Bayesian Model of the Dynamics of Motion Integration in Smooth Pursuit and Plaid Perception

K. D. Dimova

Doctor of Philosophy

2010

Bayesian model of the dynamics of motion integration in smooth pursuit and plaid perception

by

Kameliya Dragomirova Dimova

A thesis submitted to the University of Plymouth

in partial fulfilment for the degree of

Doctor of Philosophy

School of Computing and Mathematics

Faculty of Science and Technology

January 2010

Kameliya Dragomirova Dimova

Bayesian model of the dynamics of motion integration in smooth pursuit and plaid perception

Abstract. In this thesis, a model of motion integration is described which is based on a recursive Bayesian estimation process. The model displays a dynamic behaviour qualitatively similar to the dynamics of the motion integration process observed experimentally in smooth eye pursuit and plaid perception. The computer simulations of the model applied to smooth pursuit eye movements confirm the psychophysical data both in humans and monkeys, and the physiological data in monkeys. The temporal dynamics of motion integration is demonstrated together with its dependence on contrast, size of the stimulus and added noise. A new theoretical approach to explaining plaid perception has been developed, based on both the application of the model and a novel geometrical analysis of the plaid's pattern. It is shown that the results from simulating the model are consistent with the psychophysical data about the plaid motion. Furthermore, by formulating the model as an approximate version of a Kalman filter algorithm, it is shown that the model can be put into a neurally plausible, distributed recurrent form which coarsely corresponds to the recurrent circuitry of visual cortical areas V1 and MT. The model thus provides further support for the notion that the motion integration process is based on a form of Bayesian estimation, as has been suggested by many psychophysical studies, and moreover suggests that the observed dynamic properties of this process are the result of the recursive nature of the motion estimation.

List of Contents

List of Illustrations and Tables

Acknowledgements

I would like to express my deep gratitude toward my main supervisors: Prof. Michael Denham and Dr. Thomas Wennekers for their great help and support in accomplishing this work.

This thesis is dedicated to my sons Angel and Dragomir with all my love.

AUTHOR'S DECLARATION

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

The work reported here was supported by funding under the Sixth Research Framework Programme of the European Union under the grant no. 15879 (FACETS).

Relevant scientific seminars and conferences were regularly attended at which work was often presented; external institutions were visited for consultation purposes; two articles have been published.

Publications:

Dimova, K. D. & Denham, M. J. (2008). A neurally plausible model of the dynamics of motion integration in smooth eye pursuit based on recursive Bayesian estimation. *Biological Cybernetics*, **100**(3):185-201.

Dimova, K. D. & Denham, M. J. (2010). **A *model of plaid motion perception based on recursive Bayesian integration of the 1-D and 2-D motions of plaid features.*** ***Vision Research***, 50(6), 585-597. doi:10.1016/j.visres.2010.01.004

Presentation and Conferences Attended:

Poster presentation at the Optical Society of America Fall Vision Meeting, October 24 - 26, 2008, Rochester, New York. Regular attendance at the FACETS meetings regarding the work on the project and reporting the current results. (Graz, Austria, Oct. 2006; Nice, France, Oct. 2006; Freiburg, Germany, Jan. 2007; Marseille, France, Nov. 2007; Debrecen, Hungary, Jan. 2008; Leysin, Switzerland, Jan. 2009)

Words count of main body of thesis: 36 234 Signed: K Dimova

Date: 1. 09. 2010

Chapter 1. Introduction.

One of the main questions of brain research which has been highly studied for many years is how humans and animals perceive the world around them, through their sense of vision. One reason why this question has been intensely studied is because the areas of the brain which respond to visual stimulation are easily accessible and, at least in the primate brain, appear to be organised in a systematic way. It is also easy to stimulate a human or animal subject with visual stimuli whilst measuring brain responses to the stimuli. Vision is also arguably the most important of our senses, especially taking into account the amount of the information which it provides. One of the central questions in vision research is how motion is perceived, i.e. how the speed and direction of motion of moving objects is detected, both in the case of single objects moving individually and multiple objects moving relative to each another.

1.1. Functional aspects of the motion perception.

There are different functional aspects of motion perception, and presumably some of them involve different neural mechanisms. Motion integration, perception of pattern motion, relative motion estimation between objects (time to collision), depth perception, motion proprioception, image segmentation are some of the main ones. The model presented here attempts to describe the first two functional aspects of motion perception, i.e. general motion integration as reflected on smooth eye pursuit, and the perception of pattern (plaid) motion. It does so by focusing mainly on the temporal dynamics of motion integration, i.e. how the perception of motion, in particular the perceived direction of motion, changes with time in the case when the perceived object

keeps a constant velocity. Such temporal dynamics could be a valuable source of information about the underlying mechanisms defining the perception of motion.

The perception of motion is a function of highest interest and it is possible to distinguish between "perceiving relative motion between the objects" from "perceiving change of position relative to the retina", as the latter case is considered simpler. It is not that different mechanisms are assumed to be involved, but, even if they are it is necessary to examine the simplest possible case first. Thus in this thesis the estimation of relative speeds between the objects is not considered. It is interesting to mention though, that there could be motion perceived even if there is no change of the position onto the retina or, vice versa, there could be no perception of motion even if the position on the retina has been changed. An instance for the former exceptional case is the visual afterimage in the dark when there is perceived sense of motion without any change of position on the retina (Yasui and Young, 1975). The latter case is shown by the work of Rogers and Graham (1979) where a differential image motion in dot patterns did not lead to the perception of motion, but to depth perception. However, these are cases of exceptional conditions. In order to study the motion perception in a simplified way the case is taken here when the change in the position is relative to the retina only. Furthermore, the interest is focused on the simplest, but still not fully resolved problem of visual motion integration, i.e. the combination of the different motion signals coming from a coherently moving object. A more detailed description of the motion integration problem is presented in the next section.

1.2. Introduction to the problem of visual motion integration

The organisation of the visual processing areas of the brain is now quite well understood, at least in primates, and at least in respect of the early stages of processing which occur after the light stimuli activate nerve cells in the retina. Visual information

is transmitted from the eye to the cerebral cortex via the activity of the retinal ganglion cells. An individual retinal ganglion cell, one of the 1.5 million or so in the human eye, receives input from about 150 million retinal photoreceptors. Each ganglion cell is activated by light coming from a specific and very small region of the visual space (the area which is covered by our normal field of vision), which is referred to as its "receptive field" of that cell. In fact an individual ganglion cell will only respond if the light stimulus, a small spot of flashing light say, is located precisely in this specific area of the visual space, of diameter about 0.5 degrees.

So the brain actually perceives the world via a mosaic of about 1.5 million of overlapping regions, each of which corresponding to a very small part of the visual space. A major question in vision research is how these many small pieces of visual information are integrated to produce our visual perception of the whole visual space. This is of special interest in the case of visual motion perception, where the problem is referred to as that of "motion integration".

The part of the brain which first uses the visual data from the retinal ganglion cells is called the primary visual cortex (denoted as area V1 in the primate brain). This is located at the back of the skull, actually at the furthest point from the eyes, in the occipital region of the cerebral cortex. The retinal ganglion cells project their electrical activity to the neurons located in V1, which are therefore in turn activated in response to a light stimulus. The cells in V1 are activated by several ganglion cells with receptive fields which are contiguously located in the visual space. Thus they respond to a light stimulus which occurs in a specific area of the visual space, which is somewhat larger than an individual ganglion cell, say of diameter between 0.5 and 1 degree.

A neuron in V1 is activated most strongly when it is stimulated by a bar of light which moves in a particular direction, and with a particular orientation, across the cell's receptive field, as shown in Figure 1 (Hubel and Wiesel, 1962, 1965). It has been shown

that there were cells in the retina responding particularly to changes in stimulus direction and velocity independently from other physical dimensions of the stimulus such as contrast, shape or size (Barlow and Hill, 1963; Barlow and Levick, 1965).

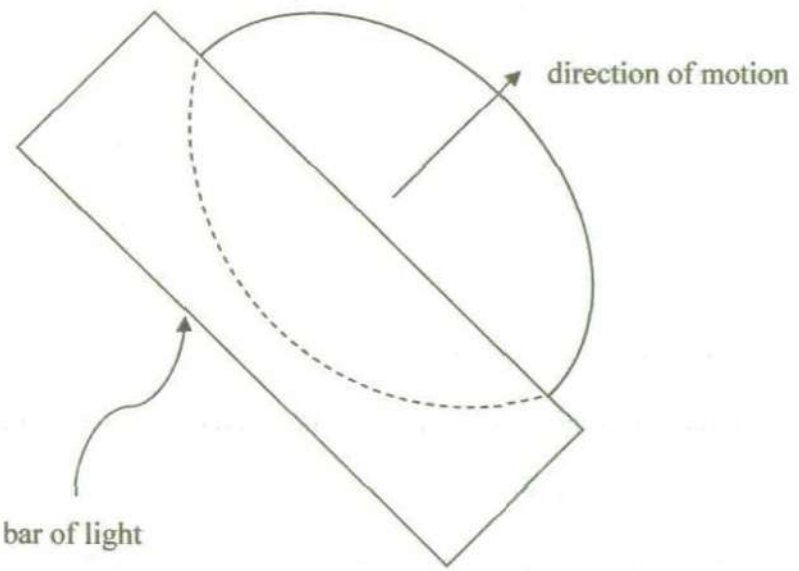


Figure 1. A bar of light with an orientation of 45°, moving upwards to the right, across the elongated, 1-2° diameter receptive field of a V1 neuron.

Thus the activity of a V1 cell is tuned to both the orientation and the direction of motion of the bar of light, although some cells respond almost equally to motion of the bar in the opposite direction (Hubel and Wiesel, 1959). The activity of the cell is only weakly tuned to speed, although the activity will rapidly diminish if the bar moves with a speed higher than a specific threshold, corresponding to the fastest moving images to which our eyes can respond.

A phenomenon which was observed some years ago in relation to the activity of V1 cells is called the "aperture effect" (Wallach, 1935 (English translation in Wuerger

et al., 1996); Fennema and Thompson, 1979; Marr and Ullman, 1981). This phenomenon can be illustrated using a simple experiment. Make a roughly circular hole of about 2 cm diameter in a piece of plain paper or card. Hold a plain pencil behind the card, so that you can see only the middle section of the pencil through the hole, with the remainder of the pencil hidden from view by the card, as shown in Figure 2a.

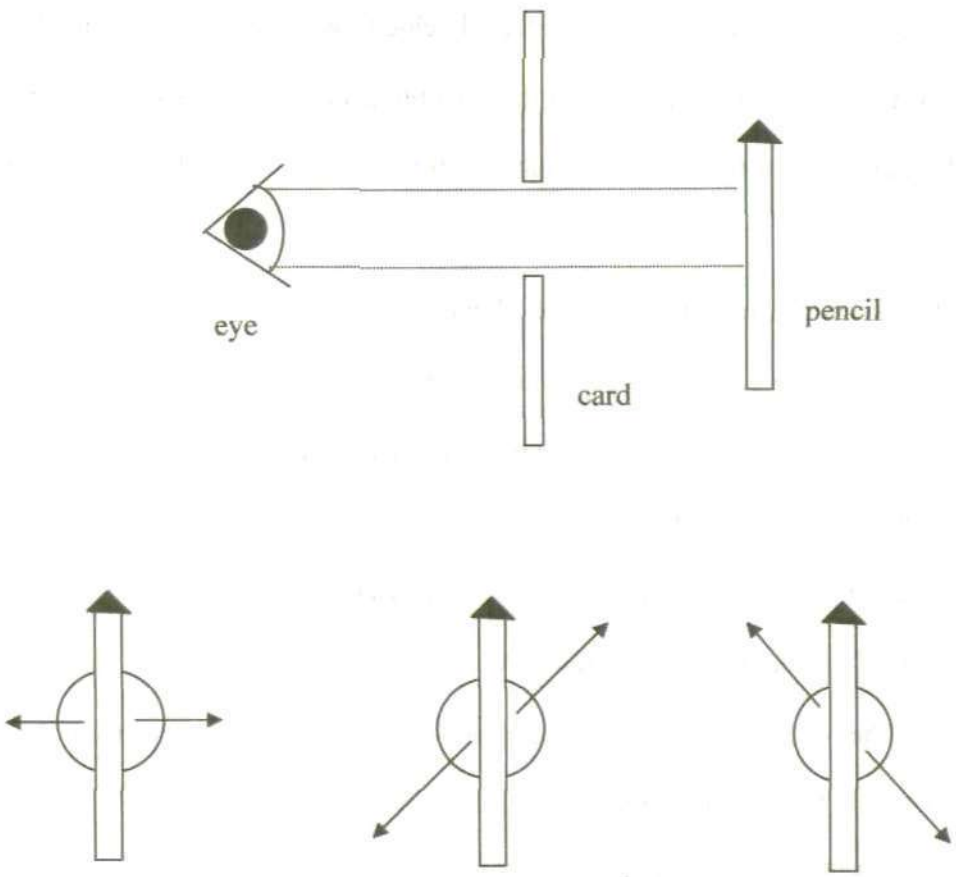


Figure 2. A simple experiment to illustrate the "aperture effect"; a) experimental set-up; b) moving the pencil.

Holding the pencil so that it is vertical, and whilst viewing it through the hole in the card, move the pencil (or the paper) first horizontally back and forth, then upwards and to the right and back at about 45°, then downwards to the right and back at around 45°, always keeping the pencil oriented vertically, as shown in Figure 2b.

What can then be observed is that when only the central section of the pencil is visible through the aperture in the card, it appears to move in the same way, namely horizontally back and forth, in all three cases. This exactly illustrates the case of a V1 neuron when a vertically oriented bar of light (the pencil) moves across its receptive field (the hole in the card). Thus V1 neurons respond only to the "local" motion of the light bar, that is the motion seen "through" their receptive field regardless of the global motion of the light bar. In other words, the detected velocity would be only the one that is orthogonal to the edge of the bar, even if the possible actual velocities of that same edge could be anywhere between approximately +90° to - 90° from the orthogonal velocity.

It follows that the perception of the global motion of objects which we see cannot be the result of the activity of individual V1 neurons, but must result from the combination, or integration, of information from many such neurons. It is presumed that this combination of information takes place at a stage of processing, or brain area, where the projections of the activities of many V1 neurons converge. One such area, amongst several, is called the medial temporal (MT) area in primates, which lies adjacent to V1 in the occipital region of the cerebral cortex of the brain. It has been first discovered in 1971 by Dubner and Zeki.

Neurons in MT are highly tuned to velocity in a similar fashion as the motion detectors in V1, however, the receptive fields of the MT neurons are roughly 10 times bigger than the ones of V1 motion detectors (Albright and Desimone, 1987). As expected for a larger receptive field, it has been shown that the MT neurons are more sensitive to higher speeds than those in V1 (Mikami *et al.*, 1986; Churchland *et al.*, 2005). However, the spatial range over which the MT neurons measure motion is the same as in V1, despite the difference in RF sizes and the velocity preferences (Pack *et al.*, 2003; Churchland *et al.*, 2005; Pack *et al.*, 2006). The input to MT comes mainly

from layers 4B and 6 of the striate cortex (Maunsell and van Essen, 1983). It seems generally accepted that the MT area plays a crucial role in the motion integration process, combining the input of a large number of V1 neurons with visio-spatially adjacent receptive fields (Rodman and Albright, 1989; Stoner and Albright, 1992; Movshon and Newsome, 1996).

To see how this might result in the MT neuron being responsive to the global motion of a moving stimulus, such as a light bar, and also its speed of motion, let us go back to the "aperture effect" experiment with the pencil and card. It should be the same experiment as before, but this time the pencil is positioned in such a way so its tip is seen through the hole, not just its middle section as before. What could be seen is that the global motion of the pencil is now very clear from the motion of the tip of the pencil seen through the aperture, compared to the case when only the edges of the pencil were visible.

Continuing the analogy with V1 neurons and their receptive fields, this simple experiment suggests that if the *end* of the light bar passes through the receptive field of a V1 cell, it may respond in a different way which depends on the direction of motion of the end point of the bar. Thus such a neuron will respond to the *global* motion of the light bar, not just its local motion. V1 neurons have been found which respond best if the end of a light bar stimulus is within the receptive field of the neuron, and many of the cells which project to the area MT have this "end-stopped" property (Hubel and Wiesel, 1965; Pack *et al.*, 2003). The end-stopped cells suppress the 1-D motion signals in favour of 2-D feature detection. When an end of an edge passes the active part of the receptive field of an end-stopping cell the response of it is maximal and at the same time the end-stopping cells could be orientation sensitive.

One way to explore the idea that combining information from both "ordinary" V1 neurons, which suffer from the aperture effect, and "end-stopped" neurons which

may respond to the direction of motion of the stimulus end-points, is to see what happens when humans observe simple moving visual stimuli which span across the receptive fields of many V1 neurons and which contain both simple moving straight edges (like the middle section of the pencil) and moving end-points (like the tip of the pencil). We say that the V1 neurons which are "seeing" only moving straight edges provide only "1-D" local motion direction information, whereas those whose receptive fields contain a moving end-point are providing "2-D" motion direction information.

Such a stimulus is a light rhombus moving horizontally against a dark background, two versions of which are illustrated in Figure 3.

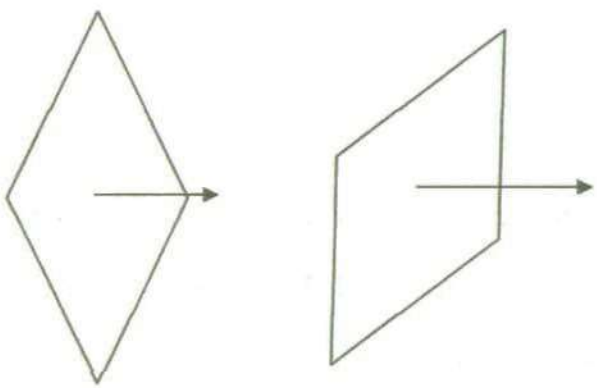


Figure 3. Rhomboid motion stimuli

Clearly, as the straight edges of these stimuli cover a much larger area of the visual space than the end-points, what could be expected is that, since V1 neuron receptive fields are evenly distributed in the central, or foveal, region of the visual field, a much larger number of V1 neurons will respond to edge, or 1-D motion of the four sides of the rhombus than to end point, or 2-D motion of the four tips of the rhombus. It could be assumed that an MT neuron integrates the information in an even-handed way from all the V1 neurons projecting to it, both 1-D and 2-D information. Thus the information received by the MT neuron will be dominated initially by the many more

V1 neurons which are responding to the local 1-D motion of the edges, as illustrated in Figure 4.

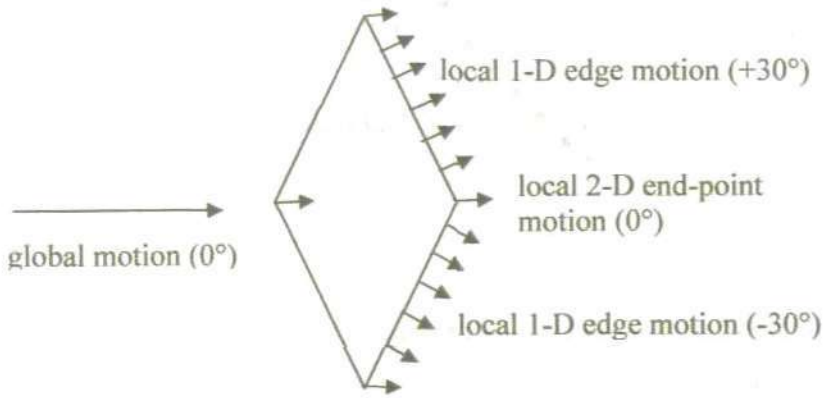


Figure 4. Dominance of 1-D local edge motion compared to the 2-D end point motion for a horizontally moving stimulus.

If the integration operation performed by the MT neurons is a simple averaging of the motion information from the convergent V1 neurons, it is clear from the diagram above that the average motion computed by the MT neuron will be in the horizontal (0°) direction, corresponding to the global motion of the stimulus.

Consider now the case of a stimulus which is the same as the one above, but tilted by 15°. This is illustrated in Figure 5.

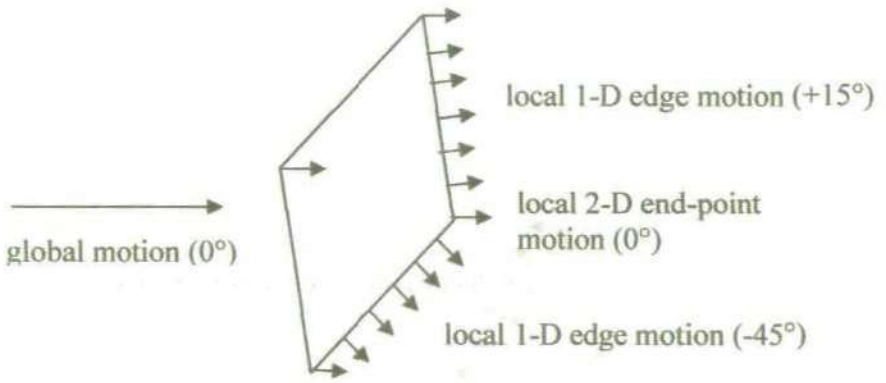


Figure 5. Dominance of 1-D local edge motion compared to the 2-D end point motion for a horizontally moving stimulus tilted at 15 degrees.

