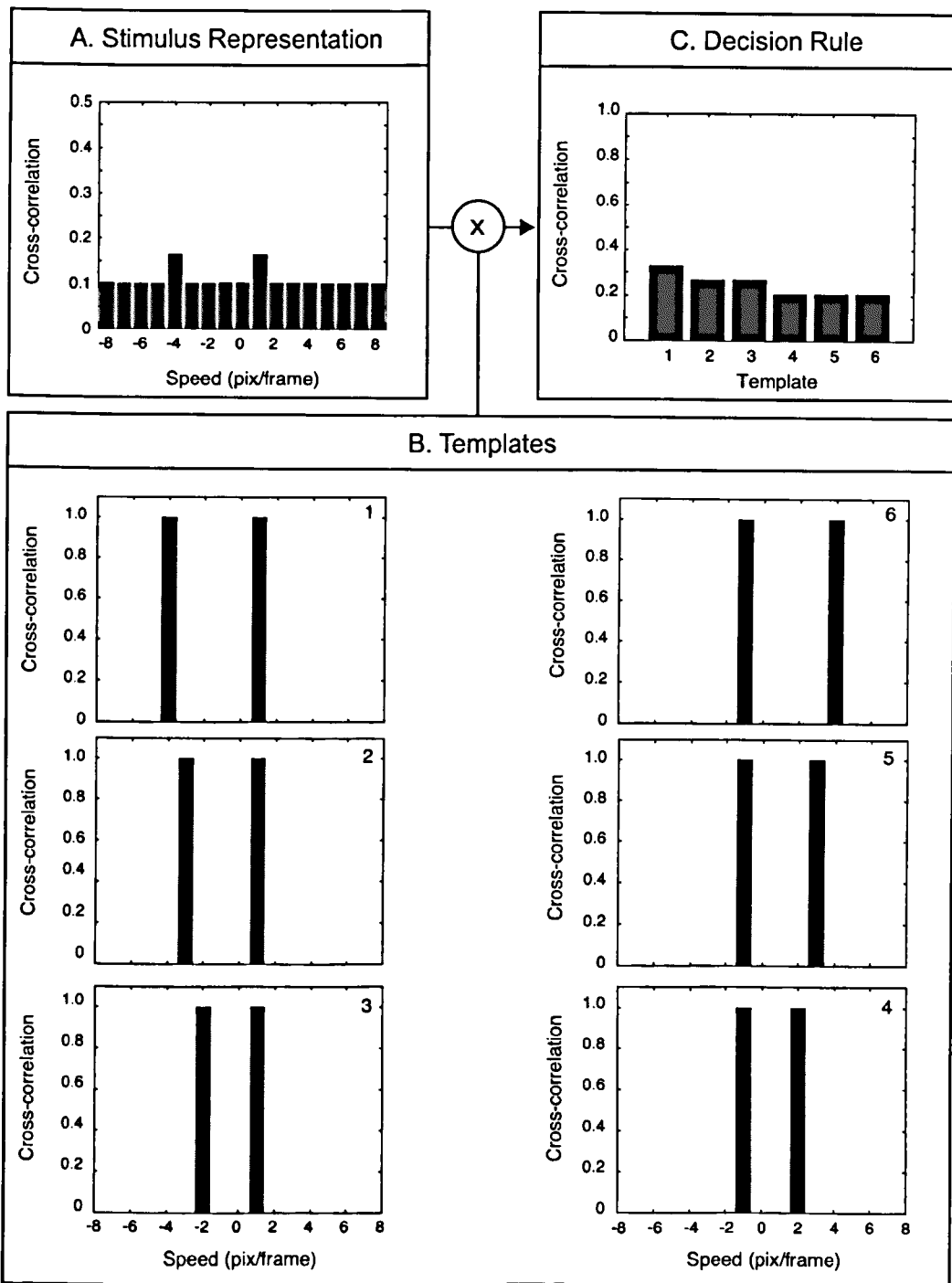
Figure 5.3. *A schematic illustration of the ideal observer algorithm.*

143

The ideal decision rule is then to choose the template that returns the largest cross-correlation value with the stimulus (Figure 5.3C), a maximum likelihood decision rule. In the case of low levels of flicker noise, the template with the highest value will correspond to the actual signal presented, and in Figure 5.3 the ideal observer indeed selects the correct template. However, at much higher levels of flicker noise the value of the incorrect template can be higher than that of the correct template leading the ideal observer to make an incorrect decision. Only these occurrences limit the ideal observer performance. Note that, in contrast to the previous experiments, the displacements in opposite directions in both the conditions are spatially separated. Therefore, as the ideal observer considers only the one-dimensional cross-correlation, ideal performance should be equivalent in the two conditions of the present experiment.

To compute ideal thresholds, simulations of the ideal observer were performed for both the transparent and coherent conditions. Simulations were performed for each of the three speed ratios, at ten noise levels for each speed ratio condition, with 30 trials (15 left faster, 15 right faster) per noise level. To compare ideal performance with pursuit performance across the different time windows, the ideal observer performance was computed for four integration windows, 5 frames, 10 frames, 15 frames and 20 frames. These integration windows are equivalent to the total length of stimulus information that was available to the pursuit response at the end of each time window, therefore the efficiency for each time window will measure the

performance relative to the total information that was available to drive the eye-movement. The ideal observer averages the correlation over the appropriate number of frames for each integration window. As in the previous chapters, efficiency is the ratio of human sensitivity to that of the ideal observer (Tanner & Birdsall, 1958; Barlow, 1978):

$$F = \left(\frac{d'_h}{d'_i}\right)^2 \qquad (1)$$

As we will see in the results section below, d' is a linear function of the proportion of signal dots presented. Therefore:

$$F = \left(\frac{\alpha_h}{\alpha_i}\right)^2 \qquad (2)$$

where α_h is the slope of the linear function of the human d' and α_i is the slope of the linear function of the ideal d'. This definition is equivalent to the ratio of human to ideal thresholds:

$$F = \left(\frac{\theta_i}{\theta_h}\right)^2 \qquad (3)$$

145

5.2.2 Results

An example of the data obtained is shown for a set of simulation of the ideal observer (Figure 5.4A) and of one human observer (Figure 5.4B). These data are for a speed ratio of 4, and were analysed in the final time window, 180 – 240ms (corresponding to 20 frames for the ideal observer). The sensitivities of the human observer and the ideal observer increase approximately linearly as the signal probability increases (therefore as the noise probability decreases), in both the transparent and corrugated conditions. The linear fits were restricted to conditions that led to a d' in the range 0.25 to 2.90. Note the much higher levels of noise required to limit the ideal observer performance. The sensitivities were then fitted with a line constrained to pass through the origin and the signal threshold was defined as the signal proportion (s) required for a d' of 1.
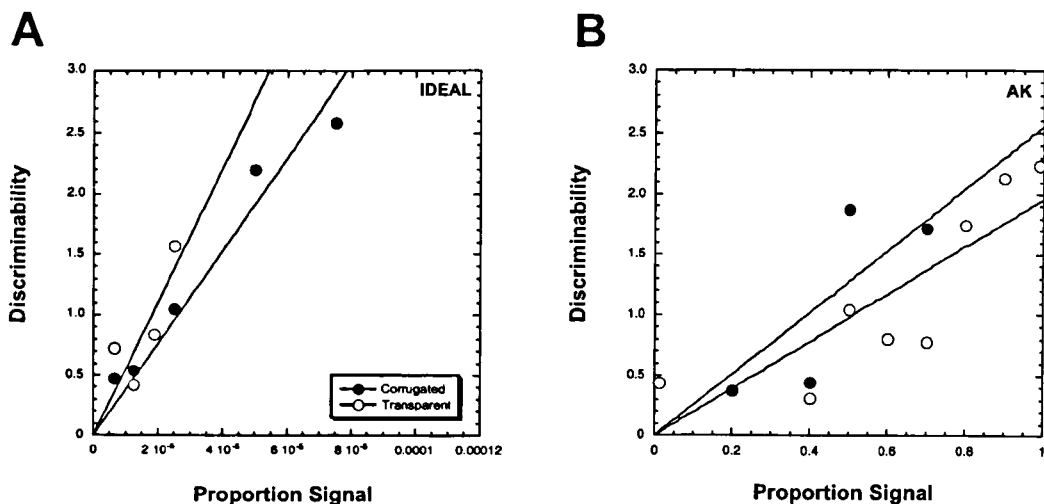
Figure 5.4. *Fitted slopes for the (A) ideal and (B) human observers. Note the fourfold difference in scale for the proportion signal between the ideal and human observer. For the ideal observer, the slope of the fitted line is 3.83 x 10⁴ in the corrugated condition and 5.53 x 10⁴ in the transparent condition. For the human observer, the slope of the fitted line is 2.55 in the corrugated condition and 1.96 in the transparent condition.*

5.2.2.1 Speed & Type of Motion

A 2 x 3 two-way analysis of variance was conducted to assess the effects of speed ratio and type of motion (corrugated versus transparent), collapsed across time window, on pursuit thresholds for each observer. Performance for the ideal observer is shown in Figure 5.5. Results for the ideal observer indicated a main effect for motion type ($F(1,18) = 31.6$, $p < .05$) and speed ratio ($F(2,18) = 11.5$, $p < .05$), but no interaction between condition and speed ratio ($F(2,18) < 1$, $p > .05$). Note that in general, the ideal observer requires only a very small proportion of signal to perform the task in both the transparent
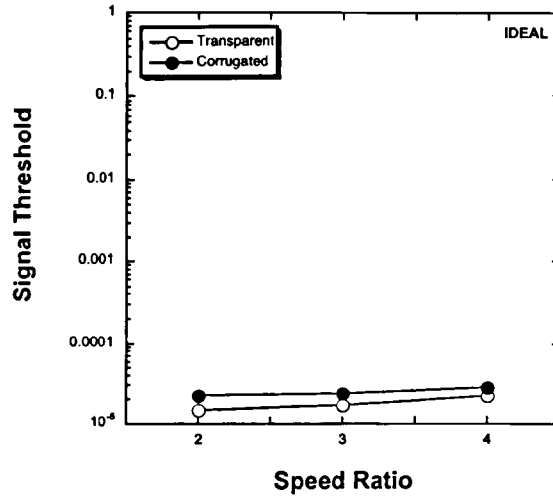
Figure 5.5. Mean transparent (open circles) and corrugated (filled circles) signal thresholds, collapsed across time window, for the ideal observer as a function of speed ratio. Error bars indicate standard error of the mean.