In this case, a simple averaging of the convergent 1-D and 2-D motion information by the MT neuron will result in a global motion of about -15°. This clearly does not correspond to the real global motion of the stimulus, which is horizontal (0°). In psychophysical experiments with eye movement recordings of monkeys and also humans, when presented with similar stimuli, it has been shown that their eyes follow initially the non-veridical direction, which is then subsequently corrected to the true one. Thus it has been suggested (Masson and Stone, 2002, Wallace *et al.*, 2002) that the observed changes in direction of the eye movements reflect the dynamics of the motion integration in the brain. The model presented in this thesis tries to capture and replicate this kind of dynamics in processing of the 1-D and 2-D motion signals. In the next section the phenomena of the motion integration as reflected by the smooth pursuit eye movements is described in more detail.

1.3. Motion perception in smooth eye pursuit.

Rashbass (1961) showed for the first time that in the case of two targets moving in the opposite directions, when one of them was moving with a constant velocity, the other with a jump, the result was a smooth eye movement following the former and a saccade following the latter. In addition, the smooth eye movement tended to be the preferred one when the both targets were presented simultaneously. Smooth pursuit eye movements are the rotations of the eye which an observer performs in order that a moving target object is held in an approximately stationary position on the foveal region of the retina so that it could be processed at a highest level of acuity. This is a valuable phenomenon for studying motion perception as it is relatively easy to access experimentally, in terms of the ease of measurement and manipulation of both the inputs to the process, the motion of the visual target, and the outputs, the eye movements. In particular, there has been a substantial effort to understand the neural mechanisms which underlie one of the earliest stages of processing in the oculomotor system, in which the local retinal motion information is integrated to provide accurate global information on motion of the object, the image of which must be held steady on the retina. As it has been described in the previous section, some local information about the global motion of the object is potentially inaccurate owing to the ambiguity introduced by the aperture effect. To ensure that object motion is accurately tracked by the oculomotor system, all the available local information, both unambiguous (2-D) and ambiguous (1-D) must be integrated, in order to provide an estimate of the object motion which will drive the eye movement in the correct target direction.

When human or monkey subjects observe presentations of the first of the two visual stimulus shown in Figure 4 under experimental conditions, and the smooth eye pursuit is monitored, during the first 70 ms or so of the subject tracking the stimulus motion with their eyes, eye motion for this stimulus is in the correct, horizontal direction

(Masson and Stone, 2002, Wallace *et al.*, 2002). For the second, tilted stimulus presented in Figure 5, however, the eye motion in this initial period is in an incorrect direction, that is approximately the direction defined by vector-averaging of the motion vectors of the edges. After this short period of erroneous eye motion, which results presumably from the brain sending a wrong global stimulus motion direction signal to the eye movement system, eye motion converges back to the correct global stimulus motion of 0° and remains there henceforth.

These experiments would seem therefore to support the theory that motion information in the primate brain, in particular in MT, is computed by averaging the local 1-D and 2-D motion information coming from V1 neurons, at least in the initial period of stimulus motion.

It has been shown (Wallace *et al.*, 2002) that there are certain factors influencing the dynamics of the velocity estimation and the magnitude of the bias toward the non-veridical direction. One of the main factors is the contrast of the stimulus, where the lower the contrast, the bigger the bias toward the non-veridical direction is. Besides, the time delay necessary for the veridical velocity to be recovered becomes longer for the lower contrast. The size of the stimulus is another factor influencing the magnitude of the non-veridical bias. The longer the edges of the tilted rhombus are, the stronger the initial bias is.

So why and how does the eye movement return to tracking the stimulus in the correct direction? Clearly the computation of the global stimulus motion direction in MT by a simple averaging operation of local motion signals from V1 is not the whole story. The theory which is put forward in this thesis is that the computation of the global motion direction is not simply an averaging of local motion signals but is based on a recursive Bayesian estimation process, and it is the recurrent nature of the motion estimation which causes a gradual shift in the estimated global motion from an

erroneous, direction of motion, biased by the predominantly 1-D local motion signals, towards the true direction of motion. When applied to the smooth pursuit experimental data obtained from Wallace *et al.*, 2002, the model was able to replicate qualitatively the observed dynamics of the velocity estimation and the effect of contrast and size of stimulus on the magnitude of the non-veridical bias. Furthermore, the effect of contrast on the time delay for the correction of the bias was also replicated by the model.

A similar experimental paradigm of non-veridical perception in the early stages of the motion integration was found in a different type of stimulus, the so-called plaid pattern. More detailed information about the results from psychophysical experiments with plaids and the performance of the proposed model in relation to them is presented in the next section.

1.4. Motion perception in coherent plaid patterns.

When two sinusoidal gratings with different orientations are superimposed and moving together this could lead to the perception of a coherent motion of the plaid pattern which is formed. Adelson and Movshon (1982) first described the conditions necessary for the coherent perception as opposed to the perception of the motion of the two separate component gratings. They found that the gratings should have similar spatial frequencies, contrasts and speeds, also not to differ in depth relative to each other in order to create a coherently moving pattern.

The velocity of the pattern motion is defined by the so-called "intersection of constraints" (IOC) rule (Fenemma and Thompson, 1979), The velocity "constraint" line corresponding to a moving edge defines all the possible velocities with which the edge could move whilst, when viewed through a circular aperture, on a perceptual level all

these velocities would result in the same actually perceived velocity, i.e. that which is in the direction orthogonal to the edge itself An example for such a constraint line is presented in Figure 6.

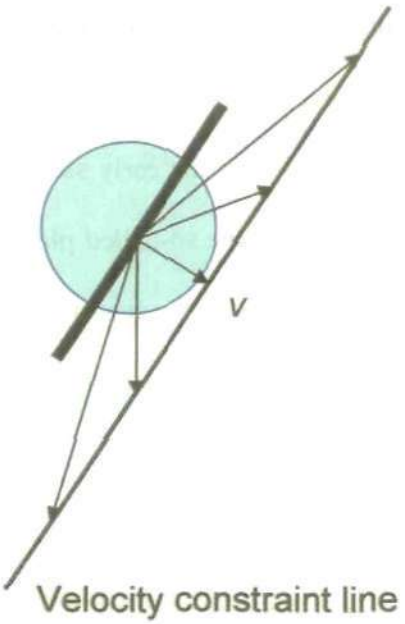


Figure 6. The velocity constraint line formed from all possible velocities of the moving bar, which would result in the same perceived velocity, in a direction orthogonal to the edge of a bar, when viewed through a circular aperture.

For two superimposed gratings forming a plaid with coherent motion, there is a single direction of motion defined by the point in velocity space at the intersection of the two velocity constraint lines corresponding to the two gratings. This is the intersection of constraints (IOC) direction. An example for the IOC direction of two gratings, forming a plaid pattern, is presented in Figure 7.

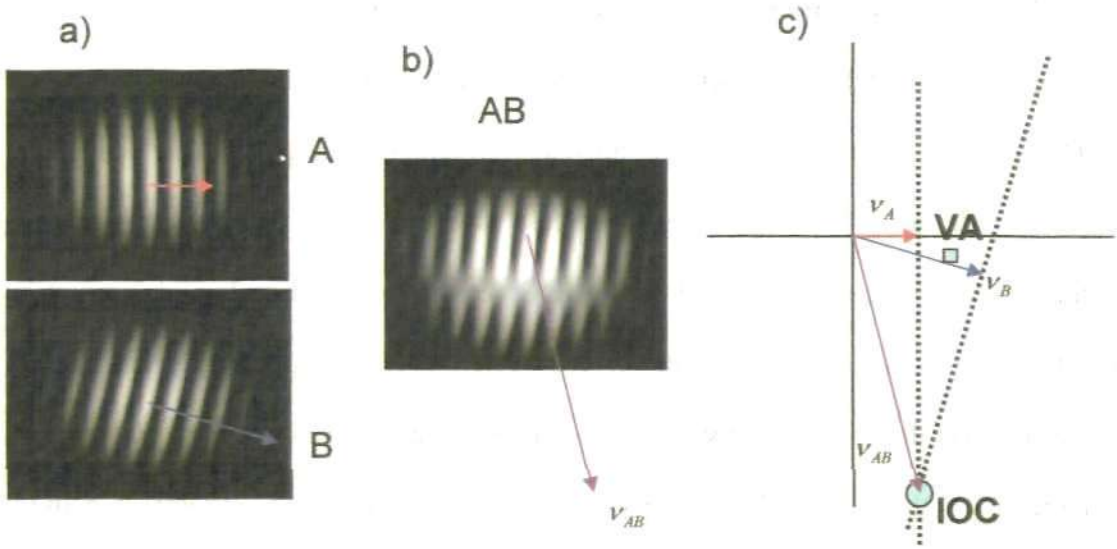


Figure 7. a) Two gratings A and B are presented, with their velocities shown, respectively, by a red and blue vector. b) When overlapped, A and B form the plaid pattern AB with direction of the resultant coherent motion shown by a purple vector (IOC). c) The intersection of the two constraint lines defines the IOC direction (toward the small blue circle), which in this case is much different from the direction of motion corresponding to the vector average (VA)of the two grating velocities (toward the small blue square).

In Figure 7 the velocities of the gratings A and B, presented by a red and blue arrow respectively, form a direction of motion of the plaid pattern AB defined by the intersections of constraints rule and denoted by a purple arrow. The main features of the plaid pattern consist of white and black high and low intensity areas, commonly called in the literature 'blobs'. The perceived direction of motion of the plaid pattern, and of the blobs, is always in the IOC direction. The IOC direction in this example is very different from the vector-average (VA) direction, formed by comuting the vector average of the two grating velocities.

Depending on whether or not there is a difference between the VA and IOC directions, plaids are divided in two types: Type I plaids and Type II plaids. Type I and

Type II plaids have different relative positions of the velocity vectors of the two gratings with regard the IOC vector. According to a classification by Ferrera and Wilson (1990) a Type I plaid is formed when the velocity vectors of the gratings are positioned on either side of the IOC vector, as in the case shown in Figure 8a. When the grating velocity vectors are symmetric about the IOC vector, they form a Type I symmetric plaid, in which case the VA and IOC velocity vectors overlap (Figure 8a). In the case of Type II plaids the two grating velocity vectors both lie on one side of the IOC vector (Figure 8b). In this case the VA and IOC velocity vectors can differ substantially both in length (speed) and direction.

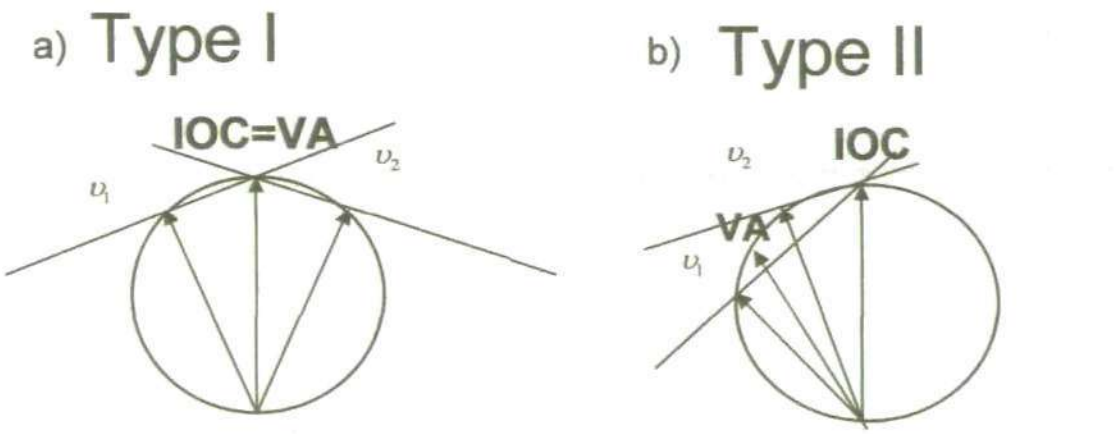


Figure 8. a) a Type I plaid in which the two velocity vectors of the gratings lie on either side of the IOC vector. b) a Type II plaid in which the velocity vectors of the gratings are both positioned on one side of the IOC vector

Although it is known that in general the perceived motion of a coherently moving plaid pattern is in the IOC direction, psychophysical experiments with Type II plaids have revealed that for short presentations of such a stimulus, the direction of the plaid motion could differ significantly from the IOC direction (Ferrera and Wilson, 1990, 1991, Yo & Wilson, 1992).

In the experiments of Yo and Wilson (1992), for example, when observing a moving Type II plaid for a short period of approximately 100ms, the human participants reported a perceived direction of plaid motion which was closer to the VA direction than to the IOC direction. Furthermore, the perceived directional bias found to be dependant on the contrast of the two gratings which formed the plaid, was also found in the experiments on the smooth pursuit eye movements for the case of the tilted rhombus (Masson and Stone, 2002, Wallace *et al.*, 2002).

As in the case of the smooth pursuit eye movement experiments cited above (Masson and Stone, 2002, Wallace *et al.*, 2002), it is shown below that the motion integration model presented in this thesis is able to simulate qualitatively the results of the perceived bias for Type II plaids (Yo & Wilson, 1992, Burke and Wenderoth, 1993, Bowns, 1996).

In the next section a number of models of motion integration which have been proposed are discussed, and the novel form of motion integration model which will be presented in this thesis is described in relation to these other models of motion integration.

1.5. Models of motion integration.

The simplest idea of how MT neurons might combine the inputs of the V1 motion detectors is related to the assumption that a MT neuron simply sums the spatially distributed inputs from detectors in V1 with similar preferred direction. Pack *et al.*, (2006) showed that everywhere in the receptive field of a MT neuron the preferred direction matches that of the V1 neurons which form the inputs to the MT neuron. Although an input summation matches well the ability of MT neurons to increase the signal-to-noise ratio, nevertheless it is not enough to explain how the motion integration

could be achieved. The main reason for that difficulty comes from the aperture problem, because even simple summation of the input could be strongly biased. Consequently, the models of motion integration have to solve conceptually the aperture problem.

There are two main types of models which address this problem. The first one (Adelson and Movshon, 1982, Heeger *et al.*. 1996, Simoncelli and Heeger, 1998) follows the idea of a two-stage processing of the motion signals Only the 1-D motion signals from V1 are measured at the first stage, and then combined non-linearly (so as to extract the 2-D information) at the second stage, presumably in the MT area, in order to estimate the true velocity, following the IOC rule. As proposed by Adelson and Movshon (1982), in the case of a coherently moving plaid the model uses in the first stage the 1-D motion signals from the component gratings, which are further combined in the second stage in such a way as to obtain the true IOC velocity of the pattern. The model does not take into account, or make use of, in the first stage of processing, any 2-D information which the moving pattern would contain, i.e the 2-D motion present in the movement of the blobs of the pattern. This kind of model could be referred to as an *integrationist* one according to a classification suggested by Pack and Born (2008), which discriminates between integrationist versus *selectionist* models. The selectionist models involve a selective process which is needed to first define the 2-D measurements from the image that are most reliable and then to include them in the final computation. They assume a competitive processing of 1-D (direction filters) and 2-D (end-stopping neurons) motion signals in V1, which is again generalised on a second, presumably MT level (Pack *et al.*, 2003). The selectionist models therefore assume that the initial bias toward the vector-averaged 1-D edge motion direction is compensated for by the 2-D cues, however time is necessary for the 2-D features to override the effect of the 1-D features. The integrationist models, on the other hand, relate the time delay in the perception of the veridical velocity to the integration process which is accomplished in

area MT. This model classification is a very general one and not always strict as the models are not necessarily mutually exclusive.

There is psychophysical and physiological evidence supporting both types of models, however the exact nature of the motion integration is still unclear. Weiss *et al.*, (2002) proposed a Bayesian model, which explained well a wide range of psychophysical data related to motion integration and misperception of the veridical velocity, including the misperception of the plaid motion. However the explanation they have given was based on the Adelson and Movshon (1982) two-stage model, taking into account the 1-D component grating velocities only.

The novel motion integration model presented in this thesis is a Bayesian Kalman filter algorithm, in which the first step is mathematically identical to the model presented in Weiss *et al.*, (2002). The Bayesian approach has been applied widely in the recent years in many areas of perception and proved to be quite successful. For instance, the Bayesian model of Weiss *et al.* (2002), described in more details in Weiss' thesis (1998) as well, could explain a spectrum of perceptual phenomena, including the barber pole illusion, biases toward VA direction for stimuli as the ones described above (tilted rhombus and plaids), the contrast effect on the bias toward VA, and others. The Bayesian approach takes into account the probability distributions of the 1-D and 2-D motion signals and also assumes a prior probability distribution for the perceived motion of the stimulus. In the case of the Weiss *et al.* (2002) model the chosen prior distribution for the stimulus motion favours 'slow' stimulus speeds, and thus takes the form of a Gaussian distribution with zero mean and small variance. The reason for such a choice is the need to solve the aperture problem, by favouring the solution with the slowest speed, from the whole family of speeds corresponding to the velocity constraint line of a moving edge (Heeger and Simoncelli, 1993). This slowest speed is in the direction orthogonal to the moving edge. In fact the response of the direction-tuned

neurons in the brain is strongest to the orthogonal velocity of the moving edge (Hubel and Weisel, 1959).

In the Kalman filter based motion integration model presented here, the prior probability distribution for the perceived stimulus velocity corresponds to the probability distribution of the initial velocity estimate in the algorithm, which was also assumed to be a Gaussian distribution with zero mean and small variance. The Kalman filter based model concentrates however on the *dynamical aspects* of the motion integration as a source of knowledge about the underlying mechanisms of the integration itself, by focusing on the recursive nature of the velocity estimation process, which starts from the prior velocity estimate and computes a sequence of updated estimates as each new measurement is made of the spatial and temporal changes in image luminance.

The model presented in this thesis is therefore novel compared to other models in that it attempts to explicitly replicate the motion integration dynamics as it is seen in the initial velocity misperception of the tilted rhombus and Type II plaids during smooth pursuit and plaid perception respectively, by combining the Bayesian approach with Kalman filter algorithm. In addition, reveals that the misperception of the direction of motion of coherent Type II plaids can be explained in terms of the 1-D and 2-D motion of the prominent features of the plaid, the so-called blobs, rather than in terms of the 1-D motion of the component gratings, as originally proposed by Adelson and Movshon (1982) and adhered to by many subsequent researchers (for the many corresponding citations, please refer to the later sections of this thesis). Thus, whilst it may seem to the reader that the plaid stimuli considered in the thesis are somewhat outdated, the work reported here reveals that the central problem of how motion in plaid patterns is perceived is still essentially unresolved, even after nearly thirty years of research. The thesis makes a significant original contribution to the resolution of the problem, by

demonstrating a simple gradient-based, Kalman filter model, the behaviour of which is explained theoretically in terms of the 1-D and 2-D motion of the plaid blobs, and which closely replicates many of the experimental results on Type II plaid misperception.

It is important to note that the presented Kalman filter based motion integration model is a strictly computational model, i.e. it reflects an abstract level approach, and any concepts regarding its neural plausibility should be taken on an abstract computational level as well, that is, in terms of abstract rather than biophysical forms of neural computation. In this context, the model is however presented in two mathematical forms, one form corresponding to the standard Kalman filter formulation; and a second form in which certain steps of the standard algorithm were approximated in order to simplify the computations involved. In particular, the approximated form of the model resulted from the attempt to create a more neurally plausible form (in the abstract neural computational sense) of the algorithm, by avoiding the calculation of inverse matrices. This form of the algorithm also allowed the presentation of the model in a distributed recurrent form, similar to how supposedly the brain might organise the recurrent flow of motion information between the primary visual cortex (V1) and the medial temporal area (MT). However, it was beyond the scope of the current work to demonstrate neural plausibility in a more specific and biophysical way, involving comparison with what is known about the anatomy and physiology, and about the neural responses and cell parameters, in areas V1 and MT.

Similarly, the parameters of the stimulus used in the simulations, the size of the background, speed of motion and size of the 'receptive fields', were not based strictly on known or assumed values of biological neuronal parameters. However, as it will be shown in Chapter 4, the choice of the parameters are quite reasonable in that they show a high level of consistency with the experimental data, not only in qualitative but also in a quantitative sense.

As with all models of natural physical phenomema, the model contains "free parameters", the choice of values for which can influence the behaviour of the model to a greater or lesser extent. In fact, for the presented model, there is only one such free parameter, i.e. the ratio of the variance of the measurement noise associated with the observation of the spatial and temporal changes in the luminance intensity of the stimulus, and the variance of the prior distribution of the perceived stimulus velocity, which in terms of the Kalman filter corresponds to the distribution of the initial velocity estimate. The values of these variances were chosen from a wide range after extensive preliminary testing, and were systematically selected in relation to the specific stimulus simulation parameters, which were subsequently held constant for all the reported model simulations. Once the optimal values for the variances were chosen from what was actually quite a small range of possible values for the selected stimulus simulation parameters, these values of variance. and hence the variance ratio "free parameter", was held constant throughout both of the main sets of model simulations, for the smooth pursuit and the plaid perception experiments. This could be cosidered as a demonstration of the robustness of the model, since each set of model simulations involved two quite different visual stimuli, a tilted rhombus in smooth pursuit and plaid patterns in plaid perception. More details on the methodology of the choice of the model parameters will be presented in Chapter 4.

1.6. Original contributions and structure of the thesis.

The main results and original contributions of this thesis are:

- An abstract-level recursive estimation model of the dynamics of motion integration based on the Kalman filter algorithm, the behaviour of which closely replicates the main characteristics of the dynamics of the misperception of stimulus velocity, in particular those observed experimentally in the smooth

pursuit of moving rhomboid stimuli and in the perception of motion of coherent Type II plaids, including the effects of stimulus changes on the dynamics: (i) changes in contrast and size of the stimulus in smooth pursuit experiments; and (ii) changes in contrast, size and relative grating angles in the plaid motion perception experiments.