and the corrugated conditions. This is a smaller figure than in the previous motion experiments, however the thresholds are not directly comparable as the for of noise is quite different between these two studies. The small but significant difference between the transparent and corrugated conditions is perplexing as this goes against the theoretical predictions, and suggests that this difference is a result of a stimulus artefact. Similarly, there is no theoretical basis for an effect of speed ratio. This may be a result of the stimulus construction, specifically at higher speeds there is in principle a greater loss of speed information at the edges of the stimulus than for smaller speeds.
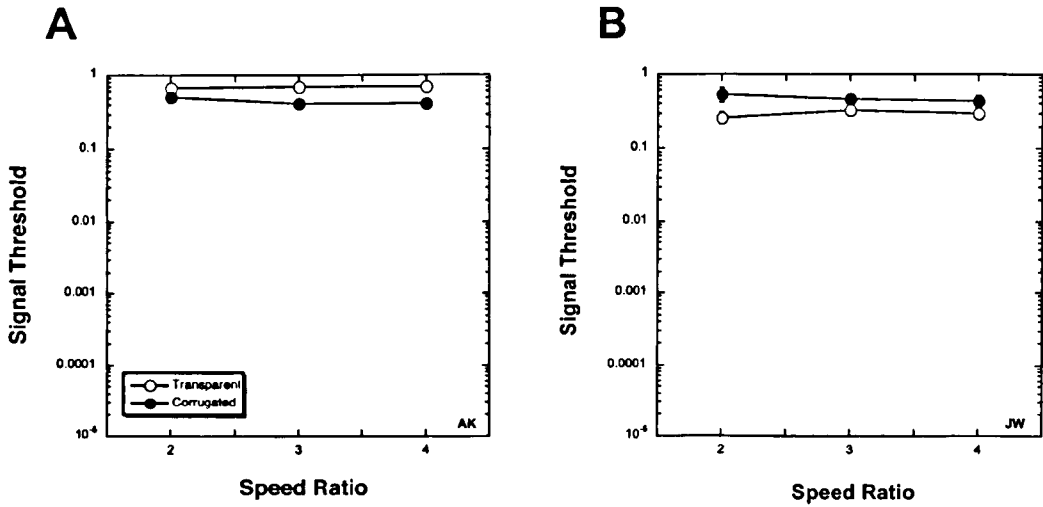
Figure 5.6. *Mean transparent (open circles) and corrugated (filled circles) signal thresholds, collapsed across time window, for the human observers as a function of speed ratio. Error bars indicate standard error of the mean.*

Figure 5.7. *Mean transparent (open circles) and corrugated (filled circles) efficiencies, collapsed across time window, for the two observers as a function of speed ratio. Error bars indicate standard error of the mean.*

149

The performance of observers AK and JW are plotted in Figure 5.6. The results of the analysis for observer AK indicated a main effect of condition, in the opposite direction to the ideal observer ($F(1,14) = 9.86$, $p < .05$), but no main effect of speed ratio ($F(2,14) < 1$, $p > .05$) and no interaction between condition and speed ratio ($F(2,14) < 1$, $p > .05$). The results of the analysis for observer JW indicated a main effect of condition ($F(1,17) = 9.39$, $p < .05$) but not effect of speed ratio ($F(2,17) < 1$, $p > .05$) and no interaction between condition and speed ratio ($F(2,17) < 1$, $p > .05$).

Now I turn to the analysis of the observers' efficiencies, which were computed following Equation (2) of this chapter. The efficiencies of both observers are plotted in Figure 5.7. A 2 x 3 two-way analysis of variance was conducted, collapsed across time window, to assess the effects of speed ratio and condition (corrugated versus transparent) on pursuit efficiencies for each observer. The results of the analysis for observer AK indicated no main effect of condition ($F(1,18) = 7.20$, $p > .05$) or speed ratio ($F(2,18) = 1.42$, $p > .05$) and no interaction between condition and speed ratio ($F(2,18) = 1.17$, $p > 0.05$). Similarly, the results of the analysis for observer JW indicated no main effect of condition ($F(1,18) < 1$, $p > .05$) or speed ratio ($F(2,18) = 1.98$, $p > .05$) and no interaction between condition and speed ratio ($F(2,18) < 1$, $p > 0.05$).

5.2.2.2   Time Window & Type of Motion

To consider the effect of the time window and condition on pursuit thresholds, a 2 x 4 two-way analysis of variance was conducted for each observer, collapsed across speed ratio.   Results for the ideal observer indicated a main effect for motion type, ($F(1,16) = 15.5$, $p < .05$), but no main effect of time window ($F(3,16) < 1$, $p > .05$) and no interaction between condition and speed ratio ($F(3,6) < 1$, $p > .05$).   Note that in general, the ideal observer requires only a very small proportion of signal to perform the task in both the transparent and the corrugated conditions.

Figure 5.8. Mean transparent (open circles) and corrugated (filled circles) signal thresholds, collapsed across speed ratio, for the ideal observer for the four different time windows of analysis.
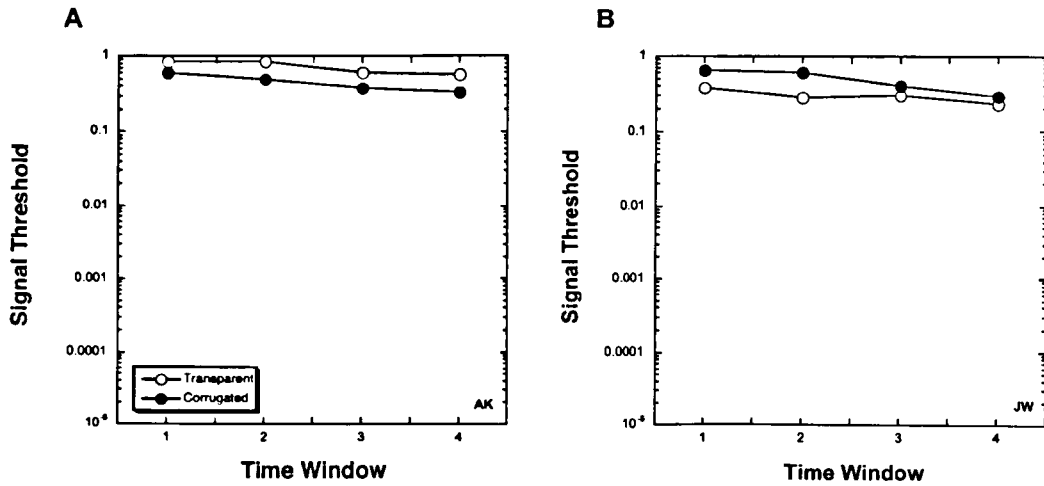
Figure 5.9. Mean transparent (open circles) and corrugated (filled circles) signal thresholds, collapsed across speed ratio, as a function of the time window of analysis for the two human observers. Error bars indicate standard error of the mean.

The human observers thresholds are plotted in Figure 5.9. The results of the analysis for observer AK indicated a main effect of motion type ($F(1,12) = 31.6$, $p < .05$) and time window ($F(3,12) = 8.53$, $p < .05$), thresholds gradually decreasing with time, but no interaction between condition and time window ($F(3,6) < 1$, $p > 0.05$). For observer JW, there was also a main effect of motion type, ($F(1,15) = 31.3$, $p < .05$), and also a main effect of time window ($F(3,15) = 10.5$, $p < .05$) with thresholds decreasing over time, and no effect of an interaction between condition and time window ($F(3,15) = 3.45$, $p > 0.05$). The means for observer JW are plotted in Figure 5.9B.
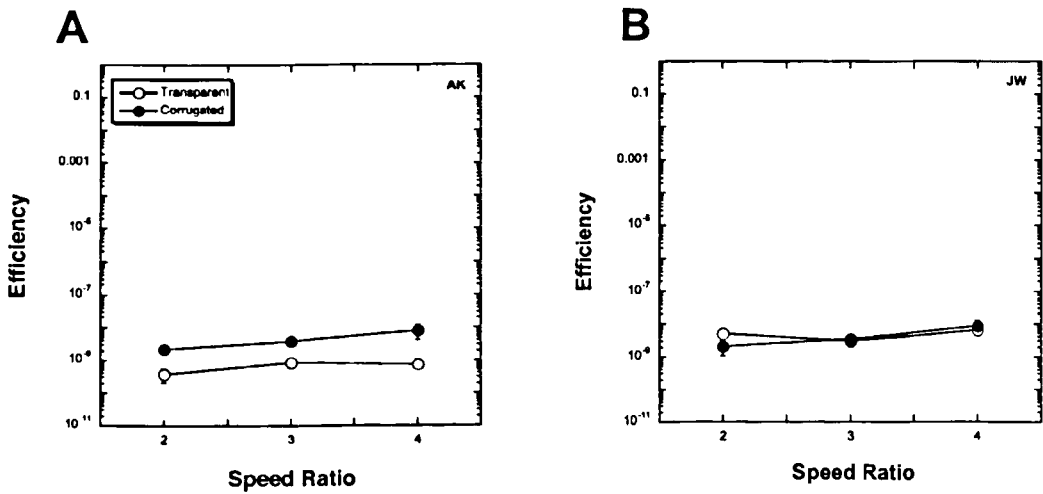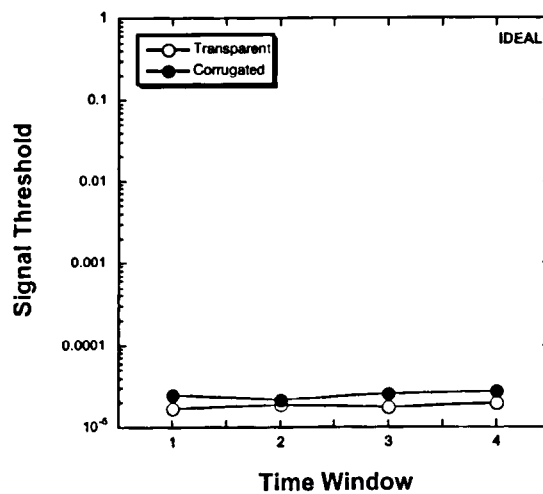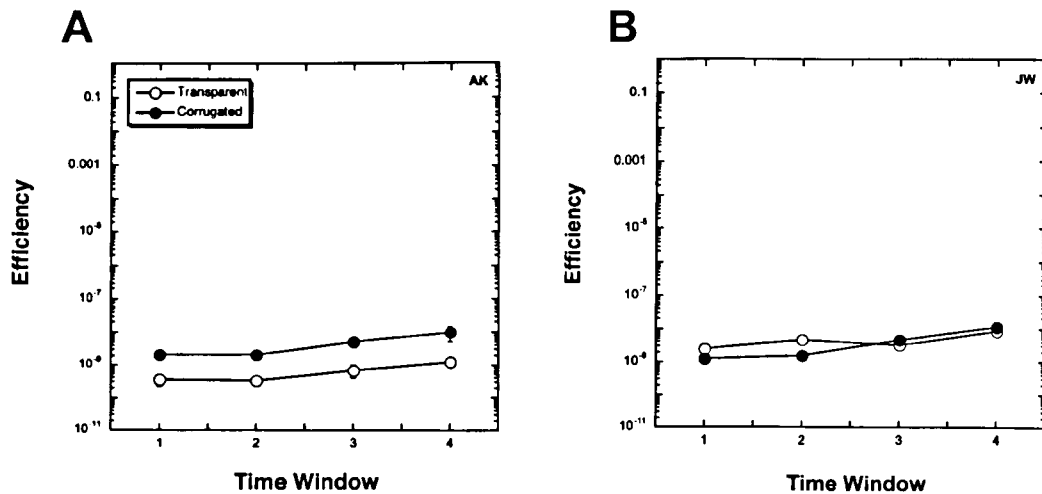
Figure 5.10. Mean transparent (open circles) and corrugated (filled circles) efficiencies, collapsed across speed ratio, as a function of the time window of analysis for the two human observers. Error bars indicate standard error of the mean.