- An approximate version of the model algorithm which eliminates the need for the computation of inverse matrices, thus forming a version of the model which is more suited for neural computation. The results from the approximate and the standard algorithm are demonstrated to be very similar, again closely replicating the experimentally observed data. In fact, the approximate version version of the model computes a first velocity estimate, which is different from the standard model and thus also from the estimate of Weiss et al. (2002) model, but more closely replicates the experimentally observed initial non-veridical bias in smooth eye pursuit for the high-contrast rhombus case, which could not be replicated by either the standard model or the Weiss et al. (2002) model.
- A complete geometrical analysis of the main features of the plaid pattern, the so called 'blobs', which has not been presented previously in the literature, and which is used to (i) theoretically predict the observed simulation behaviour of the motion integration model; and (ii) provide an alternative and novel theoretical explanation of the misperception of Type II plaid motion to that of Adelson and Movshon (1982), based on the geometry and the 1-D (edge) and 2-D (end-point) motion of the blob features.
- A demonstration of the possible analogy between the misperception of the plaid velocity based on blob geometry, and the direction bias in smooth pursuit for tilted rhombus.stimuli, showing that the edges of the blobs in the plaid could

play a perceptually similar role to the edges of the tilted rhombus in the smooth pursuit.

- Predictions for the possible motion misperception of certain forms of asymmetric Type I plaids, i.e. those in which the IOC and VA directions of motion are substantially different.
- Proposals for possible physiological experiments involving Type II plaids and predictions for their potential outcomes.

The thesis has the following structure. Following this introductory chapter, the mathematical formulation of the standard and the approximate versions of the Kallman filter based motion integration models are presented in Chapter 2. The rest of the thesis is divided into two main parts. The first part, comprising Chapters 3, 4, and 5, describes on the application of the model to the experimental data on the dynamics of motion integration in smooth eye pursuit. The second part, comprising Chapters 6, 7, 8, and 9, describes the application of the model to the experimental data on plaid motion perception. Finally Chapter 10 gives the general conclusions of the thesis and outlines some future work possibilities. Appendix A includes a general introduction to the basic principles of the Kalman filter algorithm. Appendices B and C contain copies of the two published papers which resulted from the work described in the thesis.

Chapter 2. A Bayesian model of motion integration based on the Kalman filter.

2.1. A *motion integration model based on the optimal* Kalman filter.

A model for motion integration is proposed in this thesis which is based on a recursive Bayesian estimation algorithm, commonly known as the Kalman filter after its originator (Kalman, *1960*), and uses a spatially distributed set of local observations of spatial and temporal changes in the image intensity to estimate global image velocity. For those readers who are unfamiliar with the Kalman filter, an intuitive, simplified explanation of its operation is given in Appendix A. Also introduced in Appendix A is the notation used in the following sections to describe the algorithm and the model.

At each local position (x,y) in the visual space, identified with the receptive field positions of V1 neurons, it is assumed that observations are made of the change in the intensity of the image $I(x,y,t)$ over a small change in time Δt, and over a small spatial change in the two cardinal directions (x,y), i.e. $I_t(x,y,t)$, $I_x(x,y,t)$ and $I_y(x,y,t)$, defined by

$$I_t(x,y,t) \triangleq \left[I(x,y,t+\Delta t) - I(x,y,t)\right]/\Delta t$$
$$I_x(x,y,t) \triangleq \left[I(x+\Delta x,y,t) - I(x,y,t)\right]/\Delta x$$
$$I_y(x,y,t) \triangleq \left[I(x,y+\Delta y,t) - I(x,y,t)\right]/\Delta y$$

There is a basic assumption in the formulation of the motion estimation problem regarding the intensity constancy. It is assumed that image intensity changes with time

at a location (x, y) are only the result of the image motion. Then it follows that, if v_x and v_y are the x and y components of the image velocity vector v,

$$\begin{aligned} I(x, y, t) &= I(x + \Delta x, y + \Delta y, t + \Delta t) \\ &= I(x + \Delta t.v_x, y + \Delta t.v_y, t + \Delta t) \end{aligned} \tag{1}$$

The right hand side of (1) can be approximated by its first order Taylor series expansion, giving

$$I(x, y, t) \approx I(x, y, t) + v_x I_x(x, y, t) + v_y I_y(x, y, t) + I_t(x, y, t) \tag{2}$$

This equation can be rewritten as

$$I_t(x, y, t) = -\begin{bmatrix} I_x(x, y, t) & I_y(x, y, t) \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} + \eta(x, y, t) \tag{3}$$

where $\eta(x, y, t)$ is a zero mean, Gaussian distributed noise with variance σ^2, representing the measurement error (Fennema and Thompson, 1979; Heeger and Simoncelli, 1993; Weiss *et al.*, 2002).

Equation (3) can be thought of as an observation equation for the unknown image velocity vector v, in which the intensity derivatives I_t, I_x and I_y are all measured (observed) values, and be rewritten in the discrete time form as a sequence of observations of the unknown vector v_k, i.e.

$$h_k = C_k v_k + \eta_k \tag{4}$$

where

$h_k \triangleq I_t(x, y, t)$
$C_k \triangleq -\begin{bmatrix} I_x(x, y, t) & I_y(x, y, t) \end{bmatrix}$
$t \triangleq k.\Delta t$

and where η_k is the zero mean, Gaussian distributed noise sequence with variance σ^2, representing the measurement error in the observation equation. Note that time is

expressed in the algorithm in terms of number of iterations of the basic time step Δt, i.e. $t \triangleq k.\Delta t$

The process of estimating the velocity vector v becomes one of maximising the posterior probability density function $p(v_0,...,v_N \mid h_1,...,h_N)$ with respect to $\{v_0,...,v_N\}$. Bayes' rule defines the posterior probability of a certain event given a set of observations related to this event, as the product of the likelihood of the observations given the event and the prior probability of the event, normalised by the probability of the observations. It is expressed as

$$p(v_0,...,v_N \mid h_1,...,h_N) = \frac{p(h_1,...,h_N \mid v_0,...,v_N)\,p(v_0,...,v_N)}{p(h_1,...,h_N)} \tag{5}$$

It could be written

$$p(h_1,...,h_N \mid v_0,...,v_N) = \prod_{k=1}^{N} p_{h_k}(h_k - C_k v_k) \tag{6}$$

and assuming that the image velocity is constant,

$$p(v_0,...,v_N) = p(v_0) \tag{7}$$

Thus Bayes' rule becomes

$$\begin{aligned} p(v_0,...,v_N \mid h_1,...,h_N) &\propto p(v_0)\prod_{k=1}^{N} p_{h_k}(h_k - C_k v_k) \\ &\propto \exp\left\{-\frac{1}{2}(v_0-\overline{v_0})^T P_0^{-1}(v_0-\overline{v_0}) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(h_k - C_k v_k)^T(h_k - C_k v_k)\right\} \end{aligned} \tag{8}$$

where $\overline{v_0}$ and P_0 are prior estimates of the mean and covariance matrix of the probability distribution of the image velocity vector v_k.

Rather than find the maximum of this posterior distribution directly, the process can be formulated as a recursive estimation procedure, ie in the form of a Kalman filter estimation algorithm (see Appendix A for an intuitive explanation of the derivation of this algorithm and for details of the notation used in its description):

1. initialisation:

$$v_0^0 = \overline{v_0}$$

$$P_0^0 = P_0$$

2. update of the estimate between observations (this captures the dynamics of the stimulus – in this case the stimulus velocity is constant and so the estimate is unchanged between observations):

$$v_{k+1}^k = v_k^k$$

$$P_{k+1}^k = P_k^k$$

3. update of the estimate at observation *k*

$$v_k^k = v_k^{k-1} + K_k(h_k - C_k v_k^{k-1})$$
$$P_k^k = P_k^{k-1} - K_k C_k P_k^{k-1}$$
$$K_k = P_k^{k-1} C_k^T (C_k P_k^{k-1} C_k^T + \sigma^2)^{-1}$$

where K_k is the so called Kalman gain. We can rewrite the Kalman gain as

$$K_k = [(P_k^{k-1})^{-1} + C_k^T C_k \frac{1}{\sigma^2}]^{-1} C_k^T \frac{1}{\sigma^2} \tag{9}$$

If we assume that $P_k^{k-1} = \sigma_p^2 I_2$, where I_2 denotes the 2x2 identity matrix, then

$$\begin{aligned} K_k &= [\frac{1}{\sigma_p^2} I_2 + C_k^T C_k \frac{1}{\sigma^2}]^{-1} C_k^T \frac{1}{\sigma^2} \\ &= [\frac{\sigma^2}{\sigma_p^2} I_2 + C_k^T C_k]^{-1} C_k^T \end{aligned} \tag{10}$$

Substituting for C_k in (10), we get

$$K_k = \begin{bmatrix} I_x^2 + \frac{\sigma^2}{\sigma_p^2} & I_x I_y \\ I_x I_y & I_y^2 + \frac{\sigma^2}{\sigma_p^2} \end{bmatrix}^{-1} \begin{bmatrix} I_x \\ I_y \end{bmatrix} \tag{11}$$

If $\overline{v_0}$, the prior estimate of the mean of the probability distribution of the image velocity vector v_k, is chosen to be zero in the initialisation step in the Kalman filter estimation algorithm, then the first estimate update step of the algorithm gives

$$v_1^1 = K_1 h_1 \tag{12}$$

Substituting for K_1 and h_1 in this equation yields

$$v_1^1 = \begin{bmatrix} I_x^2 + \frac{\sigma^2}{\sigma_p^2} & I_x I_y \\ I_x I_y & I_y^2 + \frac{\sigma^2}{\sigma_p^2} \end{bmatrix}^{-1} \begin{bmatrix} I_x \\ I_y \end{bmatrix} I_t \tag{13}$$

If the estimation process described above is generalised to the case where there are n observation equations for the unknown image velocity vector v at n spatial locations, i.e. in the definitions

$h_k \triangleq I_t(x, y, t + \Delta t)$
$C_k \triangleq \begin{bmatrix} I_x(x, y, t) & I_y(x, y, t) \end{bmatrix}$

h_k is now an n vector representing the local observations of the temporal derivatives of the image intensity at n spatial locations (x, y), and C_k is correspondingly an $n \times 2$ matrix of the local observations of the spatial derivatives of the image intensity at the n spatial locations (x, y), then the equation for the initial estimate of the velocity vector becomes

$$v_1^1 = \begin{bmatrix} \sum(I_x^2) + \frac{\sigma^2}{\sigma_p^2} & \sum(I_x I_y) \\ \sum(I_x I_y) & \sum(I_y^2) + \frac{\sigma^2}{\sigma_p^2} \end{bmatrix}^{-1} \begin{bmatrix} \sum(I_x I_t) \\ \sum(I_y I_t) \end{bmatrix} \tag{14}$$

in which the summations are over the local spatial and temporal derivatives of the image intensity at the n locations, and the measurement noise variance σ^2 is assumed to be the same for all n locations.

The above equation (14) is the same as that derived by Weiss *et al.* (2002) for the velocity estimate v^* of their "ideal observer" as the mean of the posterior distribution (equation (1) of Weiss *et al.*, 2002). Thus, if the initial estimate of the covariance matrix of the probability distribution of the velocity vector v_0 is given as $P_0^0 = \sigma_p^2 I_2$ in the initialisation step of the Kalman filter estimation algorithm then the algorithm will compute a velocity estimate v_1^1 in it first step which corresponds to the velocity estimate v^* of the "ideal observer" of Weiss *et al.* (2002).

However, in the model of motion integration proposed here, it is assumed that the estimation algorithm continues to use successive observations of the temporal and spatial changes in image intensity in order to recursively update the estimates of the velocity vector v_k^k and its covariance matrix P_k^k, as expressed in the Kalman filter algorithm. This implies that the Kalman filter algorithm will take some time to converge to an optimal (least squares) estimate of the mean and covariance of the image velocity vector v, based on this sequence of observations. The main proposal made in this thesis is that the dynamical behaviour of this recursive estimation process corresponds to the dynamics of the motion integration process which is seen experimentally in the initial period of motion perception both for smooth eye pursuit and plaid motion perception. It is shown in later chapters that the dynamics of the recursive estimation process reflects several key characteristics of the experimentally observed integration process dynamics.

It is clear however from the above analysis that the velocity estimate given by the first step of the Kalman filter algorithm coincides with the optimal velocity estimate of the ideal observer of Weiss *et al.* (2002). Therefore, if the "free parameter" (Weiss *et al.*, 2002) of the optimal estimate (14), the σ / σ_p ratio, is set in the algorithm to the same value as in the "ideal observer of Weiss *et al.*, our model of the motion integration process will also suffer from the fact that it will predict that a stimulus consisting of a horizontally moving high contrast "thin" rhombus will not result in any initial offset

bias in the pursuit eye motion, contrary to the experimental observations of Wallace *et al.* (2005). However, in the following section, where the computer simulations of the Kalman filter estimation algorithm based model are described, we show that for certain values of the σ / σ_p ratio, the high contrast "thin" rhombus will result in an offset bias.

In the next section, it is shown how the Kalman filter based model can be formulated so that it is amenable to implementation as a neural computation process. This results in an approximate, sub-optimal recursive estimation process which nevertheless retains its close relationship to the experimentally observed integration process. In fact, in the approximate, neural computation based form, simulations of the model show that the closeness of fit of the behaviour of the model to the experimental data improves over the optimal Kalman filter based form of the model. It is then described how this neural computation based model might be mapped onto the neural circuitry involved in the interaction between the V1 and MT areas of cortex, in a way that mimics the distributed, recurrent nature of the V1-MT circuitry. This is important since the V1-MT circuitry has been identified by many researchers as the location for the motion integration processing stage of the oculomotor system (Groh *et al.*, 1997; Simoncelli, 2003; Simoncelli and Heeger, 1998; Pack and Born, 2001; Pack *et al.*, 2003; Pack *et al.*, 2004)

2.2. A motion integration model based on an approximate version of the Kalman filter.

Computations in neural networks are generally assumed, in their simplest form, to be modelled by the weighted summation of a set of input signals, corresponding to the inner product

$$y = a \cdot x$$

where a is the vector of weights, x is the vector of inputs, and y is the scalar output of the computation. This inner product is then usually followed by some form of nonlinearity which ensures that the outputs of the network are always zero or positive, consistent with the non-negative firing rates of neurons in a biological neural network.

This simple model can be extended to the case where y is a vector of outputs

$$y = A \cdot x$$

where A is a matrix of weights, and moreover to the case where Y, A and X are all matrices, i.e. the matrix product equation

$$Y = A \cdot X$$

in which the ij th element of the matrix Y is computed from the i th row of A and the j th row of X, i.e. $y_{ij} = a_i \cdot x_j$.

All of the computations performed in the Kalman filter algorithm described in the previous section consist of such vector and matrix computations and therefore the algorithm can in principle be directly implemented in the form of a neural network computation, albeit a complex one. However there is one problem with this. Some of the matrices used in the Kalman filter algorithm are inverses of the matrices made up from the model parameters or measured variables. As far as the author is aware there is no way of computing the inverse of a matrix using an artificial neural network. Thus to make the algorithm "neurally plausible", by which is meant amenable to computation by an artificial neural network, it is necessary to avoid computation of matrix inverses.

It is clear from the description of the Kalman filter algorithm in the previous section, that a neural computational implementation of the model based on this algorithm would require the computation of an inverse matrix in order to calculate the Kalman gain matrix K_k (equation 9):

$$K_k = [(P_k^{k-1})^{-1} + C_k^T C_k \frac{1}{\sigma^2}]^{-1} C_k^T \frac{1}{\sigma^2}$$

As discussed above, such a calculation is not plausible as a neural computation, and therefore it is necessary to find a version of the algorithm which avoids this matrix inversion.

In the case of n local observations, each subject to a different, mutually independent measurement noise, with different variances σ_i^2, $i=1,...,n$, equation (9) becomes

$$K_k = [(P_k^{k-1})^{-1} + C_k{}^T \Sigma^{-1} C_k]^{-1} C_k{}^T \Sigma^{-1} \tag{15}$$

in which the measurement noise covariance matrix is denoted by $\Sigma \triangleq diag(\sigma_i^2)$, $i=1,...,n$. To avoid the two matrix inversions in (15), firstly the covariance matrix of the velocity estimate is approximated by a diagonal matrix, i.e.

$$P_k^{k-1} \triangleq \begin{bmatrix} (\sigma_{p,x}^2)_k^{k-1} & 0 \\ 0 & (\sigma_{p,y}^2)_k^{k-1} \end{bmatrix} \tag{16}$$

This assumes that the estimate of the velocity vector is uncorrelated in the x and y directions. Then the first matrix inverse term in (15) is given by

$$\left(P_k^{k-1}\right)^{-1} = \begin{bmatrix} 1/(\sigma_{p,x}^2)_k^{k-1} & 0 \\ 0 & 1/(\sigma_{p,y}^2)_k^{k-1} \end{bmatrix} \tag{17}$$

In order to avoid the second matrix inverse in (15), we must consider the term

$$C_k^T \Sigma^{-1} C_k = \begin{bmatrix} \sum_i \frac{1}{\sigma_i^2} I_{x,i}^2 & \sum_i \frac{1}{\sigma_i^2} I_{x,i} I_{y,i} \\ \sum_i \frac{1}{\sigma_i^2} I_{x,i} I_{y,i} & \sum_i \frac{1}{\sigma_i^2} I_{y,i}^2 \end{bmatrix} \tag{18}$$

It is assumed that this matrix can also be approximated by a diagonal matrix, i.e.

$$C_k^T \Sigma^{-1} C_k \approx \begin{bmatrix} \sum_i \frac{1}{\sigma_i^2} I_{x,i}^2 & 0 \\ 0 & \sum_i \frac{1}{\sigma_i^2} I_{y,i}^2 \end{bmatrix} \tag{19}$$

Thus the second matrix inversion in (15) becomes the inversion of a diagonal matrix, the inverse of which is given by

$$[(P_k^{k-1})^{-1} + C_k{}^T \Sigma^{-1} C_k]^{-1} = \begin{bmatrix} (\alpha_x)_k & 0 \\ 0 & (\alpha_y)_k \end{bmatrix} \triangleq M_k \tag{20}$$

where

$$(\alpha_x)_k \triangleq \frac{(\sigma_{p,x}^2)_k^{k-1}}{1 + (\sigma_{p,x}^2)_k^{k-1} \sum_i \frac{1}{\sigma_i^2} I_{x,i}^2} \tag{21}$$

$$(\alpha_y)_k \triangleq \frac{(\sigma_{p,y}^2)_k^{k-1}}{1 + (\sigma_{p,y}^2)_k^{k-1} \sum_i \frac{1}{\sigma_i^2} I_{y,i}^2} \tag{22}$$

With this approximation for the matrix inverse in (15), the Kalman gain matrix can be written in an approximate form which does not require any matrix inversion, i.e.

$$K_k = M_k \begin{bmatrix} \frac{I_{x,1}}{\sigma_1^2} & \cdots & \frac{I_{x,n}}{\sigma_n^2} \\ \frac{I_{y,1}}{\sigma_1^2} & \cdots & \frac{I_{y,n}}{\sigma_n^2} \end{bmatrix} \tag{23}$$

In order to update the diagonal elements of the approximate covariance matrix P_k^{k-1} defined in equation (16), it is noted that the *"at observation k"* update equation of the Kalman filter algorithm for the covariance matrix can be written as

$$\begin{aligned} P_k^k &= P_k^{k-1} - K_k C_k P_k^{k-1} \\ &= \begin{bmatrix} (\sigma_{p,x}^2)_k^{k-1} & 0 \\ 0 & (\sigma_{p,y}^2)_k^{k-1} \end{bmatrix} \\ &\quad - M_k \begin{bmatrix} \sum_i \frac{1}{\sigma_i^2} I_{x,i}^2 & \sum_i \frac{1}{\sigma_i^2} I_{x,i} I_{y,i} \\ \sum_i \frac{1}{\sigma_i^2} I_{x,i} I_{y,i} & \sum_i \frac{1}{\sigma_i^2} I_{y,i}^2 \end{bmatrix} \begin{bmatrix} (\sigma_{p,x}^2)_k^{k-1} & 0 \\ 0 & (\sigma_{p,y}^2)_k^{k-1} \end{bmatrix} \end{aligned} \tag{24}$$

Again setting the off-diagonal elements to zero results in the two independent update equations for the variances of the velocity estimate in each of the cardinal directions, i.e.

$$(\sigma_{p,x}^2)_k^k \approx \frac{(\sigma_{p,x}^2)_k^{k-1}}{1+(\sigma_{p,x}^2)_k^{k-1}\sum_i \frac{1}{\sigma_i^2} I_{x,i}^2} \tag{25}$$

$$(\sigma_{p,y}^2)_k^k \approx \frac{(\sigma_{p,y}^2)_k^{k-1}}{1+(\sigma_{p,y}^2)_k^{k-1}\sum_i \frac{1}{\sigma_i^2} I_{y,i}^2} \tag{26}$$

This approximate form of the algorithm produces a model of the motion integration process which avoids the need to compute a matrix inverse in the calculation of the Kalman gain matrix, and is therefore implementable as a neural computation.