Now I turn to the analysis of the observers' efficiencies, which were computed following Equation (2) of this chapter. These efficiencies are plotted in Figure 5.10. A 2 x 4 two-way analysis of variance was conducted, collapsed across speed ratio, to assess the effects of time window and condition (corrugated versus transparent) on pursuit efficiencies for each observer. The results of the analysis for observer AK indicated a main effect of condition ($F(1,16)$ = 7.99, $p < .05$) but no main effect of time window ($F(3,16) = 2.00$, $p > .05$) or interaction between condition and time window ($F(3,16) = 1.24$, $p > 0.05$). Though not significant, there is a trend for efficiencies to increase with time. The results of the analysis for observer JW indicated no main effect of condition ($F(1,16) < 1$, $p > .05$), but there was a main effect of time window ($F(3,16) = 5.58$, $p < .05$), efficiencies improving over time, and no effect of an interaction between condition and speed ratio ($F(3,16) < 1, p > .05$).

153

5.2.3 Discussion

The general finding is that the efficiency of smooth pursuit is *extremely* small, and is not significantly different between corrugated and transparent stimuli. This overall result is not compatible with evidence of a performance cost for smooth pursuit of transparent motion (Niemann et al., 1994). Furthermore, it is not compatible with the results of Chapter Three and a range of previous psychophysical evidence for a performance cost in transparent motion. However, here the transparent stimuli are constructed from alternating the direction of high-density random-dot strips, whereas the transparent stimuli in the psychophysical tasks of Chapter Three and of Niemann et al. (1994) were superimposed random dot patterns. It is possible that there is not an inhibitory interaction in the strip stimuli of the kind that limits performance in the psychophysical case. Mestre et al. (2001) used stimuli similar to those used here, alternating the speed (but fixing the direction) of horizontal random-dot strips. They found a cost in segmenting motions based on speed cues alone when each dot of the stimulus was 'paired' with a dot of a different speed, placed within 0.2° or less vertically and horizontally (with some random jitter) of each other, and argued that this behaviour reflects an early local pooling of motion signals. However, without this strict pairing constraint there was little cost for stimuli with bar-widths less than 0.2° even though observers perceived transparency. Similarly, in the current experiment the bar-width is 0.2° and transparency is perceived i.e. segmentation is not blocked. Furthermore, the arrangement of opposite directions of motion in horizontal strips precludes the possibility of spatially

localised inhibition between different directions of motion. Therefore, the finding of no difference in performance for transparent and corrugated stimuli is not inconsistent with directional inhibition accounts of transparency.

There are two further aspects of the analysis to consider, the manipulation of speed ratio and time window. There was an effect of time window on smooth pursuit performance, thresholds improved significantly with time (for both observers), and efficiencies tended to improve with time (though this trend was significant for only one observer). This result is further considered in the General Discussion. In contrast, speed ratio had very little effect on observer's performance. It has been shown that there are differences in the effects of speed on the perception of transparent and non-transparent motions. Psychophysical performance for uni-directional motion tasks that require the segmentation of motions is tuned to lower speeds than tasks that do not (Masson et al., 1999). Also, Mestre et al. (2001) demonstrated for unidirectional stimuli, that smaller speed differences were required to support the perception of corrugated displays of the type used here, while larger speed differences were required to support the perception of the transparent displays used here. The lack of an effect of speed for the opposite direction displays here suggests that speed cues are not required for stimuli that can be clearly segmented by directional cues.

5.3 General Discussion

5.3.1 Comparisons with Visual Efficiencies

The efficiencies found in the current study are the smallest efficiencies ever reported (see Table 2.1 for comparison)! Notably, they are much smaller than the visual efficiencies reported in Chapter Three for coherent and transparent motions. It is true that while the stimuli elicit similar percepts, and that the transparent stimuli demand both segmentation and integration of local motions, the construction of the stimuli are somewhat different. In Chapter Three the dots were located randomly anywhere on the screen, and dots of different directions were presented either simultaneously or sequentially. Here, dots of different directions were constrained to fall in different 'strips' of the stimulus. As discussed in the previous section, this difference may account for the lack of a consistent effect of transparency found here. Nevertheless, by computing efficiency performance is normalised to the information content in the stimulus, therefore in principle meaningful comparisons can be made between these two studies. There is a range of evidence that smooth pursuit is driven by the same visual mechanisms that drive perception (e.g. Krauzlis & Stone, 1999). If the pursuit system does have access to the same information as perception, the smaller efficiencies for smooth pursuit suggest that pursuit does not use all of the available information. It is possible that pursuit is driven by only a subset of the available temporal information, less than is used for perception. Indeed, the eye movement is driven after a short delay following stimulus onset for at

least a 240ms duration (the total duration of the analysis windows), while the psychophysical decision of Chapter Three occurs after the entire stimulus presentation (267ms), so it is possible that the psychophysical decision can utilise more of the available information. A further possibility is that low efficiencies are due to the high density stimuli used. It was shown in Chapter Three that perceptual efficiencies decline with increasing dot density, showing that correspondence noise constrains perceptual performance. The high-density stimuli represent the highest densities possible, and pose a severe correspondence problem. It is conceivable that perceptual efficiencies for such high-density stimuli would be comparable to the smooth pursuit efficiencies found here. Finally, the low efficiencies may be related to the flicker noise. In the case of levels of flicker noise approaching 100%, a significant proportion of the moved pixels in the stimulus will reverse polarity in successive frames, effectively producing a reverse phi stimulus. This would elicit motion percepts in the opposite direction to the actual displacements, potentially limiting the performance levels attainable in the task.

5.3.2 Temporal Dynamics

The main finding of the present experiment is that smooth pursuit performance generally improves over time, and does so similarly for the corrugated and transparent stimuli. Indeed, both these stimuli require both the segmentation of motions in different directions and the integration of

motions in same directions (with the difference that in one case transparency is perceived). Recently, there has been a range of research addressing the issue of the temporal dynamics underlying motion perception. For example, it has been shown that pursuit movements in response to 'barber-pole' stimuli are initially in the direction of the one-dimensional motion signals but over time follow the two-dimensional motion (Masson et al., 2000), and that, given two spots to track, pursuit is initiated in the vector average direction of both spots before one is selected for tracking (Lisberger & Ferrera, 1997). Recently the physiological mechanisms underlying such dynamic behaviour have been examined. For example, it has been shown that responses of single neurons in MT to motion stimuli initially suffer from the aperture problem (see Marr & Ullman, 1981 for a definition) but over time respond to the true stimulus motion (Pack & Born, 2001). Furthermore, it was shown that pursuit eye-movements initially deviated from true stimulus motion but over time are corrected (similar to the finding of Masson et al., 2000), with a comparable time course to that of the MT responses. It has also recently been suggested that the temporal dynamics of motion integration observed in area MT depends upon feedback from even higher areas (Pack et al., 2001, but see also the critiwue of Movshon et al., 2003). Indeed, cortical feedback is known to operate in figure/ground segmentation from texture cues (Super et al., 2001; Hupe et al., 1998) and may be a crucial principle of visual cortical function in general (Lamme & Roelfsema, 2000). A recent study has suggested key differences between mechanisms of integration and segmentation, specifically that motion segmentation depends upon the speed dependent responses of

V1 while motion discrimination (demanding integration without segmentation) depends upon the speed dependent responses of the higher cortical visual area MT (Masson et al., 1999). Therefore, in principle the use of the pursuit response to examine the temporal dynamics of integration and segmentation can begin to address the interplay between these fundamental processes. The results of the present experiment suggest that the mechanisms of integration and segmentation at work processing both the corrugated and transparent stimuli follow a similar time course.

5.3.3 Conclusions

I conclude this chapter with two remarks. First, efficiencies were computed for smooth pursuit of transparent and corrugated motions. These efficiencies were computed across a range of speed ratios, and over time. In principle, it is possible to directly compare the oculomotor efficiencies here with those of Chapter Three, to make meaningful conclusions about the similarities of the underlying mechanisms. There were two main impediments to this comparison. First, the stimuli across the experiments are structured differently, therefore mechanisms that come into play in one experiment may not necessarily be targeted in the other experiment. Secondly, the high dot density in the current experiment is beyond the range tested in the experiments of Chapter Three, which is an issue given the decline in efficiencies with dot density. Therefore, before drawing conclusions about the nature of the mechanisms underlying perceptual and oculomotor

performance, it would be desirable to compute perceptual efficiencies for the corrugated and transparent stimuli used here.

My second remark concerns the comparison between smooth pursuit for corrugated and transparent motions. There were two relevant questions here. Will there be a cost in performance for smooth pursuit of transparent motions, paralleling perceptual performance? Will the performance for corrugated and transparent stimuli follow a different time course? The answer to the latter question was negative, in contrast the findings suggested that the mechanisms processing both stimuli follow a similar temporal build-up. The answer to the first question is unanswered by the present study, given the inconsistencies in the results. Again, it would then be desirable to compute perceptual efficiencies for the corrugated and transparent stimuli used here and assess the similarity of the mechanisms at play for perception and action.

Chapter Six: Overall Summary and Conclusions

6.1   Summary

In this thesis I have applied the efficiency approach to examine the phenomenon of visual transparency, in which two surfaces are perceived simultaneously in the same visual direction.   In the seminal computational analysis of Marr (1982), the surface representation is a crucial stage in visual processing that links initial spatially localised descriptions to a later three-dimensional object representation.   Little is known about how this crucial intermediate representation is achieved by the cortical visual system, and it is increasingly becoming a key issue in the field.   Two processes are thought to be fundamental to the recovery of the surface layout in the environment, the segregation and integration of spatially local information (e.g. Braddick, 1993).   The complex nature of visual transparency can be used to probe the limitations of these processes on the recovery of the surface representation, as transparency demands both segregation and integration simultaneously.   In this thesis I examined the limitations on visual transparency elicited purely by motion or stereoscopic information.

As discussed in Chapters Three and Four, previous researchers have found that there is a cost in processing both motion defined transparency and stereoscopic defined transparency compared to similar non-transparent surface tasks.   This cost has been attributed to constraints in the underlying

mechanisms. However, given the acute correspondence problem of transparent stimuli compared to non-transparent stimuli, the cost could reflect a difference in the available information. By applying the efficiency approach for the first time to an analysis of visual transparency, I was able to address this issue. The efficiency approach had two key advantages: 1) performance is normalised to the available information; 2) efficiencies are absolute measures of performance and can be compared directly across tasks. Thus in this thesis I could assess performance within the domains of motion and stereo and also compare performance directly between the two. The key questions I asked were: 1) Does the cost in processing transparent motion reflect a difference in the available information or a limitation in the underlying mechanisms? 2) Does the cost in processing stereoscopic transparency reflect a difference in the available information or a limitation in the underlying mechanisms? 3) How does performance compare between motion and stereo?

I computed efficiencies for visual transparency elicited by differences in motion (Chapter Three) and differences in disparity (Chapter Four) in random dot stimuli, and in addition extended the approach to the analysis of smooth pursuit for transparent motion (Chapter Five). In each case, performance for visual transparency was compared with performance for comparable non-transparent surfaces. Generally, I found that 1) there is a residual cost in efficiency for transparent motion (across a range of dot densities and speed ratios), 2) there is no cost in efficiency for stereoscopic

transparency (across a range of dot densities and speed ratios), 3) both motion and stereo mechanisms are sensitive to correspondence noise. I have provided my interpretation of these results in motion and stereo in Chapters Three and Four. In summary these are that; 1) an inhibitory interaction between different directions of motion constrained performance for transparent motion, 2) there is no inhibitory interaction between disparity defined surfaces (although there may be local inhibitory interactions), 3) there is a correspondence noise sensitive mechanism similar or common to motion and stereo, which may be a spatial pooling of local information.