It should be noted here that the temporal and spatial derivatives of the intensity, $I_t(x,y,t)$, $I_x(x,y,t)$ and $I_y(x,y,t)$, appear in the algorithm in the form of either squared terms: I_t^2, I_x^2, I_y^2, or the products: $I_x I_t, I_y I_t, I_x I_y$. As observed in Heeger and Simoncelli (1993), there are no known cells in V1 with receptive fields which behave as products of derivatives, i.e. $I_x I_t, I_y I_t, I_x I_y$. However such products can be expressed in the form

$$\begin{aligned} I_x I_y &= \frac{1}{4}\left\{(I_x + I_y)^2 - (I_x - I_y)^2\right\} \\ I_x I_t &= \frac{1}{4}\left\{(I_x + I_t)^2 - (I_x - I_t)^2\right\} \\ I_y I_t &= \frac{1}{4}\left\{(I_y + I_t)^2 - (I_y - I_t)^2\right\} \end{aligned} \tag{27}$$

and therefore all of the required functions of intensity derivatives in the model can be expressed as the outputs of squared linear filters of the image intensity (Heeger and Simoncelli, 1993). This both simplifies the implementation of the model in a neural computational form, and allows the possibility of mapping it physiologically onto the V1-MT neural circuitry.

Whilst the above approximations make the model amenable to implementation in the form of a neural network computation, the algorithm is now further simplified to make it suitable for implementation in a *distributed, recurrent* neural computational form. Instead of continuously updating the variances of the velocity estimate as just described, the diagonal elements of the prior covariance P_0 in equations (21) and (22) are now used to calculate the elements of the M_k matrix at every *"update at observation* k*"* step of the algorithm, i.e.

$$(\alpha_x)_k = \frac{(\sigma^2_{p,x})^0_1}{1+(\sigma^2_{p,x})^0_1 \sum_i \frac{I^2_{x,i}}{\sigma^2_i}} \tag{28}$$

$$(\alpha_y)_k = \frac{(\sigma^2_{p,y})^0_1}{1+(\sigma^2_{p,y})^0_1 \sum_i \frac{I^2_{y,i}}{\sigma^2_i}} \tag{29}$$

in which, from the initial *"update between observations"* step of the algorithm, $(\sigma^2_{p,x})^0_1 = (\sigma^2_{p,x})^0_0$ and $(\sigma^2_{p,y})^0_1 = (\sigma^2_{p,y})^0_0$, the diagonal elements of P_0. This of course affects the calculation of the Kalman gain matrix using equation (23). But it avoids the need to update the elements of the covariance matrix as described in equations (24)-(26) .

In summary, the approximate form of the Kalman filter estimation algorithm which forms the basis of the proposed "neurally-plausible" implementation of the motion integration model is described as follows:

1. initialisation:

$v^0_0 = \overline{v_0}$

$P^0_0 = P_0 = diag\left[(\sigma^2_{p,x})^0_0, (\sigma^2_{p,y})^0_0\right]$

2. update of the velocity estimate between observations

(a) before the first observation

$$v_1^0 = v_0^0$$

$$(\sigma_{p,x}^2)_1^0 = (\sigma_{p,x}^2)_0^0$$
$$(\sigma_{p,y}^2)_1^0 = (\sigma_{p,y}^2)_0^0$$

(b) between subsequent observations

$$v_{k+1}^k = v_k^k$$

3. update of the velocity estimate at observation *k*

$$v_k^k = v_k^{k-1} + K_k(h_k - C_k v_k^{k-1})$$

where

$$K_k = M_k \begin{bmatrix} \frac{I_{x,1}}{\sigma_1^2} & \cdots & \frac{I_{x,n}}{\sigma_n^2} \\ \frac{I_{y,1}}{\sigma_1^2} & \cdots & \frac{I_{y,n}}{\sigma_n^2} \end{bmatrix}, \quad M_k \triangleq \begin{bmatrix} (\alpha_x)_k & 0 \\ 0 & (\alpha_y)_k \end{bmatrix}$$

$$(\alpha_x)_k = \frac{(\sigma_{p,x}^2)_1^0}{1 + (\sigma_{p,x}^2)_1^0 \sum_i \frac{I_{x,i}^2}{\sigma_i^2}}$$

$$(\alpha_y)_k = \frac{(\sigma_{p,y}^2)_1^0}{1 + (\sigma_{p,y}^2)_1^0 \sum_i \frac{I_{y,i}^2}{\sigma_i^2}}$$

The elements $(\alpha_x)_k^{k-1}$ and $(\alpha_y)_k^{k-1}$ of the M_k matrix might be thought of as *normalised forms* of the prior variances $(\sigma_{p,x}^2)_0^0$ and $(\sigma_{p,y}^2)_0^0$ of the velocity estimate.

2.3. A distributed, recurrent version of the model based on the approximate Kalman filter.

In this section it is shown that the "neurally-plausible" implementation of the motion integration model described in the previous section, leads to a distributed, recurrent neural computational version of the motion integration model.

In this version of the model, local observations of the spatial and temporal derivatives of the image intensity are made at a distributed set of spatial locations in the visual space, and are used together with the current global velocity estimate to form a corresponding set of local updates to the current estimate. The local update signals are then summed, and used to calculate a new global estimate of the velocity. The new global velocity estimate is then broadcast to each spatial location where it is again used to calculate a local update signal to the global estimate.

To describe in more detail this form of the model, we first define a local *"update"* signal for each spatially local observation window as

$$(e_k^{k-1})_i \triangleq \begin{bmatrix} \frac{I_{x,i}I_{t,i}}{\sigma_i^2} \\ \frac{I_{y,i}I_{t,i}}{\sigma_i^2} \end{bmatrix} - \begin{bmatrix} \frac{I_{x,i}^2}{\sigma_i^2} & \frac{I_{x,i}I_{y,i}}{\sigma_i^2} \\ \frac{I_{x,i}I_{y,i}}{\sigma_i^2} & \frac{I_{y,i}^2}{\sigma_i^2} \end{bmatrix} v_k^{k-1} \tag{30}$$

It then follows that the *"at observation k"* update equation for the velocity estimate in the approximate Kalman filter algorithm

$$v_k^k = v_k^{k-1} + K_k(h_k - C_k v_k^{k-1}) \tag{31}$$

where

$$K_k = M_k \begin{bmatrix} \frac{I_{x,1}}{\sigma_1^2} & \cdots & \frac{I_{x,n}}{\sigma_n^2} \\ \frac{I_{y,1}}{\sigma_1^2} & \cdots & \frac{I_{y,n}}{\sigma_n^2} \end{bmatrix}, \quad M_k \triangleq \begin{bmatrix} (\alpha_x)_k & 0 \\ 0 & (\alpha_y)_k \end{bmatrix}$$

can be written in terms of the local update signals as

$$v_k^k = v_k^{k-1} + M_k \sum_i (e_k^{k-1})_i \tag{32}$$

Thus the global update of the velocity estimate is expressed as a summation of local estimate update signals.

This results in a distributed version of the motion integration model, wherein: (i) for each spatially distributed observation window a local update signal $(e_k^{k-1})_i$ is computed based on the squared local temporal and spatial derivatives $I^2_{x,i}, I^2_{y,i}, I^2_{t,i}$, the variance of the local measurement noise σ_i^2, and the current velocity estimate v_k^{k-1} (equation (30)) ; (ii) the local update signals are summed and used to create a new velocity estimate v_k^k (equation(32)).

Clearly, the behaviour of this distributed, recurrent form of the motion integration model will be identical to that of the motion integration model based on the approximate Kalman filter estimation algorithm, as described in Section 2.2. In Figure 9 a neural network is described in diagrammatic form which is capable, in principle, of implementing the distributed, recurrent version of the motion integration model.

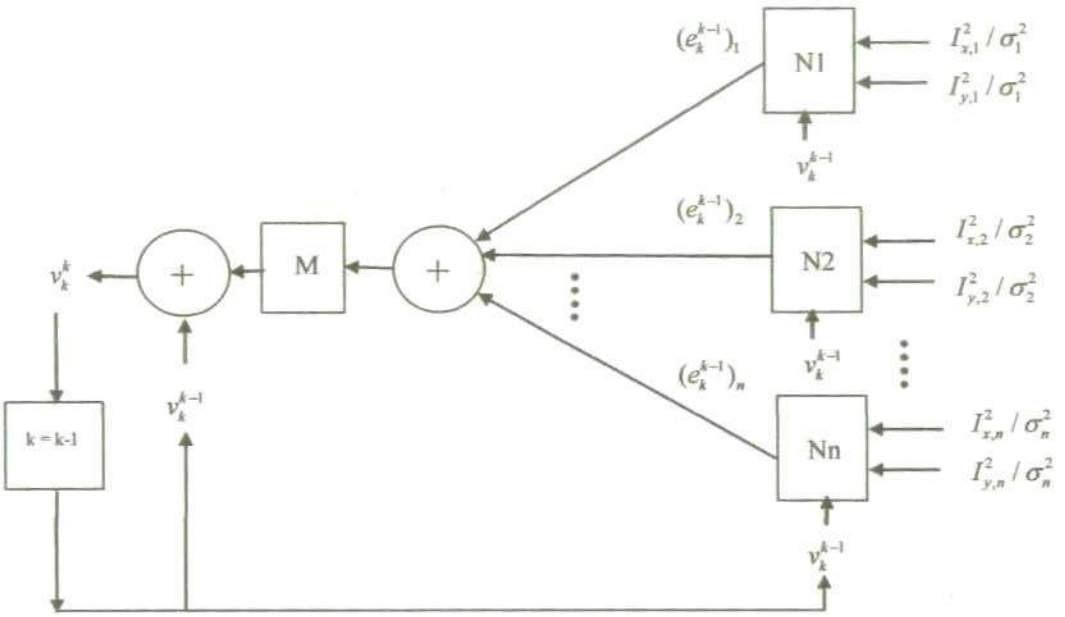


Figure 9. Neural network implementation of the distributed, recurrent form of the motion integration model. The set of distributed networks Ni each implement equation (30), with v_k^{k-1} entering each network as a modulating term. The remainder of the network implements equation (32), with the weight matrix of M corresponding to the matrix M_k defined in equation (20). The recurrent feedback path represents the update of the velocity estimate at the next step of the algorithm.

2.4. Summary.

In this chapter the mathematical form of the proposed Kalman filter based motion integration model was presented. The relationship of the change of the intensity of the image with time and the velocity was used as a basic equation. After that the Taylor series approximation was applied in order to derive the equation relating the x and y components of the velocity vector, which are unknown and the temporal and spatial derivatives of the intensity (I_t, I_x and I_y) which are observed, plus added Gaussian noise. The posterior probability of the velocity was expressed by the Bayes' rule which combines the likelihood probability function and the prior. Then the

maximum of that posterior distribution was obtained using a Kalman filter estimation algorithm, in which the estimation of the current velocity takes into account the information from the previous steps of estimation.

A version of the model based on an approximate version of the Kalman filter algorithm was also presented in this chapter. There were three main changes as compared to the optimal Kalman filter algorithm. First, to avoid the matrix inversion when calculating the Kalman gain, the covariance matrix of the velocity estimate was approximated by a diagonal matrix. This assumes that the components v_x and v_y are independent. The second change involved the assumption that the spatial intensity derivatives on the x and y axes were independent as well. This assumption simplifies the matrix which contains the local measurements of the intensity derivatives. Both approximations led to a simpler calculation of the Kalman gain, avoiding the need for computing matrix inversions. The third main change was related to the prior covariance P_0, which initial value was kept constant for every update of the M_k matrix. This assumption simplifies the algorithm and allows it to be represented in a distributed, recurrent neural computational form, which was represented in a diagrammatic form in Figure 9.

Chapter 3. Motion integration in smooth eye pursuit.

3.1. Smooth eye pursuit.

Smooth eye pursuit is a good experimental paradigm for studying the motion integration process by taking into account the latency and accuracy of the eyes response. Especially, the initiation of pursuit after the onset of the target seems most informative for studying the integration mechanisms (Lisberger and Westbrook, 1985). First, there is a latency of the eye response till the change of the target position accumulates to a certain extent. This change or error which needs to be compensated for accumulates to some degree and then it triggers the compensatory eye movements. After the eye movements start it is shown experimentally that during this motion the information about the change of the position of the target is not upgraded. It is an "open-loop" system. It takes another time delay to have the new information about motion change reflected in the eye movements. In other words, if an unpredictable or sharp change in position occurs during the latency period the eye movements are not able to respond immediately to it.

From the time of onset motion of the visual target stimulus which is to be tracked, the time delay is of about 100 ms before the visually driven eye pursuit motor response begins. As already mentioned previously, some of the motion signals may not accurately reflect the target motion, owing to the aperture effect, as is the case with edges moving in the non-orthogonal direction to themselves. Experimental evidence from human and non-human primates indicates that, as a result, the initial eye pursuit movement has both an on-axis component, i.e. in the direction of the object motion, and an off-axis bias which reflects the inaccuracies in the local motion signals (Masson & Stone, 2002; Pack and Born, 2001).

The experimental evidence suggests that the integration process begins as soon after the onset of the target motion as a visual motion signal is available, i.e. at the latency of transfer of the retinal motion signal to the site of the neural integration process. As mentioned earlier, there is strong evidence that visual area MT could be the main one for the motion integration process (Rodman and Albright, 1989; Stoner and Albright, 1992; Movshon and Newsome, 1996). Physiological experiments (Pack and Born, 2001) suggest that this latency is about 70-80 ms for the cells in MT area of alert monkeys to respond. At this point in time the eye pursuit has yet to start owing to its ~100ms latency, so the error between the target motion and the eye motion increases without compensation from any eye movement (Lisberger and Westbrook, 1985). However, the experimental data of Wallace *et al.* (2005) and of Born *et al.* (2006), suggests that the integration process is already starting to correct for the inaccurate off-axis motion signals, and this can be observed in the eye movement within ~70 ms after it begins. This is before the error signal induced by the inaccurate off-axis eye movement can be compensated for by the oculomotor feedback system, owing to the ~100 ms delay in this system. Their results could be interpreted as follows: for the first 70 ms or so, the eye movements are driven by a simple uncorrected pooling of both the unambiguous (2-D) and ambiguous (1-D) local motion measurements, with the 2-D measurements of the target object velocity only starting to dominate after this period. As a result, the correction of the off-axis direction error of the eye movements starts approximately after 60-80 ms from the onset of the target, but before the end of the 100 ms latency of the open-loop The bias error in the eye movement then decays to zero over a further period of 200 - 300 ms (see Figure 3 of Wallace *et al.*, 2005). This temporal evolution toward an accurate representation of target object motion is consistent with several experimental results: from human direction judgments (Lorenceau *et al.*, 1993); the human ocular following response (Masson and Castet,

2002; Masson *et al.*, 2000), smooth pursuit in monkeys (Pack and Born, 2001) and humans (Lindner and Ilg, 2000; Masson and Stone, 2002, Wallace *et al.*, 2005), and the responses of MT neurons in alert monkeys (Pack and Born, 2001; Pack *et al.*, 2004).

3.2. Bayesian estimation models applied to smooth eye pursuit.

Weiss *et al.* (2002) proposed an optimal Bayesian probabilistic model of motion integration in which an optimal estimate of the target object velocity is computed using the likelihoods of the local velocity measurements distributed over the visual space (Simoncelli *et al.*, 1991; Simoncelli, 2003). They used a priori distribution for the target object velocity which assumes a human preference for the assumption of slow speeds, and formulated this as a Gaussian prior centred on zero. The idea of using a slow speed preference was initially adopted by Heeger and Simoncelli (1991) in their motion itegration model. The prior probability distribution centred at zero is analogous to the assumption that the aperture problem is solved by preference to the slowest speed, i.e. the velocity orthogonal to the edge, as discussed in Section 1.5. Support for this idea comes from physiology where it is well known that the preference of the 1-D detectors of direction is toward the orthogonal to the edge direction of motion (Hubel and Weisel, 1959). Weiss et al. (2002) also assumed an additive measurement noise which was independently Gaussian distributed, with zero mean and known variance. This had the effect of making the local likelihood functions dependent on stimulus contrast. They computed the posterior distribution of the velocity as the product of the likelihoods over all the spatial locations (assuming that the likelihoods are independent) multiplied by the prior. The authors show that their model reflects several of the main characteristics of human motion perception, as observed in a range of psychophysical studies, both their own experiments which use "fat" and "thin" rhombus figures at high and low

contrast, and those of others (Burke and Wenderoth, 1993; Bowns, 1996; Stone *et al.*, 1990; Stone and Thompson, 1990; Lorenceau *et al.*, 1993) using mainly moving plaid patterns. For fixed values of measurement noise variance and prior variance (or more specifically, a fixed ratio of these quantities), and for a moving "thin" rhombus stimulus, at low contrast the posterior velocity distribution has a maximum (mean) close to the vector average of the local velocities. At high contrast, this maximum occurs at, or near, the veridical target object velocity given the intersection of the constraint lines provided by the local velocity measurements.

The model of Weiss *et al.* (2002) was not intended to replicate the experimental data from the smooth eye pursuit experiments cited above, and is not capable of reproducing the dynamics of the motion integration process as observed in these experiments. Moreover, their model, which is tuned to replicate their human perceptual experiments, would predict that a stimulus consisting of a horizontally moving high contrast "thin" rhombus would not result in any offset bias in the pursuit eye motion, whereas the experimental data of Wallace *et al.* (2005) indicates that there is always an offset bias in the initial eye motion of ~30° even in the case of a "thin" rhombus of high contrast and slow speed.

3.3. Non-Bayesian models applied to smooth eye pursuit.

The neural process of integration of the local motion signals is generally supposed to involve two stages. The first stage, which is usually attributed to neural mechanisms in primary visual cortex (V1), involves the extraction of directionally selective motion information. Since V1 neurons are subject to the ambiguities introduced by 1-D motion stimulus signals, a second stage, attributed to the medial temporal area of cortex (MT), is perceived to be the location of the neural mechanisms for integration of the local motion signals and the resolution of the ambiguities introduced by the 1-D signals. One

model based on this two stage process supposes a feed-forward mechanism which achieves the integration process by differential weighting of the feed-forward projections of the 1-D and 2-D signals (Simoncelli and Heeger, 1998). However, this model would not account for the temporal dynamics of the motion integration process observed in the above cited smooth pursuit experimental data.

An alternative model (Pack *et al.*, 2003) suggests that the dynamics of the integration process might be determined by the properties of visual neurons early in the visual pathway, in particular a temporal delay in the emergence of end-stopping in direction-selective neurons. In this formulation, the temporal evolution of 2-D motion signals for pursuit and those that have been observed in MT neurons (Pack and Born 2001; Pack *et al.*, 2004) reflect a change in the "weighting" applied to the outputs of V1 direction-selective neurons, whereby the activity of 2-D related end-stopping neurons eventually suppresses that of the 1-D contour-related neurons, and the motion signals from the 2-D terminators becomes dominant. It is suggested that this mechanism might also explain the perceptual dominance of a contour-based vector average for stimuli of *low contrast* (Weiss *et al.*, 2002), since end-stopping is weak or absent for such stimuli (Polat *et al.*, 1998; Sceniak *et al.*, 1999). This is also consistent with the experimental observations of Wallace *et al.* (2005) on pursuit initiation in humans, in which the effect of lowering the stimulus contrast is an increase in the off-axis bias in the initial transient eye movement and a lengthening of the time taken for this bias to reduce to zero. Note however that for the low contrast stimulus, the off-axis motion is eventually eliminated, albeit with a time constant up to nearly three times that for the high contrast stimulus, indicating that the suppressive mechanism, if this is indeed the case, does still operate although more weakly. As pointed out in Born *et al.* (2006), the suggested role of the end-stopping cells is also physiologically very plausible given the fact that neurons in layer 4B of V1, which is the source of the main projection from V1 to MT (Maunsell

and Van Essen, 1983; Shipp and Zeki, 1989), are strongly end-stopped (Sceniak *et al.*, 2001).

3.4. Other Bayesian motion integration models.

Other Bayesian models of motion integration and estimation have also been previously proposed (Koechlin *et al.*, 1999; Rao, 2004) but have not been applied to motion integration in smooth eye pursuit.

Koechlin *et al.*(1999) focus mainly on the segmentation mechanism of a population of neurons, where the Bayesian inference is derived from the lateral connectivity and is used to modify the feedforward signals. For instance, in Koechlin *et al.* (1999) the experiment with a tilted moving bar requires two separate sets of units (neurons); one of them detecting the orthogonal motion, and another one encoding the veridical bar-end direction. Based on the lateral connectivity and Bayesian inference, the signal from the units detecting the bar-end velocity propagates along the contour of the bar and with the time suppresses the signal from the non-veridical orthogonal velocity detectors.

This mechanism differs substantially from the model presented here. In the model presented here a pooling of the signals is performed on a higher level (supposedly MT area), where the information from the detectors signaling non-veridical motion orthogonal to the edges, are overpowered with time by the non-ambiguous, veridical information from the detectors of the motion of the ends of the bar, as a result of the recursive nature of the velocity estimation algorithm.

The work of Rao (2004) is primarily concerned with the problem of implementing the Bayesian approach in a recurrent neural network. The examples given related to motion perception are concerned with Bayesian decision making in a visual

motion detection and discrimination task, rather than the problem of motion velocity estimation based on the integration of local motion signals as addressed in this thesis.