In the following sections I further consider an account of the different effects of transparency in the motion and stereo tasks, and also consider how this may account for the additional finding of lower efficiencies in stereo compared to motion. I also consider the benefits and limitations of the efficiency approach in general. Finally, I consider the future investigations that can follow from the work in the present thesis.

6.2 The Interplay of Motion and Stereo Mechanisms in Visual Transparency

In the natural environment, it is likely that the resolution of transparency involves the interaction of multiple depth cues (Kersten, 1991). The present results indicate that there is a cost on the recovery of transparency purely from motion cues, but no cost on the recovery of transparency from

stereopsis. This echoes a recent series of psychophysical and physiological findings of Qian and colleagues (Qian & Andersen, 1994; Qian et al., 1994a). In these studies, random dot stimuli were constructed in which pairs of dots moved in opposite directions in close spatial proximity. These stimuli were perceived as 'flicker'. However, when the dots were unpaired such that they did not move in close spatial proximity, transparency was perceived. Furthermore, when paired dots were presented such that each dot had a different disparity, the previous 'flicker' percept was abolished and observers could segregate the two planes of motion. Here, the stereoscopic information resolved the conflict of the local motion information. To account for this Qian et al. (1994b) introduced the 'modified motion energy model', described in Chapter Three, a physiologically plausible account of motion and disparity processing up to the level of area MT. This model is essentially a fusion of the motion energy model (Adelson & Bergen, 1985) and the disparity energy model (Ohzawa et al., 1990). The aspect of the model relevant to the present discussion is that the opponent motion inhibition is restricted within disparity-tuned cells (and within a small spatial region hypothesised to correspond to the size of MT 'subunits'), with no inhibitory interaction between the disparity-tuned channels. This provides a physiologically plausible account of the different effects of transparency on motion and stereo efficiencies. I further suggest that the difference in processing motion and disparity may account for the higher efficiencies in the motion study i.e. the suppressive interactions between motion results in a weaker residual noise response than in the disparity case, where there is no suppressive interaction

between different disparities. Indeed, it has been found that MT responses to pure random noise (of the kind used here) are small, suggesting that the inhibition of different motion signals serves to combat the unwanted effects of correspondence noise (Qian & Andersen, 1994).

The modified motion energy model places the neural interactions that underlie perceptual transparency beyond V1 in an unidentified area, a 'sub-unit' input to area MT. Interestingly, recent physiological evidence suggests a functional interaction of motion and stereo at a sub-unit level (Pack, Born & Livingstone, 2003). In Chapters Three and Four I discussed how the similar dependencies on dot density for motion and stereo tasks suggest that a similar mechanism underlies performance in both cases. One possible mechanism is the spatial pooling of motion and disparity information that occurs within MT. Indeed, it has been found that MT cells selective to both motion and disparity are inhibited in response to different directions of motion within the same disparity, but not to different directions of motion at different disparities (Bradley et al., 1995). Therefore, an ecologically valid situation of transparency in which there are two (moving) surfaces at different depths can be represented effectively by the visual system. This implies that the inhibitory responses identified in MT in response to motion serve to average out motion noise at a given depth (i.e. within a surface). This is not to say that motion information is uninformative, indeed we do perceive depth in the transparent motion displays. It has been suggested that the system achieves this as a result of the zero-disparity motion signals in transparent

motion displays being segregated into different populations of broadly tuned MT neurons, those tuned to near and those tuned to far disparities (Bradley et al., 1998).

The above discussion provides a parsimonious account of my findings on the efficiency of visual transparency defined by motion and stereo information, and suggests that a common neural mechanism underlies the performance in both domains. Indeed, recent psychophysical evidence points to a functional interaction between motion and stereopsis in the recovery of surfaces (e.g. Lankheet & Palmen, 1998; van Ee & Anderson, 2001). However, while I have suggested that the mechanisms may lie in area MT, there is some evidence that in the motion case transparency is limited not by inhibitory interactions at MT (or the MT subunit level) but earlier fine-scale operations in V1 (Masson et al., 1999; Mestre et al., 2001). Indeed, in order to integrate different areas the boundaries must be demarcated, and the scale of V1 processing is better suited to this segmentation task than the larger receptive fields of MT. By this alternative argument, the local interaction limiting motion segmentation (transparency) is an early local averaging within V1. The data in the current thesis cannot distinguish between the two alternatives. Nevertheless, if early V1 responses do limit segmentation, and later MT responses limit integration, then there should be some interplay between these areas, as there is a necessary two-way interaction between the processes of segmentation and integration (e.g. Braddick, 1993). A signature of the interplay between feed-forward and feedback mechanisms could be found in

the temporal dynamics of the system. In the final experiment on smooth pursuit, the motor response was used to probe the temporal dynamics of motion processing. The results indicated a common temporal build-up for the corrugated and transparent stimuli, suggesting that common mechanisms of segmentation and integration operate in both cases. It would then be of interest to use stimuli that dissociate these two processes, and see whether the pursuit response reveals a different temporal build-up of the response.

In summary, while I have provided a parsimonious interpretation of my key results in this section, other possibilities remain open. In the final section 'Further Work', I consider the various research directions that can follow from the work in this thesis to address these issues. First, I further consider the efficiency approach I used in this thesis, highlighting the main benefits of the analysis and also consider the more problematic aspects.

6.3   The Efficiency Approach: Benefits and Limitations

The results of the present thesis demonstrate a number of key benefits of the efficiency approach. First, there were advantages in normalising performance to the available information. In the motion and stereo studies, there was a question as to whether the cost for transparency did reflect a visual mechanism, or alternatively a difference in the available information. The analysis revealed that while there is a residual cost in efficiency for transparent motion, there was no residual cost in efficiency for stereoscopic

transparency. Thus while a visual mechanism underlies the cost in transparent motion, in the stereo-transparency case a difference in the available information accounts for the raw performance cost. If I had not controlled for the available information in these tasks I may have come to quite different (erroneous) interpretations about the nature of the underlying visual mechanisms. Indeed, this criticism applies to any psychophysical task in which there are differences in the available information across conditions. Another aspect of controlling for the available information is that it constrains the interpretation of the results. In Chapters Three and Four, the ideal observer analysis demonstrated that there was a greater incidence of correspondence noise in the transparency conditions, but also a greater incidence of correspondence noise with increasing dot density. Therefore I could attribute the similar decline in efficiency for the motion and stereo tasks in terms of a common sensitivity to correspondence noise.

Another benefit of the efficiency approach was that, in normalising performance to the available information, it was then possible to make direct comparisons across tasks. In the current thesis, the computation of efficiencies permitted the comparison of performance between similar tasks in motion and stereo, and also to smooth pursuit eye movements. At face value comparison suggests that the visual system is better with motion than stereo information, and both are better than the performance of the pursuit motor system. To make a meaningful comparison across these domains required further consideration of the stimuli used in each domain, and the mechanisms

that may be constraining performance in each case. For example, in the previous section quite severe performance limitations were found in my laboratory experiments, compared to a theoretical observer optimised for those artificial tasks. Nevertheless, a careful consideration of the results suggested that the visual system is optimal in the natural environment where multiple cues are available to resolve complex scenes, such as transparency. Similarly, in the survey of efficiencies in Chapter Two, the comparison of efficiencies across the different tasks could be interpreted in terms of a sequence of representational stages and the possible implementation of this sequence in the cortical visual hierarchy.

The above considerations illustrate the power of the efficiency approach in the endeavour to understand the mechanisms of the visual system. However, while the approach offers more than basic psychophysical approaches, there are a number of limitations to be aware of. To apply the method, tasks must be used where performance is constrained by noise. It is important that the noise is relevant to the task at hand i.e. that it targets the same (hypothetical) mechanisms as the task. It must be possible to specify the information available to perform the task and specify optimal performance. This tends to be quite trivial for basic signal detection tasks, but for more complex tasks these limitations can pose a more formidable problem, although there are ways of applying the efficiency approach to more complex situations (Elder, 2002). A further constraint has been suggested above. *The interpretation of differences in efficiencies requires a careful consideration of the differences between*

stimuli and tasks. In the present thesis, the absolute nature of the efficiency measure permitted a direct comparison between the perceptual performance of Chapter Three and the motor performance of Chapter Five. However, while this comparison did suggest some interesting differences between motor and perceptual performance, there were clear differences in the stimuli that may have been contributing to the difference, namely the high dot density and different structure and noise of the pursuit stimuli. In general, while there is much to be gained from applying the efficiency approach, it does impose restrictions on experimentation and the results require considered interpretation.

6.4   Future Work

There are a number of directions that follow from the present findings. One direct extension of the present results is the analysis of combined stereo-motion stimuli, and a number of interesting manipulations could be performed to assess the extent of the proposed stereo-motion interaction. For example, I could assess whether the addition of stereo cues improved efficiencies in the speed discrimination task, or whether the addition of speed cues improved efficiencies in the depth discrimination task. If a local disparity-limited inhibitory interaction does limit performance in the motion case, I predict that the addition of stereo cues to the speed discrimination task will increase efficiencies. Similarly, I predict that if motion and stereo mechanisms are tightly coupled there will be an improvement in the stereo

efficiencies when given the additional motion information. A second direct extension of the present research was suggested in Chapter Three, the computation of efficiency for unidirectional transparent motions. This would directly assess the claims of motion inhibition. The prediction is that if the limiting mechanism in transparency is an inhibitory interaction between different directions of motion, there should be no cost in the efficiency of transparency in the absence of directional cues. A final direct extension of the present work concerns the comparison between perceptual and pursuit performance. As discussed in Chapter Five, it would be desirable to compare these performances for the same stimuli, and to extend the approach to stimuli that dissociate processes of segmentation and integration to assess the underlying temporal dynamics.

A further direction concerns an adaptation of the random dot stimuli used in the present study. These stimuli are spatially 'broad-band', that is to say that each dot in the stimulus contains energy at all spatial frequencies. A more controlled analysis of the mechanisms of integration and segmentation could be achieved by replacing the local elements in my random dot stimuli with Gabor elements, Gaussian windowed spatial frequency gratings. Indeed, the highest efficiencies reported in the literature have been found for Gabor stimuli that match the receptive field structure in early visual cortex. The advantage of using these Gabor stimuli as that the precise spatio-temporal content of the elements can be manipulated. These manipulations could be performed in motion and stereo to further explore the visual mechanisms that

contribute to solving the correspondence problem, segmenting and integrating local information. For example, it would be possible to test the spatio-temporal parameters of the interaction limiting transparent motion.

Further directions would be to compare these indirect behavioural experiments with comparable direct physiological studies to locate the level of the interactions. I suggested in section 6.2 that the similar effects of dot density in motion and in stereo are due to a spatial pooling of local information that occurs in area MT. This could be tested directly by isolating stereo and motion sensitive neurons in MT and assessing the effects of dot density on the neurons response. Other possibilities would be to assess MT neurons response to pure stereoscopic transparency and non-transparent surfaces, to assess whether indeed the response is similar for both and not reduced for transparency. It may also be interesting to assess responses in V1 and MT simultaneously to the transparent and non-transparent motion and stereo stimuli, as this may reveal any interplay between the areas in the process of integration and segmentation. Finally, it should also be possible to apply these designs to localise the mechanisms more directly with human observers by fMRI techniques.