The model presented in this thesis differs substantially from these other models which employ a Bayesian approach. In contrast to the Koechlin *et al.*(1999) model, the model presented here does not use Bayesian inference based on propagation of the 2-D motion signals from the end-points of the bars via the lateral neural connectivity to allow, after a propagation delay, these signals to override the 1-D edge motion signals, thus implementing the dynamics of motion integration. Instead, the model uses the recursive nature of the Kalman filter estimation algorithm to directly model the motion integration dynamics. In principle, this can then be thought of in biological neural network terms as implementing the dynamics via the recurrent connections between V1 and MT rather than the lateral connections in V1. That is, the model can be envisaged in the form where the prior global stimulus velocity estimate is formed initially in MT and distributed to local velocity detectors in V1. These local detectors then use the current global estimate to form local error signals which are projected to MT, where they are used to update the global velocity estimate, which is in turn fed back to V1 local detectors in a recurrent process. This recurrent, distributed version of the motion integration model was fully described in Section 2.3 and illustrated in Figure 9.

The model presented here also differs from the recurrent neural network implementation of the Bayesian process described by Rao (2004) which, as already noted, does not address the motion integration process at all. Rather, Rao describes how Bayesian computation in general can be carried out in a recurrent neural circuit, and how the feedforward and recurrent connections may be selected to perform Bayesian inference for arbitrary hidden Markov models. He illustrates the approach using two tasks: discriminating the orientation of a noisy visual stimulus and detecting the direction of motion of moving stimuli.

3.5. Summary.

In this chapter a review of motion integration in smooth eye pursuit was presented. It was shown that the aperture effect from the detectors of direction of motion of the edges could lead to a significant bias in the eye movement's direction during the pursuit. Instead of moving into the veridical direction of motion, the eyes tend to follow the direction orthogonal to the edge of the target stimulus. The first phase of the pursuit, during the open-loop seems to be critical, as then the error is the strongest and it is shown that during that time the eye movements could not be corrected or influenced from any external change. On the contrary, as it has been shown in Wallace *et al.*.(2005), an internal correction occurs, which takes into account the influence of the veridical information from the 1-D cues over 2-D cues of the image.

Also in this chapter both Bayesian and non-Bayesian models of motion integration were reviewed. Some of these models have been applied to the smooth eye pursuit experimental data, but are not intended or able to capture the dynamics of the integration process evident in this data.

Chapter 4. Simulation of smooth eye pursuit experiments using the Kalman filter based motion integration model.

4.1. Simulations using the optimal Kalman filter version of the model.

In this chapter it will be shown that the Kalman filter based motion integration model described in Chapter 2 captures qualitatively the temporal dynamics of the integration process in smooth eye pursuit. As has been described, the model is based on the concept that the integration process computes a Bayesian estimate of the target object velocity using the local observations of target motion provided by the directionally selective neurons in V1, the estimation process being carried out recursively, in the form of a Kalman filter (Kalman, 1960). In this chapter it is shown that the dynamics of this recursive estimation process closely replicates the dynamics of the motion integration process, as measured experimentally, under a variety of conditions involving changes in shape and contrast of the stimulus.

A computational version of the motion integration model was constructed using Matlab (version 7.3.0.267, The MathWorks) and computer simulations of the model were performed on stimuli similar to the ones used in psychophysical experiments (Wallace *et al.*, 2005). Two visual stimuli were used in the computer simulation experiments: (i) a square diamond with main axes at 0° and 90°, and (ii) an elongated tilted diamond (or "thin rhombus") with the main axis at 45° and internal angles of 10° and 170°. The stimuli were presented as solid figures on a visual space consisted of 200 by 200 pixels. Each of the stimuli was contained within a square area of 50 by 50 pixels.

Figure 10 shows an example of the tilted rhombus as used in the simulations.

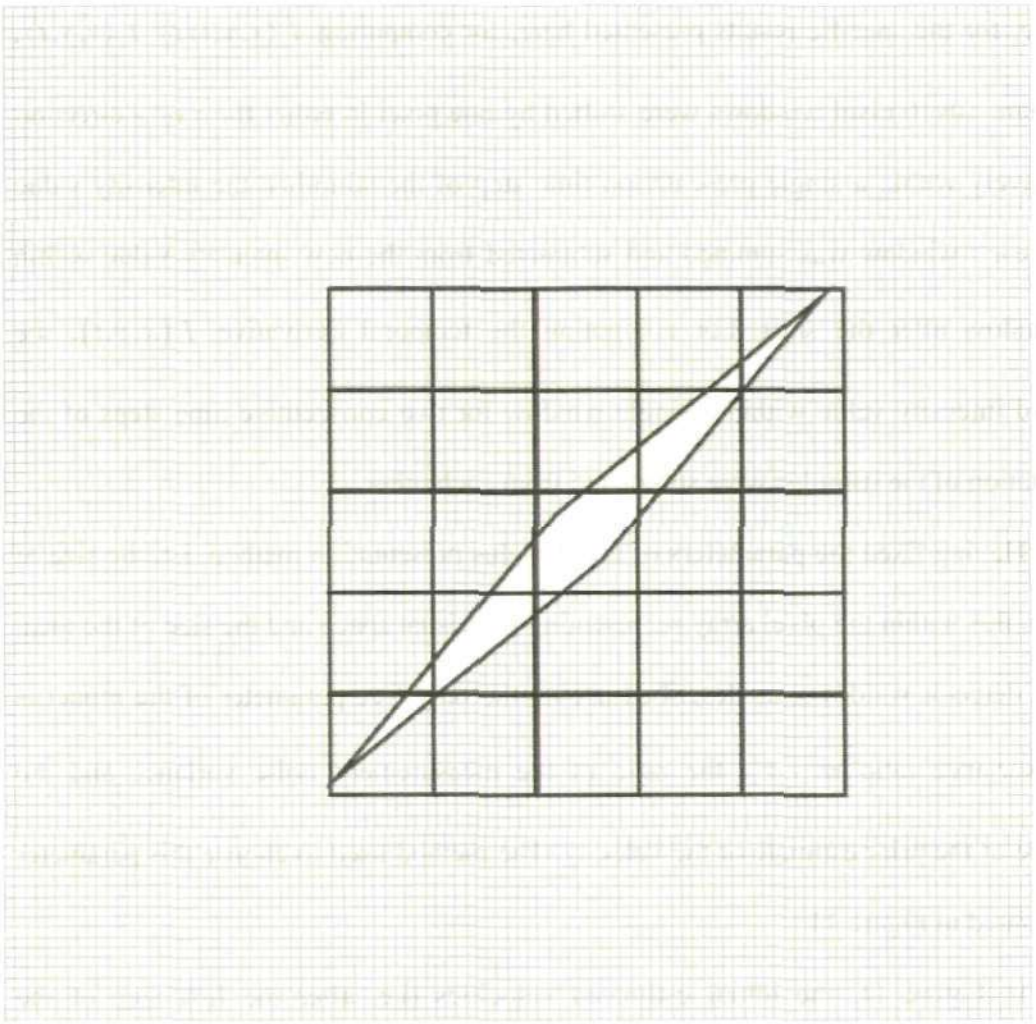


Figure 10. An example of a tilted rhombus stimulus as used in the simulations. Illustrated is a 100x100 pixel visual space (a 200x200 pixel visual space was actually used). The solid stimulus was confined within an area of 50x50 pixels contained in this background array. The square areas of 10x10 pixels covering the stimulus show the relative size of the individual local motion detectors used for calculating the spatial and temporal derivatives.

The visual space was further divided into 400 uniformly spaced and sized, non-overlapping square windows each of size 10 by 10 pixels. Thus the stimulus itself was contained within an area covered by 5x5 windows. In each window, the observations of temporal and spatial derivatives of image intensity were calculated as the image moves at a constant velocity in the horizontal direction from the left to the right of the visual space. Temporal and spatial derivatives $I_t(x,y,t)$, $I_x(x,y,t)$and $I_y(x,y,t)$ were

computed using small spatial and temporal shifts $\Delta x, \Delta y,$ and Δt of the stimulus. In practice, for the specific results presented here, for computing $I_x(x,y,t)$ or $I_y(x,y,t)$, the square 10x10 pixel windows were shifted by one pixel in either the x or y direction respectively within a single presentation time step of the stimulus; the intensity value within each window was summed and subtracted from the new intensity value within the window after the shift. For computing the temporal derivative $I_t(x,y,t)$, the summed intensity value within a single window for two consecutive time steps of the presentation of the stimulus were subtracted from each other.

Having fixed the parameters of the stimulus presentation method, i.e. the size of the stimulus within the pixel array representing the visual space and the size of the local motion detector windows, as described above, the only free parameter of the model is the covariance ratio σ^2 / σ_p^2, the ratio of the measurement noise variance and the variance of the prior estimate of the velocity. The method used to choose this parameter is illustrated in Figure 11.

In Figure 11, the initial estimated velocities (i.e. after the first step of the algorithm) in the x and y directions, v_x, v_y, for the tilted rhombus are shown for a wide range of values of σ^2 / σ_p^2, and for three levels of contrast of the stimulus: 0.25, 0.5 and 1.0. The values of the standard deviation of the prior velocity estimate σ_p were kept within a small range of values between 0.04 and 0.07, and the measurement noise standard deviation σ was varied between 0.01 and 3.

From Figure 11c in particular it can be seen that the values of the log covariance ratio σ^2 / σ_p^2 systematically determine the dependence of the estimated direction bias on contrast. When the logarithm of the σ^2 / σ_p^2 ratio is greater than 5.0 or less than -1.0 the variation of the estimated direction bias with contrast becomes considerably reduced compared to say a log covariance ratio of 2.0, for which the variation in bias with

contrast is greatest. It was therefore decided to choose a range of values of σ^2/σ_p^2 which covered this range of contrast dependence, in order to explore the full range in the subsequent simulations of the dynamics of the direction bias.

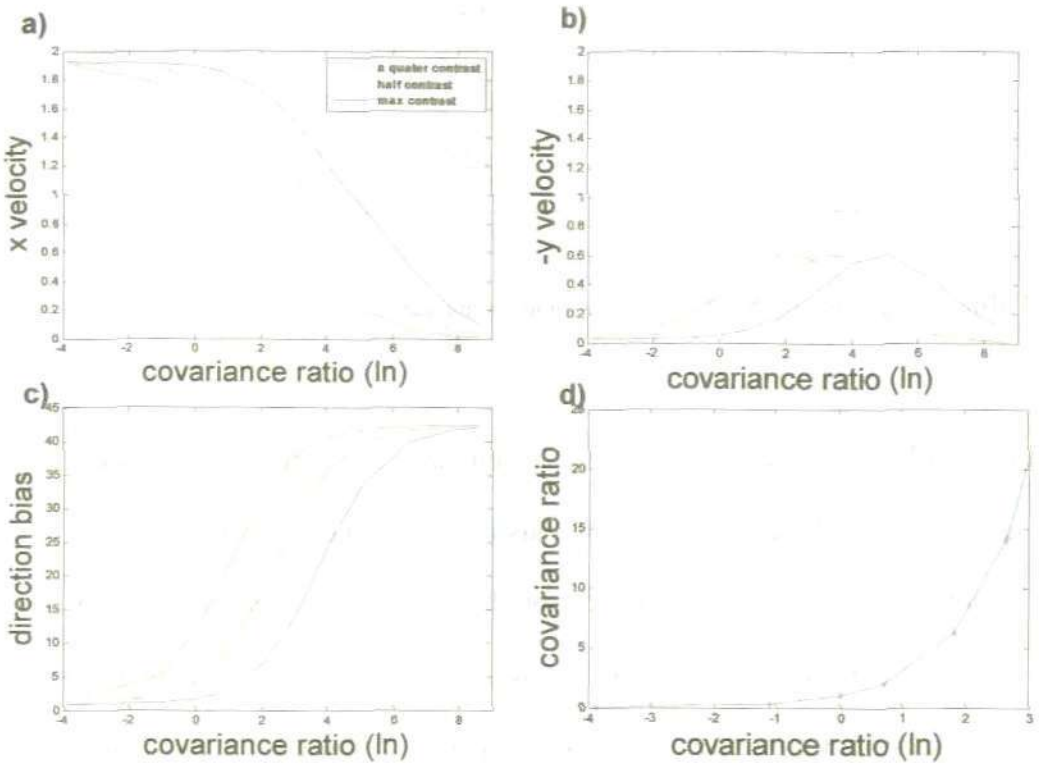


Figure 11. a,b) The initial estimates of the x and y velocity components of the tilted rhombus for 12 different values of the covariance ratio σ^2/σ_p^2 presented in a logarithmic scale. The true speed/direction of the rhombus is 2 pixels per time step in horizontal (x) direction. c) The initial estimate of the non-veridical direction bias, computed from the data in a) and b). d) Values of the covariance ratio σ^2/σ_p^2 plotted against their logarithmic representation.

The values of σ, σ_p, σ/σ_p, and $\ln(\sigma^2/\sigma_p{}^2)$ chosen to span this range in the simulations are given in Table 1 below. Thus $\ln(\sigma^2/\sigma_p{}^2)$ varies between ~ -1.0 and 5.0.

σ	σ_p	$\frac{\sigma}{\sigma_p}$	$\ln\left(\frac{\sigma^2}{\sigma_p^2}\right)$
0.04	0.07	0.6	-1.1
0.1	0.04	2.5	1.8
0.2	0.04	5	3.2
0.5	0.04	12.5	5.05

Table 1. The values of σ / σ_p used in the model simulations

Figure 12 shows the results from the computer simulation of the model in estimating the angular direction of the rhombus (0° = horizontal motion) of both the square diamond and thin rhombus stimuli, for three values of contrast: max (image intensity = 1), half (intensity = 0.5) and quarter (intensity = 0.25), and for the four different values of the ratio σ / σ_p given in Table 1, as a function of the number of iterations of the Kalman filter algorithm.

Note that here and for all simulations presented in the thesis, in displaying the temporal dynamics, time is represented in terms of number of iterations, i.e. steps, of the Kalman filter algorithm which can not be quantitatively related to any real time measurements, for instance milliseconds. Therefore the comparison of the simulation results with the experimental results must be made qualitatively and in a relative manner with respect to the variation in simulation parameters, e.g. contrast levels, rather than quantitatively, ie in terms of real values of direction bias or actual decay times., This is analagous to the manner in which the psychophysical results are often compared within a specific experimental setting.

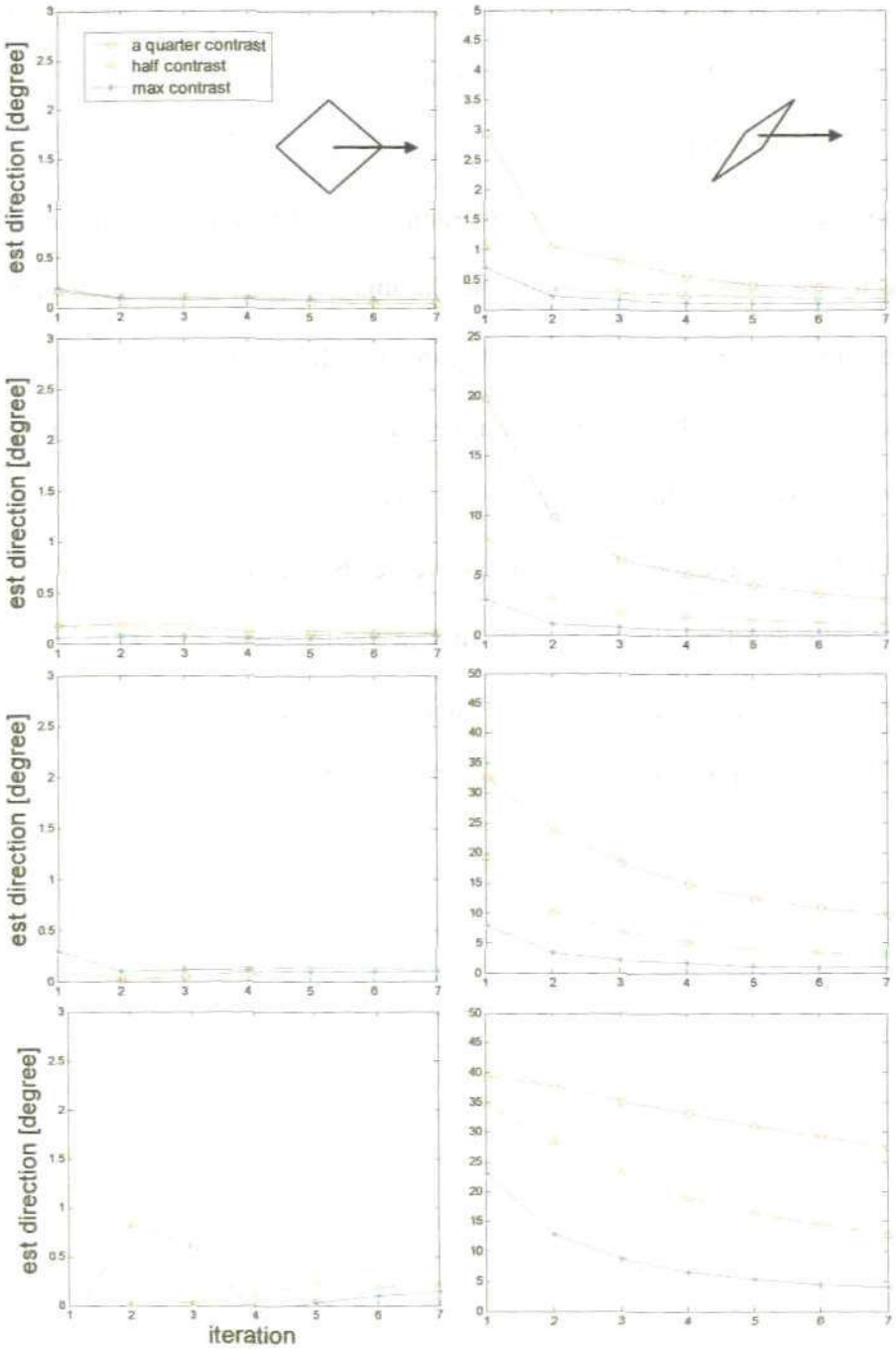


Figure 12 . Results from the computer simulations of the motion integration model based on the optimal Kalman filter algorithm, showing the estimated angular direction of the stimulus motion (0° = horizontal motion), for horizontally moving diamond and rhombus stimuli, for three different values of contrast, and for four different values of the ratio σ / σ_p increasing from the top row to the bottom row of plots. Note

the change in the vertical scale in the plots in the first two rows. The values of σ / σ_p used are given in Table 1.

Each iteration shown in the graphs in Figure 12 corresponds to five iterations/steps of the Kalman filter algorithm, and the estimated angular direction shown is the average of the estimated direction over these five iterations (thus in each plot the total number of iterations of the algorithm is thirty five).

The averaging was necessary because of sampling effects appearing as a result of the approximate and coarse way of calculating the image intensity derivatives usinf the 10x10 pixel non-overlapping windows. Figure 13 shows an example of the estimated component velocities before the averaging process, in the worst case situation of the approximate version of the algorithm (see following section), illustrating the widest observed range of variation of the estimate values between the individual steps of the Kalman filter algorithm. It was found that averaging over five steps of the algorithm was sufficient in all cases to provide a smooth variation in velocity estimates and hence in the estimates of direction, as shown in Figure 12.

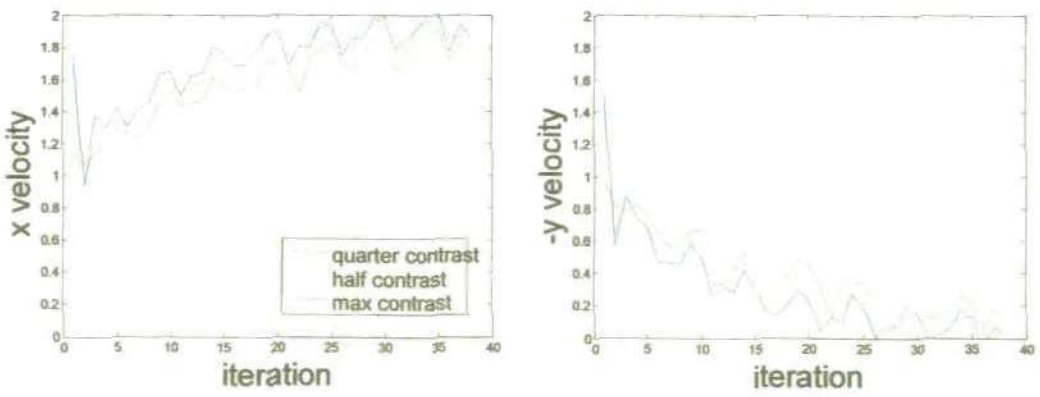


Figure 13. An example for the variation of the estimate component velocity values between the individual steps of the Kalman filter algorithm for the worst case of the approximate version of the Kalman filter algorithm, and for a log covariance ratio $\ln(\sigma^2/\sigma_p^2) = 3.2$ before the averaging over every 5 steps.

The simulation results, presented in Figure 12 have a small, but non-zero peak offset bias for all values of σ / σ_p (note the change in the vertical scale in the plots in the first two rows). The value of σ / σ_p is increased from the top to the bottom row in Figure 12, and results in a corresponding increase in the offset bias in the estimated direction. This offset bias is however very small in the case of the lowest value of the σ / σ_p ratio (top row of plots) for both the diamond and thin rhombus shapes (< 0.25° for the diamond and < 3° for the thin rhombus). In each case, the peak offset bias is followed by a temporal decay of this bias to an asymptotic value of close to 0°.

These characteristics of the model dynamics, as shown in Figure 12, i.e. an initial offset bias followed by a regular temporal decay of this bias to zero, are also the main characteristics of the dynamics of the motion integration process, as observed experimentally by Wallace *et al.* (2005). Their data show (see their Figures 5 and 6, which are presented here as Figures 14 and 15, respectively) a peak tracking direction error followed by a temporal decay toward an asymptotic value corresponding to the true target motion direction (0°).