Appendix – Ideal Observer

This appendix provides the foundations of the ideal observer for the experiments. The derivation refers specifically to the case of the motion experiments, but the equations here also apply to the stereo case. First the effective density of the display is computed from the proportion of signal and noise dots. Then the probability of matching dots from frame to frame is computed. Finally, the decision rule used by the ideal observer is derived.

Effective Density

Each frame is composed of K=10,000 (100 by 100) locations where a dot can appear, and each movie is composed of a sequence of F = 10 such frames. Each movie is produced by randomly throwing U signal dots and V noise dots on the first frame. The signal dots are then moved to the next frames according to their desired speed while the noise dots are thrown on new random locations for every frame. If X refers to the total number of dots thrown in one frame, X = U + V. Let s and n denote the probability that a location in a frame is a signal or noise dot respectively. By definition, s = U / K and n = V / K. Let now s' and n' denote the probability that a dot thrown in a frame is a signal or noise dot respectively. By definition, s' = U / X and n' = V / X, so that:

$$s' + n' = 1 . \qquad (1)$$

From these definitions, we also get:

$$s / s' = n / n'. \tag{2}$$

These definitions allow us to compute the density of the stimulus. In the coherency task, the density d is defined as the sum of the densities in the two temporal intervals. Because signal and noise dots can superimpose, the effective density in one interval is:

$$\frac{d}{2} = s + n - sn \; . \tag{3}$$

In the experiments, the density d and the proportion of noise dots n' are set, and so we can infer the density probabilities of signal and noise s and n from Equations (1), (2) and (3). Similarly, in the transparency task, the effective density of the stimulus is:

$$d = 1 - \left[(1 - n)(1 - s)^2\right]. \tag{4}$$

We can infer the density probabilities of signal and noise from Equations (1), (2) and (4).

The task involves a comparison of the slow and fast displacements of the signal dots. Because of ambiguities in matching dots across frames (the correspondence problem), each stimulus contains multiple speeds even in the noiseless condition. The multiple speeds contained in a stimulus are exactly represented in the cross-correlation computed across frames. Such a cross-correlation will present two peaks (one at the fast and the other at the slow speed) and a baseline level that corresponds to matching two unrelated dots by chance. Let c_1 and c_2 denote the peak and baseline amplitudes of the cross-correlation in the coherency task, and similarly t_1 and t_2 for the transparency task.

We now derive these values c_1, c_2, t_1 and t_2 explicitly. Let us start with c_1 and, for the sake of the argument, let's assume that one of the signals is a displacement to the right by 2 positions and the second signal a displacement to the left by 4 positions. The value c_1 is the sum of two probabilities c_{11} and c_{12} corresponding to the two intervals in a coherency trial. If the first signal is presented in the first interval, then c_{11} is the joint probability of observing a dot at location i at time t and a dot at location $(i + 2)$ at time $(t + 1)$ that we find to be:

$$c_{11} = s + (1 - s)n^2 .$$

(5)

Because the first signal was presented in the first interval, only chance will participate to the cross-correlation at the same lag of +2 in the second interval:

$$c_{12} = \left(\frac{d}{2}\right)^2 .$$

(6)

Similar reasoning for the transparency task lead to the following table:

$$\begin{cases} c_1 = s + (1-s)n^2 + d^2/4 \\ c_2 = d^2/2 \\ t_1 = s + (1-s)(s+n-sn)^2 \\ t_2 = d^2 \end{cases} .$$

(7)

Of course, Equations (3) and (4) above can be used to eliminate s from Equation (7) and obtain cross-correlation amplitudes purely in function of d and n.

The values obtained in Equation (7) are the means of the cross-correlation amplitudes for an infinite number of trials. For one particular trial, let { b_1, b_2, b_3, b_4} denote the amplitudes of the cross-correlation at the four bins of interest (in the example above, the bins at −4, −2, +2 and +4 lags). Each b_i follows a binomial distribution b(N, R), where $N = K.(F − 1)$ and R is one of the base probabilities given in Equation (7). Since N is large (90,000), the binomial distributions we are dealing with are indistinguishable from Normal

distributions with mean $\mu = R$ and variance $\sigma^2 = R \cdot (1 - R) / N$. For instance, the first bin in our example will be distributed as:

$$p(\text{bin}_1 = b_1) = \frac{1}{\sqrt{2\pi c_1(1-c_1)/N}} \exp\left[-\frac{(b_1 - c_1)^2}{2c_1(1-c_1)/N}\right] . \tag{8}$$

Decision Rule

The final stage of the ideal observer model is to combine the amplitudes of the cross-correlation and select the decision rule. Given that there are only two possible choices for the ideal observer, the optimal decision rule is to select leftward motion whenever

$$p(\text{leftward} \mid \text{stimulus}) > p(\text{rightward} \mid \text{stimulus}) \tag{9}$$

where the 'stimulus' is represented by the four amplitudes of the cross-correlation function as described above. These posterior conditional probabilities can be rewritten as functions of likelihoods using Bayes' rule:

$$p(\text{leftward} \mid \text{stimulus}) = \frac{p(\text{stimulus} \mid \text{leftward})\, p(\text{leftward})}{p(\text{stimulus})} . \tag{10}$$

Given that the denominator in Equation (10) is a constant for a particular trial, and that leftward and rightward motions are equally likely, the decision rule in Equation (9) can be rewritten in terms of the likelihood ratio D:

$$D = \frac{p(\text{stimulus} \mid \text{leftward})}{p(\text{stimulus} \mid \text{rightward})} > 1 . \tag{11}$$

We therefore only need to focus on the likelihoods. If we assume independence between the bins of the cross-correlation, the likelihood for the coherency task becomes:

$$\begin{aligned}
p(\text{stimulus} \mid \text{leftward}) &= p(\text{bin}_1 = b_1, \text{bin}_2 = b_2, \text{bin}_3 = b_3, \text{bin}_4 = b_4 \mid \text{leftward}) \\
&= p(b_1 = c_1, b_2 = c_2, b_3 = c_1, b_4 = c_2) \\
&= p(b_1 = c_1)\, p(b_2 = c_2)\, p(b_3 = c_1)\, p(b_4 = c_2) \\
&= \frac{1}{4\pi^2 \sigma_1^2 \sigma_2^2} \exp\left[-\frac{(b_1 - c_1)^2}{2\sigma_1^2} - \frac{(b_2 - c_2)^2}{2\sigma_2^2} - \frac{(b_3 - c_1)^2}{2\sigma_1^2} - \frac{(b_4 - c_2)^2}{2\sigma_2^2} \right]
\end{aligned} \tag{12}$$

If we further assume that near threshold, the variances of the peak and baseline amplitudes of the cross-correlations will be approximately equal ($\sigma_1^2 = \sigma_2^2 = \sigma^2$), the likelihood can be further simplified:

$$p(\text{stimulus} \mid \text{leftward}) = \frac{1}{4\pi^2 \sigma^4} \exp\left[-\frac{(b_1 - c_1)^2 + (b_2 - c_2)^2 + (b_3 - c_1)^2 + (b_4 - c_2)^2}{2\sigma^2} \right] . \tag{13}$$

The likelihood ratio then becomes:

178

$$D = \exp\left[-\frac{(b_1 - c_1)^2 + (b_2 - c_2)^2 + (b_3 - c_1)^2 + (b_4 - c_2)^2}{2\sigma^2} \right.$$

$$\left. + \frac{(b_1 - c_2)^2 + (b_2 - c_1)^2 + (b_3 - c_2)^2 + (b_4 - c_1)^2}{2\sigma^2} \right].$$

$$= \exp\left[\frac{(c_1 - c_2)}{\sigma^2} \cdot (b_1 - b_2 + b_3 - b_4) \right]$$

(14)

Since $c_1 \geq c_2$ the decision rule from Equation (11) simplifies to:

$$(D > 1) \Leftrightarrow (b_1 + b_3) > (b_2 + b_4) .$$

(15)

Similar reasoning lead to the same decision rule in the transparency task.

The decision rule in Equation (15) is equivalent to template matching with two templates. The leftward template has only two peaks at bins 1 and 3 (in our example, speeds −4 and +2) and the rightward template has peaks at bins 2 and 4. This template matching procedure is the one that is implemented in the simulations of the ideal observer.

179

References

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

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal Coherence of Moving
Visual-Patterns. *Nature, 300*(5892), 523-525.

Akerstrom, R. A., & Todd, J. T. (1988). The Perception of Stereoscopic
Transparency. *Perception & Psychophysics, 44*(5), 421-432.

Andrews, D. P., Butcher, A. K., & Buckley, B. R. (1973). Acuities for spatial
arrangement in line figures: human and ideal observers compared.
Vision Research, 13, 599-620.

Anzai, A., Ohzawa, I., & Freeman, R. D. (1999a). Neural mechanisms for
encoding binocular disparity: receptive field position versus phase.
Journal of Neurophysiology, 82, 874-890.

Anzai, A., Ohzawa, I., & Freeman, R. (1999b). Neural mechanisms for
processing binocular information. I. Simple cells. *Journal of
Neurophysiology, 82*, 891-908.

Anzai, A., Ohzawa, I., & Freeman, R. (1999c). Neural mechanisms for
processing binocular information. II. Complex cells. *Journal of
Neurophysiology, 82*, 909-924.

Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The Physical Limits of
Grating Visibility. *Vision Research, 27*(11), 1915-1924.

Barlow, H. (1978). The efficiency of detecting changes of density in random

dot patterns. *Vision Research, 18,* 637-650.

Barlow, H. (1980). The absolute efficiency of perceptual decisions. *Philosophical Transactions of the Royal Society of London Series B, 290,* 71-82.

Barlow, H., & Tripathy, S. P. (1997). Correspondence noise and signal pooling in the detection of coherent visual motion. *Journal of Neuroscience, 17*(20), 7954-7966.

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *Journal of Physiology (London), 119,* 69-88.

Barlow, H. B. (1962). Measurements of the quantum efficiency of discrimination in human scotopic vision. *Journal of Physiology, 160,* 169-188.

Barlow, H. B. (1962). A method of determining the overall quantum efficiency of visual discriminations. *Journal of Physiology, 160,* 155-168.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception, 1,* 371 - 394.

Barlow, H. B. (1977). Retinal and central factors in human vision limited only by noise. In H. B. Barlow & P. Fatt (Eds.), *Vertebrate photoreception* (pp. 337-358). London: Academic Press.

Barlow, H. B., & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symettry in random dot displays. *Vision Research, 19,* 783-793.

Bennett, P. J., Sekuler, A. B., & Ozin, L. (1999). Effects of aging on calculation efficiency and equivalent noise. *Journal of the Optical Society of America a-Optics Image Science and Vision, 16*(3), 654-668.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer-Verlag.

Berkeley, G. (1709/1975). *Works on Vision* (New ed.). London: Dent.

Beutter, B. R., & Stone, L. S. (1998). Human Motion Perception and Smooth Eye Movements Show Similar Directional Biases for Elongated Apertures. *Vision Research, 38*(9), 1273-1286.

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

Born, R. T., Groh, J. M., Zhao, R., & Lukasewycz, S. J. (2000). Segregation of Object and Background Motion in Visual Area MT: Effects of MIcrostimulatioon on Eye Movements. *Neuron, 26,* 725-734.