Moreover, the model simulations show (Figure 12) that decreasing the contrast of the stimulus, for any fixed value of the ratio σ / σ_p, results in an increase in both the peak offset bias in the estimated velocity and the time constant for the decay of the bias to zero. This is again in close qualitative agreement with the data of Wallace *et al.* (2005), (here shown as Figure 14) in which the initial tracking bias was reduced from ~44° to ~30° when the contrast was increased from 10% to 40%, after which there was little further reduction. Similarly, increasing the target contrast resulted in a decrease in the decay time constant from ~168 ms at 10% contrast to an asymptotic value ~60 ms for contrasts >40%.

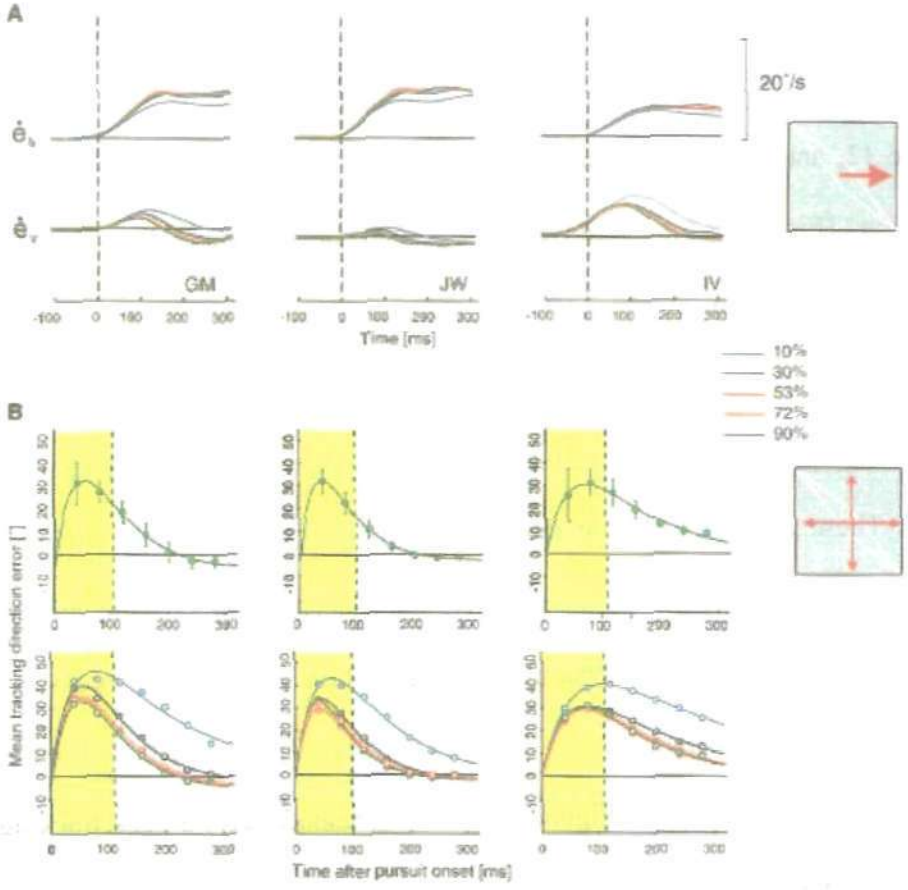


Figure 14. (FIG. 5 from **Wallace *et al.*, 2005** with the original legend).. Smooth pursuit responses to a 10°/s object motion: effects of object contrast. Layout is the same as for Fig. 2. *A*: mean velocity profiles of pursuit eye movements driven by a type II diamond drifting rightward for a range of object contrasts. For this experiment, the background was set to a gray level corresponding to the mean luminance level. *B*: mean tracking direction errors as a function of time for a single contrast condition (90%, *top*) and for a range of object contrasts (*bottom*). The same color code is used for *A* and *B*.

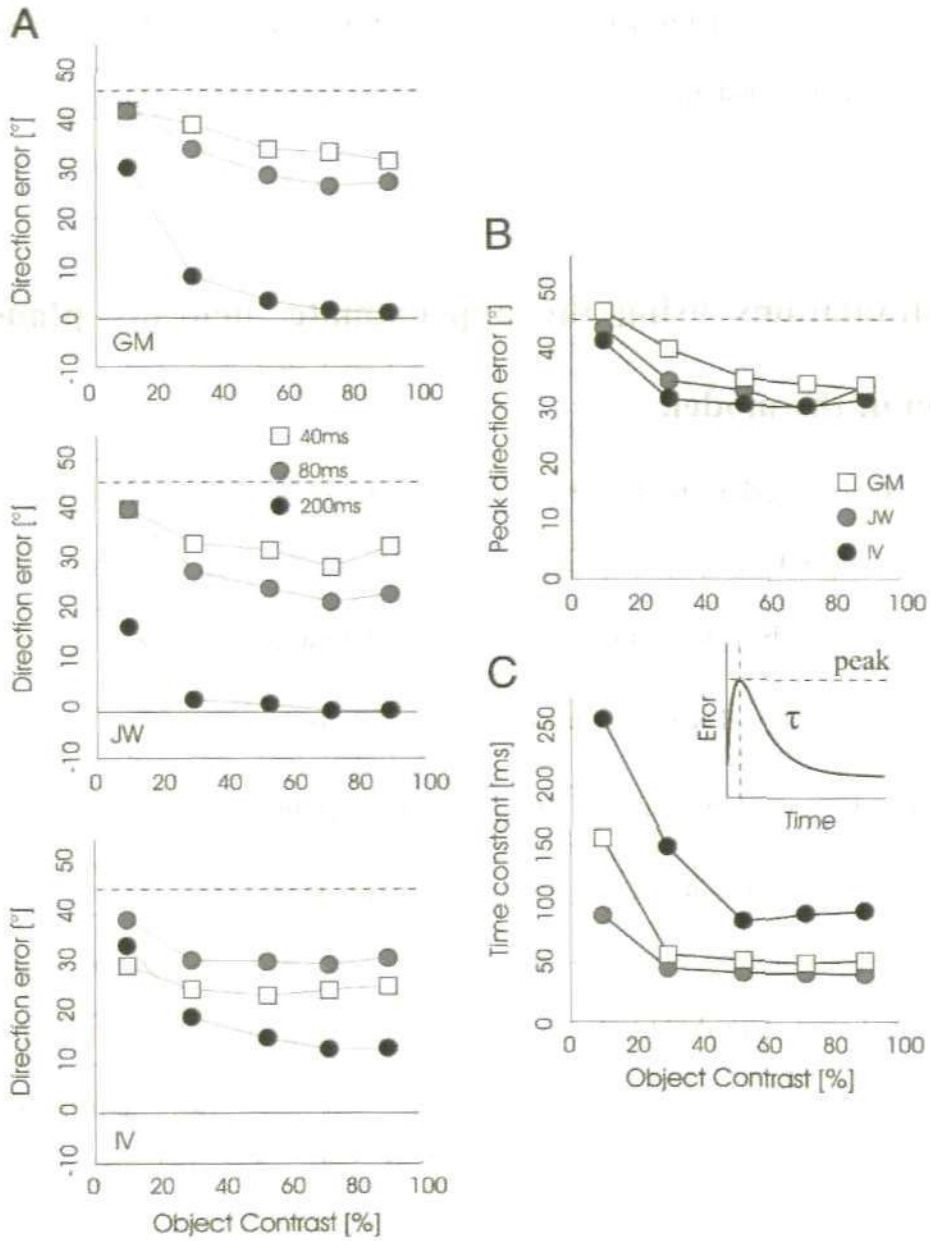


Figure 15 (FIG. 6 from **Wallace *et al.*, 2005** with the original legend). Effects of object contrast on pursuit dynamics. *A*: relationships between mean (across direction) tracking direction errors and target contrast, for 3 different time bins and each subject. *B*: best-fitting peak direction errors, as a function of object contrast, for each subject. - - -, the vector average prediction, 44° away from the actual object-motion direction. *C*: for the same 3 subjects, the best-fitting decay time constants are plotted against contrast.

Overall, the model simulation results show a qualitatively, and in specific instances quantitatively, high level of correspondence with the experimental results of Wallace *et al.* (2005), in particular in respect of the variation of both the peak offset bias in the target velocity and the decay of this bias to zero, with contrast level.

4.2. Simulations using the approximate, neurally plausible version of the model.

The simulation experiments described in Figure 12 were repeated with the motion integration model based on the approximate form of the Kalman filter estimation algorithm, as described in section 3 above, where for simplicity $\sigma_i = \sigma$ for all i, and $(\sigma^2_{p,x})^0_0 = (\sigma^2_{p,y})^0_0 = \sigma^2_p$. The results are shown in Figure 16.

As in Figure 12, the value of σ / σ_p in Figure 16 increases from the top row to the bottom, and results in a corresponding increase in the peak offset bias in the estimated direction for both types of stimuli. There is again a small, but non-zero peak offset bias in the case of the diamond stimulus, which is less than 1° in all but the case of the highest value of σ / σ_p and the lowest contrast (lower left-hand plot).

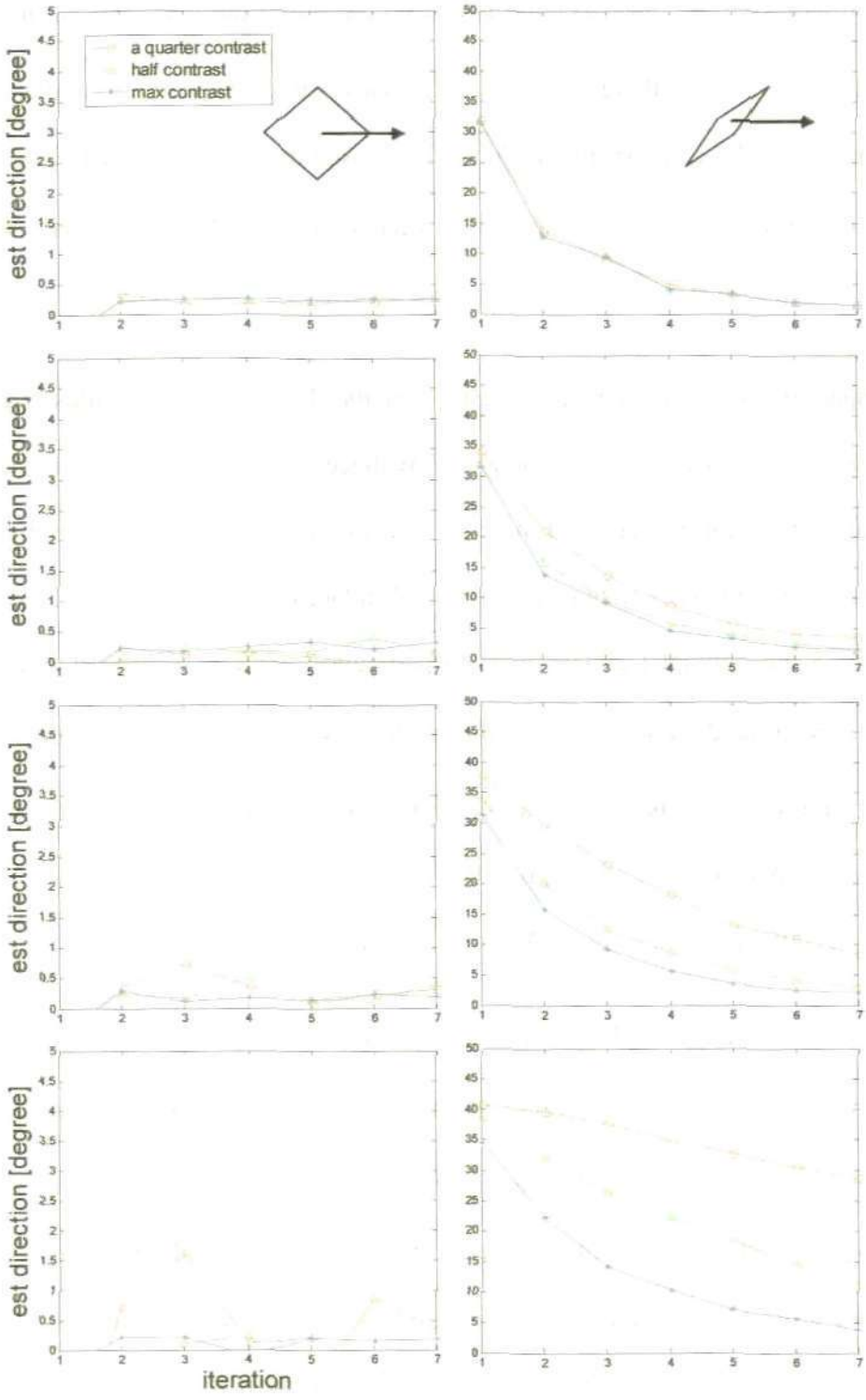


Figure 16. Results from the computer simulations of the motion integration model as in Figure 12, but for the model based on the approximate Kalman filter algorithm.

For the thin rhombus stimulus however, there is a substantial peak offset bias (>30°) for all values of σ / σ_p (note that there is no change in the vertical scale in the plots in the first two rows as there is in Figure 12). This is in contrast to the simulations described in Figure 12 for the optimal Kalman filter based model, where such values of direction error are obtained only for the highest value of σ / σ_p (lower right-hand plot in Figure 12).

Comparing the model simulation results for the thin rhombus stimulus (right-hand column of plots) with Figures 5 and 6 of Wallace *et al.* (2005) (here shown as Figures 14 and 15), it can be seen that the peak direction error in the model simulation now corresponds more closely to that described by Wallace *et al.* than is the case for the model based on the optimal Kalman filter. In particular, in the case of all values of σ / σ_p used, the peak direction error of the model has a variation with contrast of between 30° and 40°. This is in close agreement with the variation in peak direction error observed by Wallace *et al.*, which was between 30° and 45° (Figures 6A and 6B of Wallace *et al.* (2005), presented here as Figure 15A and 15B)

It can also be seen from Figure 16 that the decay time constant of the direction error also increases with both the value of the σ / σ_p ratio and with reducing the contrast, as in the case of Figure 12, and again closely mimics qualitatively the variation in decay time constant with contrast observed experimentally by Wallace *et al.* (Figure 6C of Wallace *et al.* (2005), here Figure 15C). In this case, however, compared to the optimal filter based model (Figure 12), the time constants are generally larger. Interestingly, the model mimics a significant characteristic of the Wallace *et al.* data, in that the time constants stay approximately constant for all but the lowest levels of contrast where they increase substantially (~50% increase – see Figure 6C of Wallace *et al.* (2005), here as Figure 15C). The model results also display this feature of the

experimental data for all but the lowest value of the σ / σ_p ratio, as seen in the three lower right-hand plots of Figure 16. Overall, in the case of the approximate filter based version of the model there is a higher level of correspondence between the simulation results and the experimental results of Wallace *et al.* (2005), than in the case of the optimal filter based version of the model.

A further set of simulation experiments were carried out with the model based on the approximate Kalman filter, using the thin rhombus with varying length, in order to see the effect of the relative influence of 1-D versus the 2-D local motion measurements. In principle, a longer stimulus will contain a larger proportion of 1-D cues, compared to 2-D cues, and thus should result in an increase in the peak offset bias in the direction estimate and a longer time of decay of this bias to zero. This prediction was confirmed experimentally in a study for smooth pursuit initiation in monkeys (Born *et al.*, 2006). They showed that the directional error is more pronounced and lasts longer for the longer length tilted bars.

The model simulation results are presented in Figure 17, using only one level of covariance ratio σ / σ_p equal to 5 ($\ln(\sigma^2 / \sigma_p^2) = 3.2$). for simplicity. When the length of the long diagonal of the 'thin' rhombus increases, both the peak directional error and the decay time increase, approaching the value of the vector average direction for the longest stimulus used. These results compare well in a qualitative sense with those of Born *et al.* (2006) (see their Figure 4, which is shown here as Figure 18).

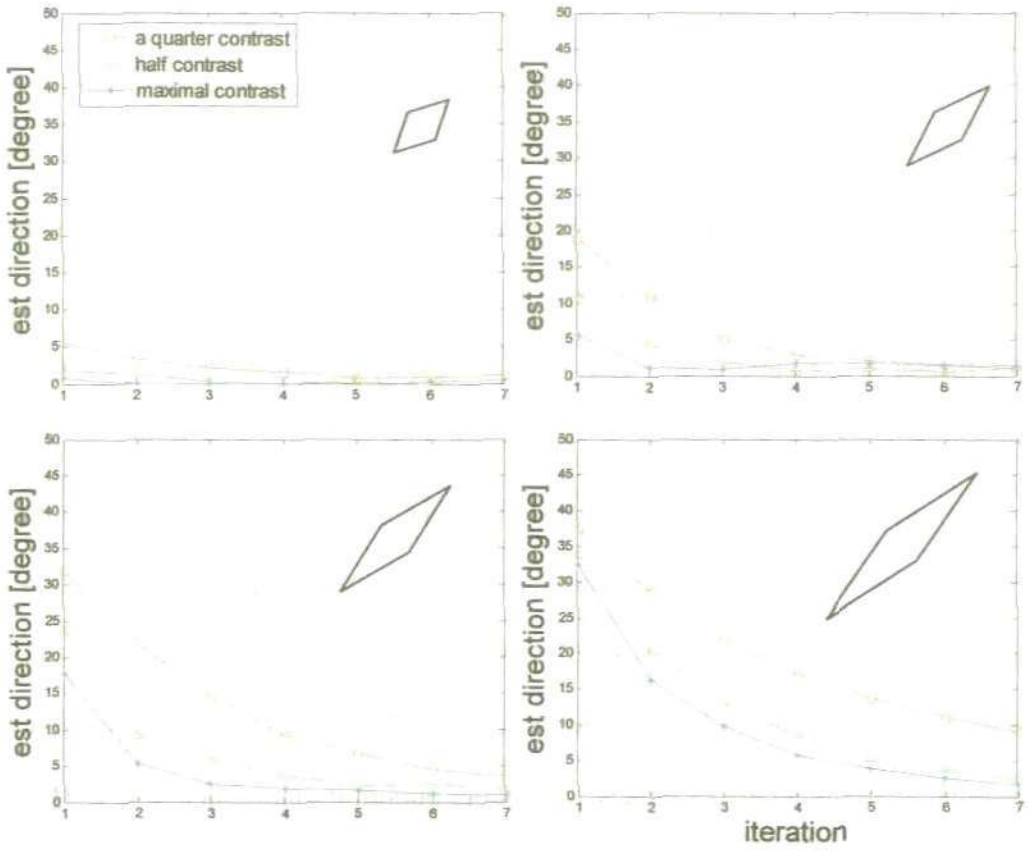


Figure 17. Results from the computer simulations of the motion integration model based on the approximate Kalman filter algorithm, for four different sizes of the thin rhombus and three values of contrast. The sizes are: 10, 20, 30 and 50 pixels length for the main diagonal and a constant size of 3 pixels for the short diagonal. The value of σ / σ_p ratio used was equal to 5 ($\ln(\sigma^2/\sigma_r^2) = 3.2$).

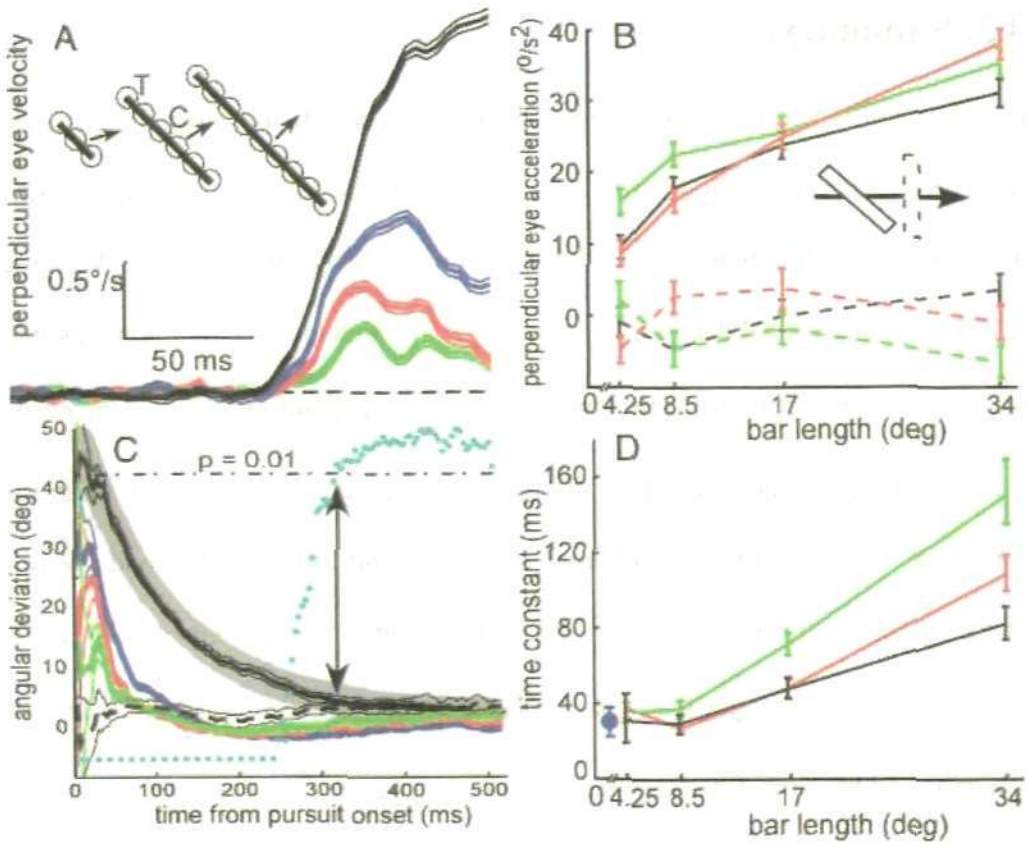


Figure 18 (FIG. 4 from **Born *et al.*, 2006** with the original legend). Results of varying bar length. *A*: average eye velocity perpendicular to the direction of target motion for bars of different lengths for *monkey H*. Each thick line is the average of _900 trials; thin lines represent SE. *B*: initial (from 0 to 40 ms after pursuit onset) perpendicular eye acceleration as a function of bar length for 3 different monkeys (*C*, green; *H*, red; *G*, black) for bars that were either tilted (solid lines) or not (dashed lines). The error bars indicate SE. *C*: time course of the angular deviation for bars of different lengths in *monkey H* (same data as in *A*). The thick lines represent the direction of the mean vector and the thin lines represent the 95% confidence interval about the mean direction. The cyan circles plot the *P* values of a 2-sample test (Watson-Williams test) at successive time points comparing the deviation induced by the longest tilted bar (solid black line) with the nontilted bar of the same length (dashed black line). The time at which the difference becomes nonsignificant (arrows) was defined as the duration of the deviation. The significance criterion, $P = 0.01$, is represented by the horizontal dash-dot line. Symbols near the bottom of the plot correspond to a $P < 10^{-14}$. *D*: time constants of the best-fitting single-exponential decay as a function of bar length for the same 3 monkeys (colors as in *B*). Error bars indicate 95% confidence intervals determined using a bootstrap procedure (see METHODS). The filled blue circle indicates the best-fitting time constant for the population data for 60 MT cells recorded from 2 alert macaque monkeys. The fit was to the angular deviation of the mean neuronal direction vector in response to fields of tilted bars, each bar being 3° long (Fig. 2C from Pack and Born 2001).