Born, R. T., Pack, C. C., & Zhao, R. (2002). Integration of motion cues for the initiation of smooth pursuit eye movements. In J. Hyona & D. P. Munoz & W. Heide & R. Radach (Eds.), *Progress in Brain Research* (Vol. 140, pp. 225-237).

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research, 14,* 519-527.

Braddick, O. (1993). Segmentation Versus Integration in Visual-Motion Processing. *Trends in Neurosciences, 16*(7), 263-268.

Braddick, O. (1997). Local and global representations of velocity: transparency, opponency, and global direction perception. *Perception, 26*(8), 995-1010.

Braddick, O., Wishart, K. A., & Curran, W. (2002). Directional performance in

motion transparency. *Vision Research, 42,* 1237-1248.

Bradley, D. C., & Andersen, R. A. (1998). Center-Surround Antagonism Based on Disparity in Primate Area MT. *Journal of Neuroscience, 18*(18), 7552-7565.

Bradley, D. C., Qian, N., & Andersen, R. A. (1995). Integration of Motion and Stereopsis in Cortical Area MT of the Macaque. *Nature, 373,* 609-611.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Braje, W. L., Tjan, B. S., & Legge, G. E. (1995). Human-Efficiency for Recognizing and Detecting Low-Pass Filtered Objects. *Vision Research, 35*(21), 2955-2966.

Bravo, M. J., & Watamaniuk, S. N. J. (1995). Evidence For 2 Speed Signals - a Coarse Local Signal For Segregation and a Precise Global Signal For Discrimination. *Vision Research, 35*(12), 1691-1697.

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. *Journal of Neuroscience, 12*(12), 4745-4765.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of Neurons in Macaque Mt to Stochastic Motion Signals. *Visual Neuroscience, 10*(6), 1157-1169.

Burgess, A. (1985). Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America Series A, 2*(9), 1498-1507.

Burgess, A., & Barlow, H. B. (1983). The Precision of Numerosity

Discrimination in Arrays of Random Dots. *Vision Research, 23*(8), 811-820.

Burgess, A. E. (1990). High level visual decision inefficiencies. In C. Blakemore (Ed.), *Vision: coding and efficiency* (pp. 431-440). Cambridge: Cambridge University Press.

Burgess, A. E. (1999). Visual signal detection with two-component noise: low-pass spectrum effects. *Journal of the Optical Society of America a-Optics Image Science and Vision, 16*(3), 694-704.

Burgess, A. E., & Ghandeharian, H. (1984). Visual Signal-Detection .2. Signal-Location Identification. *Journal of the Optical Society of America a-Optics Image Science and Vision, 1*(8), 906-910.

Burgess, A. E., Li, X., & Abbey, C. K. (1997). Visual signal detectability with two noise components: Anomalous masking effects. *Journal of the Optical Society of America a-Optics Image Science and Vision, 14*(9), 2420-2442.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of Human Visual Signal Discrimination. *Science, 214*(4516), 93-94.

Burns, N. R., & Zanker, J. M. (2000). Estimating internal noise for human visual-motion-detection mechanisms. *Perception, 29*, 92-93.

Burt, P., & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science, 208*, 615 - 617.

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

Carl, J. R., & Gellman, R. S. (1987). Human Smooth Pursuit: Stimulus-Dependent Responses. *Journal of Neurophysiology, 57*(5), 1446-1463.

Chawla, D., Buechel, C., Edwards, R., Howseman, A., Josephs, O., Ashburner,

J., & Friston, K. J. (1999). Speed-dependent responses in V5: A replication study. *Neuroimage, 9*(5), 508-515.

Chawla, D., Phillips, J., Buechel, C., Edwards, R., & Friston, K. J. (1998). Speed-dependent motion-sensitive responses in V5: An fMRI study. *Neuroimage, 7*(2), 86-96.

Chen, Y. Z., Matthews, N., & Qian, N. (2001). Motion rivalry impairs motion repulsion. *Vision Research, 41*(27), 3639-3647.

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. *Journal of Neurophysiology, 71*(6), 2269-2280.

Cohen, B., Matsuo, V., & Raphan, T. (1977). Quantitative Analysis of the Velocity Characteristics of Optokinetic Nystagmus and Optokinetic After-nystagmus. *Journal of Physiology (London), 270,* 321-344.

Collewijn, H., & Tamminga, E. P. (1984). Human smooth and saccadic eye movements during voluntary pursuit of different target motions on different backgrounds. *Journal of Physiology (London), 351,* 217-250.

Collewijn, H., van der Mark, F., & Jansen, T. C. (1975). Precise recordings of human eye-movements. *Vision Research, 15,* 447-450.

Cook, E. P., & Maunsell, J. H. R. (2002). Dynamics of neuronal responses in macaque MT and VIP during motion detection. *Nature Neuroscience, 5*(10), 985-994.

Cormack, L. K., Landers, D. D., & Ramakrishan, S. (1997). Element density and the efficiency of binocular matching. *Journal of the Optical Society of*

America a-Optics Image Science and Vision, 14(4), 723-730.

Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1991). Interocular

   Correlation, Luminance Contrast and Cyclopean Processing. *Vision*

   Research, 31(12), 2195-2207.

Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1993). Disparity-tuned

   channels of the human visual system. *Visual Neuroscience, 10,* 585-596.

Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1994). An Upper Limit to the

   Binocular Combination of Stimuli. *Vision Research, 34*(19), 2599-2608.

Crick, F., & Koch, C. (1998). Constraints on cortical and thalamic projections:

   the no-strong-loops hypothesis. *Nature, 391,* 245-250.

Cumming, B. G., & DeAngelis, G. C. (2001). The Physiology of Stereopsis.

   Annual Review of Neuroscience, 24, 203-238.

Curran, W., & Braddick, O. J. (2000). Speed and direction of locally-paired dot

   patterns. *Vision Research, 40*(16), 2115-2124.

De Bruyn, B., & Orban, G. A. (1999). What is the speed of transparent and

   kinetic-boundary displays? *Perception, 28,* 703-709.

De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Perceptual filling-in:

   a paramteric study. *Vision Research, 38,* 2721-2734.

DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area

   MT and the perception of stereoscopic depth. *Nature, 394,* 677-680.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1998). Depth is encoded in

   the visual cortex by a specialized receptive field structure. *Nature,*

   352(156-159).

DeValois, R. L., & DeValois, K. K. (1988). *Spatial Vision.* New York: Oxford

University Press.

Eagle, R. A., & Rogers, B. J. (1996). Motion detection is limited by element density not spatial frequency. *Vision Research, 36*(4), 545-558.

Eagle, R. A., & Rogers, B. J. (1997). Effects of dot density, patch size and contrast on the upper spatial limit for direction discrimination in random-dot kinematograms. *Vision Research, 37*(15), 2091-2102.

Eckstein, M., Beutter, B. R., & Stone, L. S. (1998). *Quantitative Metrics Relating Human Perceptual and Oculomotor Performance During Visual Search* (NASA/TM-1998-208762): NASA.

Eckstein, M. P., Ahumada, A. J., & Watson, A. B. (1997). Visual signal detection in structured backgrounds .2. Effects of contrast gain control, background variations, and white noise. *Journal of the Optical Society of America a-Optics Image Science and Vision, 14*(9), 2406-2419.

Eckstein, M. P., Beutter, B. R., & Stone, L. S. (2001). Quantifying the performance limits of human saccadic targeting during visual search. *Perception, 30*, 1389-1401.

Elder, J. H., Morgenstern, Y., & Tabone, R. H. (2002). A new ideal-observer formulation for perceptual completion. *Perception, 31*, 109-110 Suppl.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkoff und Hartel.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.

Fodor, J. A. (1983/1996). *The Modularity of Mind*. London: MIT Press.

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

Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar,

hyperbolic, and Cartesian gratings in macaque visual cortex. *Science, 259,*

100 - 103.

Geisler, W. S. (1989). Sequential Ideal-Observer Analysis of Visual

Discriminations. *Psychological Review, 96*(2), 267-314.

Gepshtein, S., & Cooperman, A. (1998). Stereoscopic transparency: a test for

binocular vision's disambiguating power. *Vision Research, 38*(19), 2913-

2932.

Gibson, J. J. (1979/1986). *The ecological approach to visual perception.* New Jersey:

Lawrence Erlbaum.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999a). Identification of band-pass

filtered letters and faces by human and ideal observers. *Vision Research,*

39(21), 3537-3560.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999b). Signal but not noise changes

with perceptual learning. *Nature, 402*(6758), 176-178.

Gray, M. S., Pouget, A., Zemel, R. S., Nowlan, S. J., & Sejnowski, T. J. (1998).

Reliable disparity information through selective integration. *Visual*

Neuroscience, 15, 511-528.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.*

New York: Wiley.

Grimson, W. E. L. (1985). Computational Experiments with a Feature Based

Stereo Algorithm. *IEEE Transactions on Pattern Analysis and Machine*

Intelligence, 7(1), 17-34.

Groh, J. M., Born, R. T., & Newsome, W. T. (1997). How is a Sensory Map

Read Out? Effects of Microstimulation in Visual Area MT on Saccades and Smooth Pursuit Eye Movements. *The Journal of Neuroscience, 17*(11), 4312-4330.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology, 35*(1), 96 - 111.

Hallett, P. E. (1987). Quantum efficiency of dark adapted human vision. *Journal of the Optical Society of America Series A, 4*(12), 2330-2335.

Harris, J. M., & Parker, A. J. (1992). Efficiency of Stereopsis in Random-Dot Stereograms. *Journal of the Optical Society of America a-Optics Image Science and Vision, 9*(1), 14-24.

Harris, J. M., & Parker, A. J. (1994). Objective Evaluation of Human and Computational Stereoscopic Visual Systems. *Vision Research, 34*(20), 2773-2785.

Hays, A. V., Richmond, B. J., & Optican, L. A. (1982). A UNIX-based multiple process system for real-time data acquisition and control. *WESCON Conference Proceedings, 2,* 1-10.

Heeger, D. J., Boynton, G. M., Demb, J. B., Seidemann, E., & Newsome, W. T. (1999). Motion opponency in visual cortex. *Journal of Neuroscience, 19*(16), 7162-7174.

Heinen, S. J., & Watamaniuk, S. N. J. (1998). Spatial integration in human smooth pursuit. *Vision Research, 38,* 3785-3794.

Helmholtz, H. von (1867/1925). *Treatise on physiological optics.* 3rd edition. New York: Dover publications.

Hilgetag, C.-C., O'Neill, M. A., & Young, M. P. (1996). Indeterminate

Organization of the Visual System. *Science, 271*(5250), 776.

Hiris, E., & Blake, R. (1996). Direction repulsion in motion transparency.

Visual Neuroscience, 13(1), 187-197.

Hoffman, D. D. (1998). *Visual Intelligence.* New York: Norton.

Horn, B. K. (1975). Obtaining shape from shading information. In P. H.

Winston (Ed.) *The Psychology of Computer Vision.* (p 115 – 135). New

York: McGraw Hill.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interactions,

and functional architecture in the cat's visual cortex. *Journal of Physiology

(London), 160,* 106 - 154.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional

architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal

of Neurophysiology, 28,* 229 - 289.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional

architecture of monkey striate cortex. *Journal of Physiology (London), 195,*

215 - 243.

Hubel, D. H., & Wiesel, T. N. (1998). Early exploration of the visual cortex.

Neuron, 20(3), 401-412.