4.3. Summary.

In this chapter the results were shown of the model simulations using a square diamond shape stimulus and a thin 45° tilted rhombus stimulus. For the tilted rhombus, moving in a horizontal direction, the results were qualitatively similar to the Wallace *et al.* (2005) psychophysical data, i.e. the error bias in the velocity estimation was bigger for the lower contrast stimuli. At the same time the bias depends directly on the ratio σ / σ_p

The results obtained from the approximate version of Kalman filter algorithm for both the horizontally moving rhombus and square diamond were closer quantitatively to the Wallace *et al.* (2005) data. The directional bias for all levels of the σ / σ_p ratio was stronger than the one calculated with the optimal Kalman algorithm. The bias and the time decay constant were the biggest for the case of the lowest contrast of the stimuli, which was strongly with accordance with the Wallace *et al.* (2005) results. This approximated algorithm gives better results than the optimal one, as it predicts a directional bias even for the case of low σ / σ_p ratio and high contrast, which is in conformation with the experimental data on smooth pursuit. In the discussion of these results in the following Chapter, it will be pointed out how they differ from the study of Weiss *et al.*. (2002), which is the closest to the work presented here.

The other important result described in this chapter regarding the approximate version of the Kalman filter algorithm, is the direct dependence of the directional bias on the stimulus length. It shows that the algorithm needs time to take into account the relative strength of the information provided from the 1-D and 2-D motion detection.

The simulation results of the model as performed by both optimal and approximate versions of the Kalman filter gave similar qualitative results and they were in accordance with the psychophysical data on smooth pursuit regarding the initial non-veridical bias of the tracking eye movements. In general, the bias showed direct

dependence on the ratio σ / σ_p , and on the stimulus size, and an inverse relationship with the stimulus contrast. These results are in accordance with the psychophysical data of Wallace *et al.* (2005) and Born *et al.* (2006).

The difference, between the approximate and the optimal version appeared mainly in the effect of the ratio σ / σ_p on the direction bias. Whereas in the case of the approximate version the bias was significant for all levels of σ / σ_p, for the optimal algorithm the bias was less sensitive to the values of this ratio, This suggests that for the approximate version it is easier to obtain a directional bias for all contrast levels and σ / σ_p ratios, which is closer to the psychophysical data on smooth eye pursuit.

Chapter 5. Discussion of the application of the motion integration model to smooth eye pursuit.

5.1. Simulation results.

The aim in this part of the thesis has been to develop a theoretical, neurally plausible computer simulation model of the motion integration process which is intrinsic to the control of eye movements in the smooth pursuit of moving target stimuli. In particular psychophysical data for initiation of smooth pursuit eye movements suggest that the integration of local measurements of the velocity of the stimulus starts before the closing of the oculomotor control loop and is correcting for inaccurate, ambiguous local motion signals prior to the onset of any compensatory eye movement (Wallace *et al.*, 2005; Born *et al.*, 2006). This data also demonstrates a dynamical behaviour for the correction of the off-axis directional error in the global motion signal used to initiate eye movement, presumably reflecting the dynamics of an underlying motion integration and global velocity calculation process in the brain and the subsequent correcting eye movements. To develop the model, a recursive Bayesian estimation of the global stimulus velocity was used, based on the local 1-D and 2-D motion measurements. The estimation process was implemented as a Kalman filter algorithm for optimally determining the posterior distribution of the global velocity estimate.

This algorithm makes some basic assumptions about the nature of the motion observation process, notably that it is linear and subject to additive Gaussian distributed measurement noise. These assumptions were used in the original description of the Kalman filter (Kalman, 1960), and has subsequently been commonly used in engineering problems because it leads to a tractable solution to the estimation problem. This was pointed out by Weiss *et al.* (2002), who made the same linearity and Gaussian

measurement noise assumptions in the formulation of their motion integration model. It is also assumed that (i) the stimulus is moving with constant velocity, as in the psychophysical experiments, and (ii) changes in image intensity at any location in visual space are due only to motion of the stimuli, the so-called "standard principle of intensity conservation", a standard assumption in several models of motion processing (Heeger and Simoncelli, 1993; Weiss *et al.*, 2002).

Given these assumptions the simulation results from the Kalman filter based motion integration model have been shown to closely match the experimental data of smooth pursuit. In particular it was shown that such a model has a dynamic behaviour, the result of the recurrent Bayesian estimation process, which possesses qualitatively the same characteristics as the experimentally observed dynamics of the motion integration process during the initial stages of smooth eye pursuit (Wallace *et al.*, 2005; Born *et al.*, 2006). Specifically, the variations in the stimulus contrast in the model simulations lead to changes in the peak directional error and the time constant for the decay of this error to zero which are qualitatively consistent with the changes in the dynamics of the directional error induced by similar contrast variations in the experimental situation (Wallace *et al.*, 2005). Similar consistent results were obtained from the model in response to changes in stimulus length and the corresponding ratio of 1-D and 2-D local motion measurements (Born *et al.*, 2006). These results strongly suggest that the brain may be using some form of Bayesian estimation process to correct for the presence of ambiguous 1-D local motion cues in the calculation of a veridical global stimulus velocity for smooth eye pursuit.

Furthermore, the neural plausibility of the model was demonstrated, in terms of the feasibility, in principle, of its implementation using neural computational methods, i.e. avoiding the computation of matrix inversions as in the optimal Kalman filter based model. This required the development of an approximate version of the Kalman filter

which eliminated the need for matrix inversion. The simulation results from the motion integration model based on the approximate Kalman filter algorithm presented here show an even closer correspondence with the experimental data, suggesting that the brain may possibly be adopting a form of sub-optimal Bayesian approach to the estimation of stimulus velocity in the integration of local motion cues.

5.2. The origin of motion integration dynamics in the initiation of smooth eye pursuit.

For some time it has been unclear as to the origin of the temporal dynamics which have been observed in the initiation of smooth eye pursuit in both humans and monkeys (Masson and Stone, 2002, Wallace *et al.*, 2005) and the associated dynamic response properties of monkey MT neurons (Pack and Born, 2001; Born *et al.*, 2002). One recent proposal has been that it is the consequence of a delayed response of end-stopped cells resulting in a change in the "weighting" applied to the outputs of V1 direction-selective neurons as end-stopping eventually suppresses contour-related motion signals and emphasizes those from terminators (Pack *et al.*, 2003; Born *et al.*, 2006). However this suggestion has not resulted in the development of a specific model which can be tested against the behaviourally and physiologically observed dynamics.

It has also been proposed that the independent computation of Fourier and non-Fourier motion signals followed by the computation of their vector average will lead to a dynamic response in motion integration owing to the delay in computing non-Fourier motion signals relative to Fourier motion signals (Löffler and Orbach 1999; Wilson *et al.*. 1992). The phenomenon of non-Fourier motion perception, the visual perception of motion that cannot be explained simply on the basis of the autocorrelation structure of the visual stimulus, is well recognised, and is generally considered to be due to nonlinear preprocessing of the visual stimulus prior to standard motion analysis. A

Fourier motion stimulus is one whose spatiotemporal luminance function has directional motion energy, and thus, can be identified as moving from its autocorrelation structure, e.g. by a Reichardt motion detector. For n = 2, 3, 4,..., an *n-th*-order, non-Fourier, motion stimulus is a stimulus whose spatiotemporal luminance function has directional motion energy only if it is first passed through a polynomial of order n, and not of lower order. Thus motion of a second order, non-Fourier stimulus can be detected by first subjecting the stimulus to a squaring operation, or full-wave rectification. The effect of this nonlinear pre-processing is thus to translate "texture-defined" motion, motion defined by translations of some texture property such as contrast or spatial frequency, into luminance motion so that it becomes accessible to standard motion analysis (Chubb & Sperling, 1988). There are several examples of texture-defined motion which may be readily perceived by human observers, yet cannot supposedly be analysed by luminance-based motion models (Chubb and Sperling, 1988, 1989; Cavanagh & Mather 1989; Benton & Johnston 1997). However, more recently it has been shown that the motion of several texture-defined motion stimuli can be detected by luminance-based mechanisms (Benton et al, 2001).

The model of Löffler and Orbach (1999) closely follows that of Wilson *et al.* (1992), with additional components which increase physiological plausibility. In the model of Löffler and Orbach (1999), two parallel pathways, Fourier and non-Fourier, extract the motion of luminance boundaries and texture boundaries respectively. The Fourier pathway extracts motion using directionally-tuned Reichardt detectors. In the non-Fourier pathway, the responses of model V1 simple cells are squared and then filtered at a second stage, presumably by cells in area V2. These filters are tuned to a lower spatial frequency and oriented orthogonal to the initial V1 cell filters to extract texture boundaries. Qualitatively the same motion extraction steps then follow as for the Fourier pathway. Finally, the signals of the Fourier and non-Fourier pathways are

combined, supposedly at the level of MT pattern detection cells. Thus the non-Fourier pathway in the model has three qualitative differences from the Fourier pathway. First, the second stage filters do not operate on the local contrast differences of the stimulus itself, but on the squared responses of the initial filters of the stimulus. Second, this second stage filtering is carried out by lower spatial frequency filters with orientation tuning which is perpendicular to the orientation tuning of the initial filters. Third, the second stage filters have opposite polarity: they exhibit off-centre characteristics.

In support of the two-pathway model, the time-delay present in the non-Fourier pathway relative to the Fourier pathway has been found to be approximately 60 ms (Yo and Wilson, 1992), which is consistent with the observed dynamics of motion integration. Recent work by Barthélemy *et al.* (2008) also shows that the Fourier and non-Fourier motion signals display different contrast dynamics, which may underlie the variation of the dynamics of motion integration with contrast in such a model.

Loffler and Orbach (1999) tested their model on stimuli corresponding to stationary terminators, plaids, and moving terminators. Although they did not simulate the temporal dynamics of their model explicitly, they make a prediction for the dynamics of direction perception of briefly presented terminator stimuli, which suggests an initial directional bias approximately equal to the output of the Fourier pathway alone. As far as the author is aware, the model has not been used to explain the dynamics of motion integration in smooth pursuit initiation, and it is not clear whether or not the results of Wallace *et al.* (2005) and Born *et al.* (2006), in particular the variation in magnitude and decay time of the directional bias with stimulus shape and length could be readily explained by this model, since such variations would appear to have little or no effect on the time delay of the non-Fourier pathway. Furthermore, as pointed out in Smith *et al.* (2005), there is no clear evidence of a separate cortical pathway for the computation of non-Fourier "pattern" motion cues, such as in areas V2

or V3, which do not apparently make an important separate contribution to the behaviour of pattern direction selective neurons in MT.

A number of models of motion integration have been based on the idea of the spatial propagation of 2-D motion signals from line terminators and their suppression of the ambiguous 1-D motion signals. The model of Lidén and Pack (1999) employs recurrent networks to spatially propagate motion signals across model MT cells, the propagation dynamics providing a good qualitative fit to the temporal dynamics of motion integration as observed in MT (Pack and Born, 2001). Similarly, the motion integration and segmentation model of Grossberg *et al.* (2001) uses feedback connections between model MT and MST cells to select winning directions and suppress losing directions on a top-down matching process, resulting therefore in a temporal dynamics. It also uses lateral connectivity in MT cells to amplify 2-D feature tracking signals and suppress 1-D ambiguous direction signals, in a manner similar to that of Lidén and Pack (1999). Neither of these two models has been used to simulate the smooth pursuit initiation data of Wallace *et al.* (2005), but in principle could provide an explanation of the dynamic nature of this data. Indeed, Lidén and Pack (1999) have used their model to simulate the perception of a horizontally oriented line translating at 45° relative to its orientation. The model initially signals horizontal motion, and then gradually recovers the true line motion direction by propagating unambiguous motion signals generated by the terminators along the contour of the line. The model dynamics is therefore based on the dynamics of the spatial propagation of the motion signals, which the authors relate anatomically to the lateral connectivity between motion selective cells in MT. Some evidence exists for laterally extending fibres in MT (Van Essen, Maunsell & Bixby 1981), but further detailed physiological examination of these connections and their propagation delays would be necessary to reveal whether or not the dynamics conveyed by this connectivity is consistent with the psychophysically

(Wallace *et al.*, 2005) or physiologically observed (Born *et al.*, 2006) dynamics of the perceptual bias in the initiation of smooth pursuit.

The model of motion disambiguation proposed by Bayerl and Neumann (2004) also suggests that the dynamics of motion integration results from the spatial propagation of signals in MT, and whilst it has not been used to explain specifically the data on smooth pursuit initiation, such propagation dynamics would similarly predict the observed dynamics of smooth pursuit initiation. An additional feature of this model compared to that of Lidén and Pack (1999) is that it uses feedback from MT to V1 to attentionally gate by excitatory modulation the unambiguous V1 motion signals, and thus takes some account of the recurrent nature of the MT-V1 circuitry.

In the Bayesian estimation framework, Koechelin *et al.* (1999) describe a model of motion integration based on the V1-MT circuitry that also employs mechanisms of recurrent lateral interactions. Their model proposes a multiplicative combination of feed-forward input and the result of lateral integration, which leads to the proposal that the model represents a neural implementation of Bayesian motion estimation.

Also using a Bayesian approach, a recursive extension of the Bayesian model of motion estimation of Weiss *et al.* (2002) has recently been proposed (Montagnini *et al.*, 2007). They used human experimental data on smooth pursuit eye movements in response to dot and line stimuli to derive, respectively, the variances of two likelihood functions, one for the 1-D cues and one for the 2-D cues. These likelihood functions are used together with the same prior as in Weiss *et al.* (2002) in a recursive version of the Weiss *et al.* model, to produce a discrete evolution of velocity estimates. They show that the temporal evolution of the velocity estimates expressed in terms of tracking error, coarsely matches their experimental data for mean eye velocity and different target stimulus speeds. However they have not attempted to explain the observations of Wallace *et al.* (2005) on smooth pursuit initiation.

In contrast to the above models, the model proposed in this thesis uses a neurally plausible recursive estimation process based on an approximate form of the Kalman filter, and has been specifically addressed to the issue of motion integration dynamics in the initiation of smooth eye pursuit. As has been shown, the model proposed here demonstrates the main characteristics of the experimentally observed dynamics, and the dependence of the model dynamics on contrast qualitatively replicates the experimentally observed contrast variation in the motion integration dynamics (Wallace *et al.*, 2005; Born *et al.*, 2006). It has also been possible to show similar qualitative agreement between the model simulations and the experimental data of Born *et al.* (2006), for the dependence of the motion integration dynamics on stimulus length.

An important feature which distinguishes the model proposed here from several of the others described above is that it does not depend on the concept of lateral propagation of unambiguous motion signals along line contours in order to achieve a veridical estimation of motion direction in the presence of ambiguous direction cues. Rather, it is suggested here that true motion perception is achieved through a recursive estimation process which produces successively improved velocity estimates by eliminating the initial estimation errors introduced by the ambiguous motion signals of local detectors. It is suggested that this can be achieved by the divergent feedback connections which are known to exist from MT motion selective cells back to V1 cells, and which connect to a retinotopic area of V1 which corresponds to the receptive field size of the MT cell (Shipp and Zeki, 1989). At the present time, more is known about these connections than is about the lateral connections in MT, in particular the fast transmission times of V1-MT connections (Raiguel *et al.*, 1989; Schmolesky *et al.*, 1998; Hupe *et al.* 2001). It is very clear however that much further work will be needed, in relating the models to their proposed physiological and anatomical basis in the visual system, before it becomes clear as to whether the origin of motion integration dynamics

lies predominantly in: (i) the delayed activation of end-stopped cells (Pack *et al.*, 2003; Born *et al.*, 2006); (ii) separate Fourier and non-Fourier motion pathways (Löffler and Orbach 1999); (iii) lateral connectivity in MT (Lidén and Pack, 1999; Bayerl and Neumann, 2004; Koechelin *et al.*, 1999)); (iv) feedback connectivity between MT and MST (Grossberg *et al.*, 2001); or (v) feedback connectivity between V1 and MT.

The model proposed here provides support for the notion that motion integration in the brain might be based on a Bayesian estimation process, as has been suggested by many psychophysical studies, and suggests that the observed motion integration dynamics, and their dependence on stimulus contrast and length, may result from the recursive nature of this motion estimation process. The model also reflects the recurrent MT-V1 circuitry, the pooling of information from local V1 motion detectors, the convergence of the polled local motion signals in projections to MT, and the divergent feedback of MT velocity signals to V1 cells. However far more work needs to be done before the recurrent MT-V1 circuitry can be shown to be capable of supporting the kind of recursive Bayesian estimation algorithm inherent in the model.

It is also clear that the proposed model, along with all the models reviewed here, is limited by its ability to explain the dynamics of motion integration in the initiation of smooth eye pursuit only in relation to intensity-based motion stimuli, and not moving targets defined in other ways, e.g. cyclopean targets defined by means of random-dot stereograms. We note however that although it is likely that different neural pathways are involved in luminance motion and stereomotion perception, it is possible that the roles played by the pooling of local motion detectors, and by moving stereoscopic line terminators are similar in both cases (Donnelly *et al.*, 1997; Patterson *et al.*, 1998).

5.3. The effect of the free parameter σ / σ_p.

The free parameter in the model, σ / σ_p, the ratio of standard deviations of the probability density functions of the measurement noise and the prior velocity estimate, has been varied as shown in Table 1, in order to demonstrate the dependence of the model behaviour on this parameter. As has been pointed out by Stocker and Simoncelli (2006), Bayesian models of visual motion perception are difficult to validate quantitatively owing to the fact that it is hard to attribute precise values to these variables. Even though a reasonable estimate of the prior velocity estimate can be made on theoretical grounds, or based on the statistics of natural visual stimuli, uncertainty in the value of the measurement noise variance remains. It is interesting however to note that this variance always appears in the model as a divisive modulation of the squared image intensity derivatives, that is, as $\frac{I^2_{x,t}}{\sigma_i^2}, \frac{I^2_{y,t}}{\sigma_i^2}$. Thus we can think of the measurement noise variance in neural terms as a signal which increases or decreases the magnitude of the output of the squared linear filter representation of the neurons which code for $I^2_{x,t}, I^2_{y,t}$. The obvious suggestion for a neural implementation of the model is that the influence of the measurement noise variances in the Bayesian formulation could be identified with the divisive normalisation role of the intracortical connections between orientation hypercolumns (Carandini *et al.*, 1997). Thus neurons reporting the directional derivatives (or their orientated versions) of the image intensity would be suppressed by the activity of neurons in the surrounding spatial region which were responding to stimuli with a different orientation/direction, as observed physiologically (Morrone, Burr and Maffei, 1982), indicating an increased uncertainty in the local motion measurement consistent with a decrease in signal to noise ratio for this measurement. If this approach is adopted, there is no need for a precise measurement of the measurement noise variance, which in any case is highly unlikely to be represented

in any V1 neuronal responses, but instead the divisive role of this variance is used in the model to mimic the effect of divisive normalisation via intracortical lateral connections (Heeger, 1992; Carandini and Heeger, 1994; Heeger *et al.* 1996; Carandini *et al.*, 1997)

5.4. Predictions of the model.

The recursive nature of the calculation of the velocity estimate in the model suggests that blanking of the stimulus for short periods of time will not result in the reintroduction of an offset directional error in the eye movement when the stimulus reappears at the end of the blanking period. This follows from the fact that the update of the velocity estimate, as expressed in equation (31), depends on the error signal $(h_k - C_k v_k^{k-1})$. Owing to the definitions of h_k and C_k (following equation (4)) in the absence of any stimulus during the blanking period, both h_k and C_k will be zero. Thus no updates of the velocity estimate will take place during the blanking period, and at the end of the period the original velocity estimate will be used in further updates. Thus the estimation procedure will continue as if it had not been interrupted. This "prediction" of the model has already been confirmed by the experiments of Masson and Stone (2002), in which the motion of a tilted elongated diamond stimulus (thin rhombus) was transiently blanked for a period of 90 ms during steady state pursuit. They observed a small decrease in the eye velocity in the veridical direction during the blanking period and a transient increase in this velocity immediately after the blanking period ended, but no post-blanking reappearance of an offset direction error. The small decrease in the eye motion in the veridical direction could be explained by a slow decay of the velocity estimate in the model in the absence of any updating, which would maintain the directional component of the estimate but decrease its magnitude. The update of the estimate post-blanking would then rapidly correct for any magnitude error without introducing any directional error.