Huk, A. C., & Heeger, D. J. (2001). Pattern-motion responses in human visual

cortex. *Nature Neuroscience, 5*(1), 72-75.

Hupe, J.-M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J.

(1998). Cortical feedback improves discrimination between figure and

ground by V1, V2 and V3 neurons. *Nature, 394,* 784-787.

Julesz, B. (1964). Binocular Depth Perception without Familiarity Cues. *Science, 145,* 356 - 362.

Kaufman, L., Bacon, J., & Barraso, F. (1973). Stereopsis without image segregation. *Vision Research, 13,* 137-147.

Kawano, K. (1999). Ocular Tracking: behavior and neurophsyiology. *Current Opinion in Neurobiology, 9,* 467-473.

Kersten, D. (1984). Spatial Summation in Visual Noise. *Vision Research, 24*(12), 1977-1990.

Kersten, D. (1987). Statistical Efficiency for the Detection of Visual Noise. *Vision Research, 27*(6), 1029-1040.

Kersten, D. (1990). Statistical limits to image understanding. In C. Blakemore (Ed.), *Vision: coding and efficiency* (pp. 32-44). Cambridge: Cambridge University Press.

Kersten, D. (1991). Transparency and the cooperative computation of scene attributes. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing.* Cambridge, MA: MIT Press.

Kingdom, F., Moulden, B., & Hall, R. (1987). Model for the detection of line signals in visual noise. *Journal of the Optical Society of America Series A, 4*(12), 2342-2354.

Kingsmith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and Unbiased Modifications of the Quest Threshold Method - Theory, Simulations, Experimental Evaluation and Practical Implementation. *Vision Research, 34*(7), 885-912.

Knill, D. C. (1998a). Surface orientation from texture: ideal observers, generic

observers and the information content of texture cues. *Vision Research, 38*(11), 1655-1682.

Knill, D. C. (1998b). Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Research, 38*(11), 1683-1711.

Knill, D. C. (1998c). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research, 38*(17), 2635-2656.

Knill, D. C., Field, D., & Kersten, D. (1990). Human Discrimination of Fractal Images. *Journal of the Optical Society of America a-Optics Image Science and Vision, 7*(6), 1113-1123.

Komatsu, H., Kinoshita, M., & Murakami, I. (2000). Neural responses in the retinotopic representation of the blind spot in the macaque V1 to stimuli for perceptual filling-in. *Journal of Neuroscience, 20*(24), 9310-9319.

Komatsu, H., & Wurtz, R. H. (1988). Relation of Cortical Areas MT and MST to Pursuit Eye Movements. I. Localization and Visual Properties of Neurons. *Journal of Neurophysiology, 60*(2), 580-603.

Komatsu, H., & Wurtz, R. H. (1989). Modulation of Pursuit Eye Movements by Stimulation of Cortical Areas MT and MST. *Journal of Neurophysiology, 62*(1), 31-47.

Kowler, E., van der Steen, J., Tamminga, E. P., & Collewijn, H. (1984). Voluntary selection of the target for smooth eye movement in the presence of superimposed, full-field stationary and moving stimuli. *Vision Research, 24*(12), 1789-1798.

Krauzlis, R. J., & Stone, L. S. (1999). Tracking with the mind's eye. *Trends in*

Neurosciences, 22(12), 544-550.

Kuffler, S. W. (1953). Discharge patterns and functional organization of

mammalian retina. *Journal of Neurophysiology, 16,* 37 - 68.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision

offered by feedforward and recurrent processing. *Trends in Neurosciences,*

23(11), 571-579.

Land, E. H.& McCann, J. J. (1971). Lightness and retinex theory. *Journal of the*

Optical Society of America, 61, 1 –11.

Landy, M. S., Cohen, Y., & Sperling, G. (1984). HIPS: Image processing under

UNIX. Software and applications. *Behavioral Research. Methods,*

Instruments & Computers, 16, 199-216.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995).

Measurement and modeling of depth cue combination: In defense of

weak fusion. *Vision Research, 35*(3), 389-412.

Legge, G. E., Gu, Y., & Luebker, A. (1989). Efficiency of graphical perception.

Perception and Psychophysics, 46(4), 365-374.

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast Discrimination in

Noise. *Journal of the Optical Society of America a-Optics Image Science and*

Vision, 4(2), 391-404.

Lindsey, D. T., & Todd, J. T. (1998). Opponent motion interactions in the

perception of transparent motion. *Perception & Psychophysics, 60*(4), 558-

574.

Lisberger, S. G., & Ferrera, V. P. (1997). Vector Averaging for Smooth Pursuit

Eye Movements Initiated by Two Moving Targets in Monkeys. *The*

Journal of Neuroscience, 17(19), 7490-7502.

Lisberger, S. G., Morris, E. J., & Tychsen, L. (1987). Visual Motion Processing and Sensory-Motor Integration for Smooth Pursuit Eye Movements. *Annual Review of Neuroscience, 10*(97-129).

Lisberger, S. G., & Westbrook, L. E. (1985). Properties of Visual Inputs that Initiate Horizontal Smooth Pursuit Eye Movements in Monkeys. *The Journal of Neuroscience, 5*(6), 1662-1673.

Liu, Z. L., & Kersten, D. (1998). 2D observers for human 3D object recognition? *Vision Research, 38*(15-16), 2507-2519.

Liu, Z. L., Kersten, D., & Knill, D. C. (1999). Dissociating stimulus information from internal representation - a case study in object recognition. *Vision Research, 39*(3), 603-612.

Liu, Z. L., Knill, D. C., & Kersten, D. (1995). Object Classification For Human and Ideal Observers. *Vision Research, 35*(4), 549-568.

Livingstone, M., & Hubel, D. (1988). Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception. *Science, 240,* 740-749.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577 - 621.

Mamassian, P., Landy, M. & Maloney, L.T. (2001). Bayesian Modelling of Visual Perception. In R. P. N. Rao, B. A. Olshausen & M. S. Lewicki, M. S. (Eds.) *Probabilistic Models of the Brain: Perception and Neural Function* (pp. 13 – 36). Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* New York: W. H.

Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London Series B, 207,* 187-217.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London Series B, 200,* 269-294.

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

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London Series B-Biological Sciences, 204,* 301-328.

Marr, D., & Ullman, S. (1981). Directional Selectivity and Its Use in Early Visual Processing. *Proceedings of the Royal Society of London Series B-Biological Sciences, 211*(1183), 151- 180.

Marshak, W., & Sekuler, R. (1979). Mutual repulsion between moving visual targets. *Science, 205,* 1399-1401.

Masson, G. S., & Mestre, D. R. (1998). A look into the black box: eye movements as a probe of visual motion processing. *Current Psychology of Cognition, 17,* 809-829.

Masson, G. S., Mestre, D. R., & Stone, L. S. (1999). Speed tuning of motion segmentation and discrimination. *Vision Research, 39*(26), 4297-4308.

Masson, G. S., Rybarczyk, Y., Castet, E., & Mestre, D. R. (2000). Temporal dynamics of motion integration for the initition of tracking eye movements at ultra-short latencies. *Visual Neuroscience, 17,* 753-767.

Masson, G. S., & Stone, L. S. (2002). From following edges to pursuing objects. *Journal of Neurophysiology, 88*, 2869-2873.

Mather, G., & Moulden, B. (1980). A simultaneous shift in apparent direction: further evidence for a "distribution-shift" model of direction coding. *Quarterly Journal of Experimental Psychology, 32*, 325-333.

Mather, G., & Moulden, B. (1983). Thresholds for movement direction: Two directions are less detectable than one. *Quarterly Journal of Experimental Psychology, 35A*, 513-518.

Maunsell, J., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience, 10*, 363 - 401.

Maunsell, J. H. R., & Van Essen, D. C. (1983). The connections of the middle temporal area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neurophysiology, 3*, 2563-2586.

Maunsell, J. H. R., & van Essen, D. C. (1983). Functional-Properties of Neurons in Middle Temporal Visual Area of the Macaque Monkey .1. Selectivity for Stimulus Direction, Speed, and Orientation. *Journal of Neurophysiology, 49*(5), 1127-1147.

Mckee, S. P., Silverman, G. H., & Nakayama, K. (1986). Precise Velocity Discrimination Despite Random Variations in Temporal Frequency and Contrast. *Vision Research, 26*(4), 609-619.

Mestre, D. R., & Masson, G. S. (1997). Ocular Responses to Motion parallax Stimuli: The Role of Perceptual and Attentional Factors. *Vision Research, 37*(12), 1627-1641.

Mestre, D. R., Masson, G. S., & Stone, L. S. (2001). Spatial scale of motion

segmentation from speed cues. *Vision Research, 41,* 2697-2713.

Metelli, F. (1974). The perception of transparency. *Scientific American, 230,* 91 - 98.

Miles, F. A. (1998). The neural processing of 3-D visual information: evidence from eye movements. *European Journal of Neuroscience, 10,* 811-822.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action.* Oxford: Oxford University Press.

Minksy, M. & Papert, S. (1969). *Perceptrons: An introduction to computational geometry.* Cambridge, MA: MIT Press.

Mishkin, M., Ungerleider, L., & Macko, K. (1983). Object vision and spatial vision: Two central pathways. *Trends in Neuroscience, 6,* 414 - 417.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1986). The analysis of moving visual patterns. In C. Chagas & R. Gattass & C. Gross (Eds.), *Pattern Recognition Mechanisms.* New York: Springer Verlag.

Movshon, J. A., Albright, T. D., Stoner, G. R., Majaj, N. J., & Smith, M. A. (2003). Cortical responses to visual motion in alert and anesthetized monkeys. *Nature Neuroscience, 6*(1), 3-4.

Mulligan, J. B. (1993). Nonlinear Combination Rules and the Perception of Visual-Motion Transparency. *Vision Research, 33*(14), 2021-2030.

Murakami, I. (1997). Motion transparency in superimposed dense random-dot patterns: psychophysics and simulation. *Perception, 26*(6), 679-692.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual Surface Representation: A critical link between Lower-Level and Higher-Level Vision. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual Cognition.* Cambridge, MA: MIT

Press.

Newsome, W. T., & Pare, E. B. (1988). A Selective Impairment of Motion
Perception Following Lesions of the Middle Temporal Visual Area (Mt).
Journal of Neuroscience, 8(6), 2201-2211.

Niemann, T., Ilg, U. J., & Hoffmann, K. P. (1994). Eye movements elicited by
transparent stimuli. *Experimental Brain Research, 98*, 314-322.

Ohzawa, I. (1998). Mechanisms of stereoscopic vision: the disparity energy
model. *Current Opinion in Neurobiology, 8*, 509-515.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. (1996). Encoding of binocular
disparity by simple cells in the cat's visual cortex. *Journal of
Neurophysiology, 75*, 1779-1805.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth
discrimination in the visual cortex: neurons ideally suited as disparity
detectors. *Science, 249*, 1037-1041.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1997). Encoding of binocular
disparity by complex cells in the cat's visual cortex. *Journal of
Neurophysiology, 77*, 2879-2909.

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

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.

Pack, C. C., Born, R. T., & Livingstone, M. S. (2003). Two-Dimensional

Substructure of Stereo and Motion Interactions in Macaque Visual

Cortex. *Neuron, 37,* 525-535.

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial

frequencies, noise, and the efficiency of letter discrimination. *Vision

Research, 31(7/8),* 1399-1415.

Parker, A. J., & Yang, Y. (1989). Spatial Properties of Disparity Pooling in

Human Stereo Vision. *Vision Research, 29(11),* 1525-1538.

Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.),

Vision: coding and efficiency (pp. 3-24). Cambridge: Cambridge University

Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics:

Transforming numbers into movies. *Spatial Vision, 10(4),* 437-442.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of

America a-Optics Image Science and Vision, 16(3),* 647-653.

Perret, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to

faces in the monkey temporal cortex. *Experimental Brain Research, 47,* 329

- 342.

Perrone, J. A., & Thiele, A. (2001). Speed skills: measuring the visual speed

analyzing properties of primate MT neurons. *Nature Neuroscience, 4(5),*

526-532.

Pola, J., & Wyatt, H. J. (1985). Active and passive smooth eye movements:

effects of stimulus size and location. *Vision Research, 25(8),* 1063-1076.

Pollard, S. B., & Frisby, J. P. (1990). Transparency and the Uniqueness

Constraint in Human and Computer Stereo Vision. *Nature, 347(6293),*

553-556.

Pollard, S. B., Mayhew, J. E. W., & Frisby, J. P. (1985). PMF - a Stereo
Correspondence Algorithm Using a Disparity Gradient Limit. *Perception,*
14(4), 449-470.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental*
Psychology, 32(1), 3 - 25.

Prazdny, K. (1985). Detection of Binocular Disparities. *Biological Cybernetics,*
52, 93 - 99.

Qian, N. (1994). Computing stereo disparity and motion with known
binocular cell properties. *Neural Computation, 6,* 390-404.

Qian, N., & Andersen, R. A. (1994). Transparent Motion Perception as
Detection of Unbalanced Motion Signals. II. Physiology. *The Journal of*
Neuroscience, 14(12), 7367-7380.

Qian, N., Andersen, R. A., & Adelson, E. H. (1994). Transparent Motion
Perception as Detection of Unbalanced Motion Signals. I. Psychophysics.
The Journal of Neuroscience, 14(12), 7357-7366.

Qian, N., Andersen, R. A., & Adelson, E. H. (1994). Transparent Motion
Perception as Detection of Unbalanced Motion Signals. III. Modeling.
The Journal of Neuroscience, 14(12), 7381-7392.

Rashbass, C. (1961). The relationship between saccadic and smooth tracking
eye movements. *Journal of Physiology (London), 159,* 326 - 338.

Rayner, K. (1977). Visual Attention in Reading: Eye movements reflect
cognitive processes. *Memory and Cognition, 5,* 443-448.

Rayner, K. (1998). Eye movements in reading and information processing:

Twenty years of research. *Psychological Bulletin, 124,* 372-422.

Read, J. C. A. (2002). A Bayesian model of stereopsis and motion direction

discrimination. *Biological Cybernetics, 86,* 117-136.

Rizzolatti, G., Riggio, L., Dascola, I., & Umilta, C. (1987). Reorienting of

attention across the horizontal and vertical meridians - evidence in favor

of a premotor theory of attention. *Neuropsychologia, 25*(1A), 31-40 Special

Issue.

Rose, A. (1948). The Sensitivity Performance of the Human Eye on an

Absolute Scale. *Journal of the Optical Society of America, 38*(2), 196-208.

Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of*

brain mechanisms. Washington DC: Spartan.

Shimojo, S., Kamitani, Y., & Nishida, S. (2001). Afterimage of a perceptually

filled-in surface. *Science, 293,* 1677-1680.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in

visual area MT. *Vision Research, 38*(5), 743-761.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and

neural representation. *Annual Review of Neuroscience, 24,* 1193-1216.

Simpson, W. A., & Manahilov, V. (2001). Matched filtering in motion

detection and discrimination. *Proceedings of the Royal Society of London*

Series B-Biological Sciences, 268(1468), 703-709.

Simpson, W. A., Manahilov, V., & Mair, M. S. (1999). Is motion processing

unitary? *Journal of the Optical Society of America a-Optics Image Science and*

Vision, 16(12), 2836-2844.

Smith, A. T., Curran, W., & Braddick, O. J. (1999). What motion distributions

yield global transparency and spatial segmentation? *Vision Research,*
39(6), 1121-1132.

Snowden, R. J. (1989). Motions in Orthogonal Directions Are Mutually
Suppressive. *Journal of the Optical Society of America a-Optics Image Science
and Vision, 6*(7), 1096-1101.

Snowden, R. J. (1990). Suppressive Interactions Between Moving Patterns -
Role of Velocity. *Perception & Psychophysics, 47*(1), 74-78.

Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The
Response of Area Mt and V1 Neurons to Transparent Motion. *Journal of
Neuroscience, 11*(9), 2768-2785.

Solomon, J. A., & Pelli, D. G. (1994). The Visual Filter Mediating Letter
Identification. *Nature, 369*(6479), 395-397.

Steinman, R. M. (1986). The need for an eclectic, rather than systems,
approach to the study of the primate oculomotor system. *Vision Research,
26*(1), 101-112.

Stevenson, S. B., Cormack, L. K., & Schor, C. M. (1989). Hyperacuity, Super-
resolution and Gap Resolution in Human Stereopsis. *Vision Research,
29*(11), 1597-1605.

Stevenson, S. B., Cormack, L. K., & Schor, C. M. (1991). Depth Attraction and
Repulsion in Random Dot Stereograms. *Vision Research, 31*(5), 805-813.

Stevenson, S. B., Cormack, L. K., & Schor, C. M. (1992). Disparity tuning in
mechanisms of human stereopsis. *Vision Research, 32*(9), 1685-1694.

Stone, L. S., Beutter, B. R., & Lorenceau, J. (2000). Visual motion integration
for perception and action. *Perception, 29,* 771-787.

Stoner, G. R., & Albright, T. D. (1992). Neural Correlates of Perceptual Motion

Coherence. *Nature, 358*(6385), 412-414.

Stoner, G. R., Albright, T. D., & Ramachandran, V. S. (1990). Transparency

and coherence in human motion perception. *Nature, 344,* 153-155.

Super, H., Spekreijse, H., & Lamme, V. A. F. (2001). Two distinct modes of

sensory processing observed in monkey primary visual cortex (V1).

Nature Neuroscience, 4(3), 304-310.

Swets, J. A. (Ed.). (1964). *Signal detection and recognition by human observers.*

New York: Wiley.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in

perception. *Psychological Review, 68*(5), 301-340.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1964). Decision Processes in

Perception. In J. A. Swets (Ed.), *Signal detection and recognition by human*

observers. New York: Wiley.

Tanaka, H., & Saito, H. (1989). Analysis of motion of the visual field by

direction, expansion/contraction, and rotation cells clustered in the

dorsal part of the medial superior temporal area of the macaque

monkey. *Journal of Neurophysiology, 62,* 626-641.

Tanner, W. P., & Birdsall, T. G. (1958). Definitions of d' and η as Pschophysical

Measures. *The Journal of the Acoustical Society of America, 30*(10), 922-928.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual

detection. *Psychological Review, 61*(6), 401-409.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human-Efficiency

For Recognizing 3-D Objects in Luminance Noise. *Vision Research, 35*(21),

3053-3069.

Todd, J. T., & Norman, J. F. (1995). The Effects of Spatiotemporal Integration

on Maximum Displacement Thresholds for the Detection of Coherent

Motion. *Vision Research, 35*(16), 2287-2302.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability

of signals in single neurons in cat and monkey visual-cortex. *Vision

Research, 23*(8), 775-785.

Treue, S., Hol, K., & Rauber, H. J. (2000). Seeing multiple directions of motion

- physiology and psychophysics. *Nature Neuroscience, 3*(3), 270-276.

Tsai, J. T., & Victor, J. D. (2003). Reading a population code: a multi-scale

neural model for representing binocular disparity. *Vision Research, 43*,

445-466.

Turing, A. M. (1950). Computational machinery and intelligence. *Mind,

59(236)*, 433 – 460.

Tyler, C. (1991). Cyclopean Vision. In D. Regan (Ed.), *Binocular Vision* (pp. 38-

74). London: MacMillan.

Underleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. G.

Ingle & M. A. Goodale & R. J. Q. Mansfield (Eds.), *Analysis of visual

behaviour* (pp. 549 - 586). Cambridge, MA: MIT Press.

van Doorn, A. J., & Koenderink, J. J. (1982a). Temporal properties of the visual

detectability of moving spatial white noise. *Experimental Brain Research,

45*, 179-188.

van Doorn, A. J., & Koenderink, J. J. (1982b). Spatial Properties of the Visual

Detectability of Moving Spatial White Noise. *Experimental Brain Research,*

45, 189-195.

van Ee, R., & Anderson, B. L. (2001). Motion direction, speed and orientation in binocular matching. *Nature, 410,* 690-694.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information Processing in the Primate Visual System: An Integrated Systems Perspective. *Science, 255,* 419-423.

Van Essen, D. C., & Maunsell, J. H. R. (1980). Two-dimensional maps of the cerebral cortex. *Journal of Comparative Neurology, 191,* 255-281.

Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neuroscience, 6(9),* 370 - 375.

Van Meeteren, A., & Barlow, H. B. (1981). The Statistical Efficiency for Detecting Sinusoidal Modulation of Average Dot Density in Random Figures. *Vision Research, 21(6),* 765 - 777.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224(4654),* 1260 - 1262.

Watamaniuk, S. N. J. (1993). Ideal observer for discrimination of the global direction of dynamic random-dot stimuli. *Journal of the Optical Society of America Series A, 10(1),* 16-28.

Watamaniuk, S. N. J., & Heinen, S. J. (1999). Human smooth pursuit direction discrimination. *Vision Research, 39,* 59-70.

Watanabe, K. (1999). Optokinetic Nystagmus with spontaneous reversal of transparent motion perception. *Experimental Brain Research, 129,* 156-160.

Watson, A. B., Barlow, H. B., & Robson, J. G. (1983). What Does the Eye See

Best. *Nature, 302*(5907), 419-422.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive

psychometric method. *Perception & Psychophysics, 33*(2), 113-120.

Watson, A. B., & Turano, K. (1995). The Optimal Motion Stimulus. *Vision*

Research, 35(3), 325-336.

Watt, R. J., & Andrews, D. P. (1982). Contour curvature analysis: hyperacuities

in the discrimination of detailed shape. *Vision Research, 22,* 449-460.

Weinshall, D. (1989). Perception of multiple transparent planes in stereo

vision. *Nature, 341,* 737-739.

Weinshall, D. (1991). Seeing "ghost" planes in stereo vision. *Vision Research,*

31(10), 1731-1748.

White, B. W. (1962). Stimulus-conditions affecting a recently discovered

stereoscopic effect. *American Journal of Psychology, 75,* 411-420.

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from

stochastic local motions. *Vision Research, 24*(1), 55- 62.

Zanker, J. M. (2001). Combining Local Motion Signals: A Computational

Study of Segmentation and Transparency. In J. M. Zanker & J. Zeil

(Eds.), *Motion Vision - Computational, Neural, and Ecological Constraints*

(pp. 113-124). New York: Springer Verlag.

Zanker, J. M., Srinivasan, M. V., & Egelhaaf, M. (1999). Speed tuning in

elementary motion detectors of the correlation type. *Biological*

Cybernetics, 80, 109-116.