5.5. Summary.

This chapter was a discussion on the application of the presented model to the smooth eye pursuit experimental data. The simulation results and other models were discussed in regard to the origin of the motion integration dynamics. The role of the free parameter σ / σ_p was examined in terms of image intensity derivatives normalization with suggestion about the possible role of the lateral connectivity in V1 area in this process. Finally, some predictions of the model were derived, in particular about the interruption of the stimulus presentation during the smooth pursuit initial stage.

Chapter 6. Motion integration in the perception of plaid patterns.

6.1. Models of motion integration in the perception of plaid motion.

The problem of how the visual system combines the motion of two moving gratings to form the percept of a coherent moving plaid pattern is still unsolved after nearly thirty years of research. It has long been known that the plaid motion can be computed by a velocity space construction, known as the "intersection-of-constraints" (IOC) (Fennema & Thompson, 1979). The intersection-of-constraints rule was described in the introduction of the thesis (Figure 6 and 7); nevertheless it will be briefly restated: the motion of an edge viewed through an aperture has a family of possible velocities, which form a so-called constraint line of velocities. The intersection of the constraint lines corresponding to the motion of all the edges forming an object gives the point in velocity space which defines the true velocity vector of the object. Based on this rule, Adelson & Movshon (1982) proposed a two-stage model for the analysis of plaid motion in which the one-dimensional (1-D) motions of the plaid's two component gratings are first determined, and then combined in a weighted summation corresponding to the IOC construction. This model has dominated research in the area for almost thirty years, despite the psychophysical (Welch, 1989; Derrington & Suero, 1991; Derrington & Badcock, 1992; Stone, Watson & Mulligan, 1990) and physiological (Movshon *et al.*, 1985; Movshon & Newsome, 1996; Tinsley *et al.*, 2003) evidence being equivocal. In particular, the available evidence is based entirely on experiments using symmetric Type I plaids (Ferrera & Wilson, 1990), for which the

plaid velocity vector lies between the velocity vectors of the two component gratings, which have equal magnitude. The strongest evidence against the Adelson and Movshon (1982) model was obtained when Type II plaids, the velocity vector of which lies outside of the velocity vectors of the two component gratings, were used in psychophysical experiments (Yo & Wilson, 1992). These experiments demonstrated that the direction of the plaid motion during the initial period (up to ~60 ms) of stimulus presentation is misperceived, with a strong bias in the perceived direction towards the vector sum (VS) direction of the velocities of the component gratings. Whilst it is possible that the Adelson and Movshon (1982) model is correct for Type I plaids, and that another mechanism is responsible for Type II plaid motion perception, this would seem highly unlikely.

Subsequent to the Yo and Wilson (1992) experiments, and prior experiments which showed identified misperceptions in the direction and speed of Type II plaids (Ferrera and Wilson, 1990, 1991), several models have been proposed which attempt to explain these misperceptions. Wilson, Ferrera and Yo (1992) suggested a model, subsequently extended by Wilson and Kim (1994), which consisted of two parallel processing pathways, one signalling the direction of the component gratings (presumed to be mediated by neurons in area V1 of visual cortex) and the other (presumed to be end-stopped neurons in area V2) signalling, after a hypothesised delay of ~77 ms, the direction of "the motion of illusory lines formed by the nodes of the Type II pattern" (Yo and Wilson, 1992). The signals of the first pathway are combined (by neurons in extrastriate area MT to which both V1 and V2 neurons project) to form a cosine-weighted sum of the component grating velocities. The signals of the second pathway are derived after full-wave rectification of the stimulus and orientation filtering at a lower spatial frequency than that of the component gratings (postulated to take place in V2). A cosine-weighted sum of the two pathways is then followed by competitive

feedback inhibition in order to predict the perceived plaid direction. The delay in the second pathway accounts for the initial misperception of the plaid direction towards the vector sum direction of the component gratings' velocities. Whilst this model offers a compelling explanation of the observed misperception, it is deficient in several respects, as discussed in Alais *et al.*. (1997), who carried out experiments on the effect of the size and number of plaid features, or "blobs", the "nodes of the Type II pattern" referred to above, on the misperception. They concluded that a more likely explanation is based on "a feature sensitive mechanism which responds to the motion of plaid features and which is tuned to their various qualities" (Alais *et al.* 1997). The plaid blobs which they examined and refer to are the high luminance regions which are formed at the intersection of the component gratings and which, in particular for Type II plaids, are the most visually salient features in the plaid pattern for a human observer.

6.2. Bayesian model of motion integration in plaid perception.

In this part of the thesis, the aim is to demonstrate that the misperception of the plaid direction, its dependence on the angular separation and contrast of the component gratings, and its decrease with lengthening stimulus duration, can all be fully explained by a two-stage model which is based on the detection of both the one-dimensional (1-D) and two-dimensional (2-D) motion of the blobs, and their combination by a recursive Bayesian velocity estimation process This is the same model, which has already been presented in Chapters 2 and 3 and applied successfully on smooth eye pursuit dynamics.

In the first stage of the model, local motion detectors respond to both the 1-D and 2-D motion of the blobs within the plaid. It is hypothesised that these detectors are based on the complex and hypercomplex (end-stopped) neurons in V1 (Hubel & Weisel, 1965; Pack, Livingstone, Duffy, and Born, 2003). This stage of the model differs from

that of Wilson *et al.* (1992) in that: (i) the 1-D motion signals are derived not from the motion of the component gratings but from the edge motion of the blobs; (ii) there is no requirement for the separate combination or any explicit weighting, cosine or otherwise, of the 1-D signals; (iii) the 2-D motion signals are derived directly from the terminations (end-points) of the blobs, without the requirement for full-wave rectification (squaring) of the plaid stimulus. In the second stage of the model, the 1-D and 2-D motion signals are combined using a recursive Bayesian least squares estimation process, which might be postulated to occur in the recurrent V1-MT circuitry. This also differs from the Wilson *et al.* (1992) model in that cosine weighting of the 1-D and 2-D signals is not required, nor is there the need for a final stage of *competitive inhibition.*

In the next Chapter of the thesis, the specific geometric properties of the blobs which, it is claimed, play the main role in the perception of plaid motion are examined in detail. In particular it is shown that the shape of the blobs, specifically the extent of their elongation, is defined by the angular difference in the directions of motion of the component gratings, and that the orthogonal direction of motion of the longer edges of the elongated blobs is given by the *mean* of the directions of motion of the component gratings. It is also shown that as the blobs become more elongated, the orthogonal direction of motion of the longer edges of the blobs tends towards the VS of the directions of motion of the component gratings. Although the blobs have been implicated in the perceptual process by several authors (Wilson *et al.*, 1992; Burke and Wenderoth, 1993; Wenderoth, Alais, Burke and van der Zwan, 1994; Alais, Wenderoth and Burke, 1994, 1997), as far as the author is aware this is the first time that the geometric properties of the blobs and their relationship to the directions of motion of the component gratings have been precisely defined.

It is then shown theoretically how these particular properties of the blobs can lead to the misperception of Type II plaid direction which has been observed psychophysically (Yo and Wilson, 1992; Burke and Wenderoth, 1993). To demonstrate this more fully, a computer simulation of the model is used to mimic the observed misperception, and show that the magnitude of the direction bias, its dependence on angular separation and contrast, and the convergence of the perceived plaid direction towards the veridical direction with increasing stimulus presentation duration, are all accurately predicted by the model.

Finally a discussion is included of how the model differs from the two-stage model of Adelson and Movshon (1982), yet is consistent with the available physiological and psychophysical evidence, and how it relates to a recent Bayesian extension of the Adelson and Movshon model (Weiss and Adelson, 1998; Weiss *et al.*, 2002), and the models proposed by Bowns (1996, 2006).

6.3. Summary.

A brief overview of the main models relating to plaid motion was presented in this chapter. It includes the wide accepted idea of the 'intersection-of-constraints' rule (Adelson and Movshon, 1982) versus the numerous psychophysical results showing the violation of the same rule. It was briefly described how the model presented here tries to explain these contradictory results and replace the intersection-of constraints rule with one based on the detection of the dominant local motions in the image.

Chapter 7. Application of the motion integration model to plaid perception.

7.1. Geometric analysis of the plaid blobs.

The high luminance regions of the plaid, i.e. the blobs, which are formed at the intersections of the component gratings, can be precisely defined by representing the plaid as the *product* of two gratings rather than as a *sum* of two gratings, its normal form of representation. Typically a plaid is described by the sum of two sine or cosine gratings, i.e. the spatiotemporal luminance intensity function of the stimulus is defined by

$$I(x,y,t) = \sin(\omega_1) + \sin(\omega_2) \tag{33}$$

where $\omega_i = 2\pi s_i(x\cos\theta_i + y\sin\theta_i + v_i t)$; s_i = spatial frequency (cycles/°); θ_i = direction of motion (°); and r_i = speed (°/sec), for the i th. grating, $i \in \{1,2\}$. Using a simple trigonometric identity, this expression can be rewritten as,

$$I(x,y,t) = 2\sin((\omega_1+\omega_2)/2)\cos((\omega_1-\omega_2)/2) \tag{34}$$

i.e. as the product of two anti-phase gratings, henceforth referred to as the *product* gratings to distinguish them from the *component* gratings used in the summation form (equation (33)) of the plaid. The two product gratings comprise: (i) a sine grating which moves in the direction $\phi = (\theta_1+\theta_2)/2$, and which has a spatial frequency $s_\phi = s(\cos\theta_1+\cos\theta_2)/2\cos\phi$, and a speed $r_\phi = (r_1+r_2)\cos\phi/(\cos\theta_1+\cos\theta_2)$; and (ii) a cosine grating which moves in the direction $\varphi = \phi - 90^\circ$, has a spatial frequency

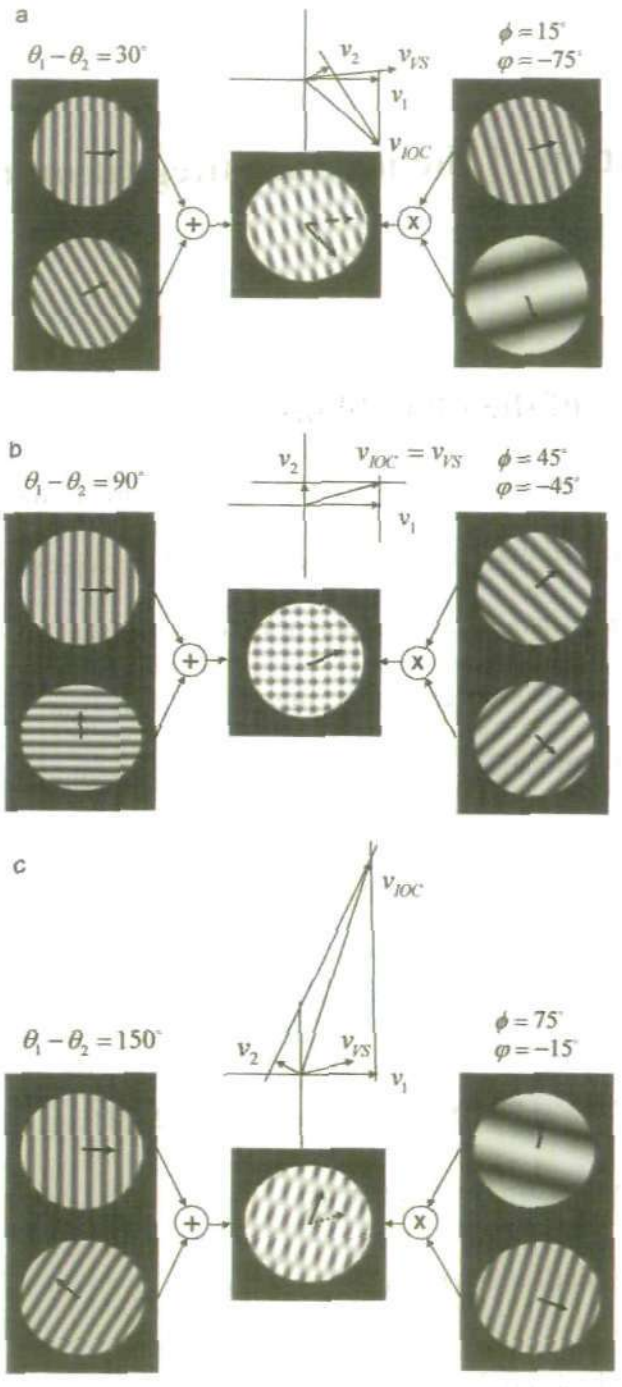


Figure 19. Three examples (a-c) of the representation of a plaid (centre) as the sum (left) or the product (right) of two gratings. The velocity space diagram above each plaid shows the velocity vectors for each component grating, v_1 and v_2 together with the IOC and vector sum velocity vectors v_{IOC} and v_{VS}. The arrows on the gratings and plaids also show their directions of motion, with the dashed arrow on the plaid showing the vector sum direction.

$s_\varphi = s(\cos\theta_1 - \cos\theta_2)/2\cos\varphi$, and a speed $r_\varphi = (r_1 - r_2)\cos\varphi/(\cos\theta_1 - \cos\theta_2)$. For simplicity it is assumed henceforth that $s_1 = s_2 = s$.

Derrington and Ukkonen (1999) used this representation to describe a specific instance of a plaid with component gratings oriented symmetrically about the vertical and a specific relationship between the spatial frequency of the gratings and their orientations. In this case they obtain a vertically oriented, horizontally moving product grating and horizontally oriented, stationary product grating.

Figure 19 shows three examples of equivalent plaid representations in terms of their component and product gratings. These clearly demonstrate that the high luminance regions, or blobs, in the plaid which occur at the intersections of the component gratings are precisely defined by the anti-phase modulation of one product grating by the other. In particular, where the spatial frequencies of the product gratings differ substantially (Figs. 19a and 19c), the blobs are readily seen to correspond to the luminance peaks of the higher frequency product grating modulated by the lower frequency product grating. The shape of the blobs, in terms of the ratio of their long and short edges, is thus determined by the ratio of the spatial frequencies of the product gratings, which can be expressed in terms of the angular difference $\theta_1 - \theta_2$ in the directions of the component gratings as

$$s_\phi / s_\varphi = 1/\tan((\theta_1 - \theta_2)/2) \tag{35}$$

The direction of motion of the blobs (and therefore the IOC direction of the plaid) is given by the VS direction of the two product gratings.

Since Type II plaids (e.g. Figure 19a) are the major interest, and in order to simplify the presentation, the main characteristics of the blobs will be derived here only for this case, although similar equations describing the characteristics of Type I plaids (e.g. Figures 19b and 19c), can be easily obtained. For Type II plaids, the IOC velocity vector lies outside of the two component grating velocity vectors. In this case, the ratio of the speeds of the component gratings, r_1 / r_2, must be greater than one, and the difference in their directions of motion, $\theta_1 - \theta_2$, must be less than 90°. It follows from (35) that, as the difference in the direction of motion of the two component

gratings, $\theta_1 - \theta_2 < 90^\circ$, decreases, the ratio of the spatial frequencies of the product gratings, s_ϕ / s_φ, will increase, and the blobs will become more elongated in shape. Moreover, the shape of the blobs only depends on the difference between the directions of the two component gratings, $\theta_1 - \theta_2$, and not on the ratio of their speeds. It can be seen from (35) that $s_\phi \geq s_\varphi$ when $0^\circ \leq \theta_1 - \theta_2 \leq 90^\circ$, and $s_\varphi \geq s_\phi$ when $90^\circ \leq \theta_1 - \theta_2 \leq 180^\circ$. Also s_ϕ / s_φ increases as the difference between the component grating angles decreases, to a theoretical limit of ∞ when $\theta_1 = \theta_2$, and s_φ / s_ϕ increases as the difference between the component grating angles increases, to a theoretical limit of ∞ when $\theta_1 - \theta_2 = 180^\circ$. It also follows from (35) that $s_\phi > s_\varphi$, and thus the motion of the longer edges of the blobs orthogonal to their orientation will be in the direction $\phi = (\theta_1 + \theta_2)/2$, the mean of the directions of motion of the component gratings.

Most importantly, the difference between the orthogonal direction of motion of the longer edges of the high luminance regions, ϕ, and the vector sum (VS) direction of the velocities of the component gratings, denoted by θ_{VS}, can be expressed as:

$$\phi - \theta_{VS} = \arctan\left(\frac{(r_1/r_2)-1}{(r_1/r_2)+1} \cdot \frac{s_\varphi}{s_\phi}\right), \qquad (36)$$

When the angle $\theta_1 - \theta_2$ is greater than 90°, this expression changes to:

$$\varphi - \theta_{VS} = \arctan\left(\frac{(r_1/r_2)+1}{(r_1/r_2)-1} \cdot \frac{s_\phi}{s_\varphi}\right), \qquad (37)$$

due to the fact that the product gratings change their places (the edge of the high intensity regions which used to be the longer one when $(\theta_1 - \theta_2) < 90°$ now becomes the

shorter one). In particular, the case when the component gratings speeds form an angle greater than 90° is an interesting one, as then the plaid which is formed is Type I, for which the IOC vector lies between the component gratings' velocity vectors. However, the relationship in (37) is similar to that in (36). This means that for asymmetric Type I plaids our model would predict analogous behaviour for the estimated angle as in the case of Type II plaids. This should be especially true for the case when the difference in the directions of the IOC and VS vectors is of similar magnitude to that in the case of Type II plaids, which as described above are perceived non-veridically. Examples will be shown later of the predicted direction misperception for Type I plaids under certain conditions.

Equation (36) shows that, for a fixed ratio of component grating speeds, r_1 / r_2, as the difference between directions of the two component gratings, $\theta_1 - \theta_2$, decreases, and the shape of the blobs become more elongated, the angular difference between the orthogonal direction of motion of the longer blob edges and the VS direction of the component gratings will decrease. It is also worth noting that for a fixed difference in the directions of the component gratings, $\theta_1 - \theta_2$, as the speed ratio r_1 / r_2 increases, the angular difference expressed by (36) will increase, causing the orthogonal direction of motion of the longer edges of the blobs to move away from the VS direction of the component gratings.

7.2. Theoretical predictions of the model.

The geometric analysis of the blobs, as expressed by equations (35) – (37) give rise to theoretical predictions about the behaviour of the model in response to Type I and Type II plaids. In the first stage of the model, it is proposed that local motion detectors signal

both the 1-D (edge) and 2-D (end-point) motion of the blobs present within the plaid. Thus in the case of Type II plaids, for which the blobs are elongated, the majority of the local motion detectors will respond to the 1-D motion of the longer edges of the blobs. Since a local motion detector signals the velocity of 1-D edge motion in the direction orthogonal to the orientation of the edge, owing to the aperture effect (Wallach, 1935; Wuerger, Shapley and Rubin, 1996, Marr and Ullman, 1981), the majority of the local motion detectors will signal motion in the orthogonal direction of motion of the long edges of the blobs. The geometric analysis of the previous section shows that for a fixed ratio of component grating speeds, r_1 / r_2, as the difference between the directions of the two component gratings, $\theta_1 - \theta_2$, decreases and the shape of the blobs become more elongated, the orthogonal direction of motion of the longer edges of the blobs, ϕ, will tend towards the VS direction of the component gratings. Thus the majority of the local motion detectors will signal motion in a direction which is increasingly biased, as $\theta_1 - \theta_2$ decreases, towards the VS direction.

In the second stage of the model, it is proposed that the outputs of the local motion detectors are combined using a recursive Bayesian estimation process. The estimate formed in the first iteration of the estimation process will thus form the model's prediction of the perceived plaid velocity in a short initial period of stimulus presentation. As has been already discussed, this estimate will be dominated by the majority of local motion detectors which signal the orthogonal motion of the blobs in the ϕ direction. It has also been shown, in equations (35) and (36) respectively, that as the difference between component grating directions, $\theta_1 - \theta_2$, decreases: (i) the long edge of the blob will become longer and therefore drive an increasing majority of local motion detectors; and (ii) the orthogonal motion of the blobs in the ϕ direction approaches the VS direction of the component gratings. Hence it follows that, as the angle between the component gratings decreases, the first velocity estimate formed by

the model, and therefore the initial plaid velocity predicted by the model, will be increasingly biased towards the ϕ direction, which itself will approach the VS direction of the component gratings. This is precisely what Yo and Wilson (1992) observed in their psychophysical experiments.

For example, consider one of the Type II plaids used by Yo and Wilson in their experiments. The parameters of the component gratings of this plaid are: $\theta_1 = 70.5^\circ$, $\theta_2 = 48.2^\circ$, $v_1 = 1.33$, and $v_2 = 2.67$. Then, for these values: $\theta_1 - \theta_2 = 22.3^\circ$, $\theta_{IOC} = 0.2^\circ$, $v_{IOC} = 3.9$, $\theta_{VA} \approx 55.6^\circ$, $v_{VA} = 4.0$, $\phi = 59.4^\circ$, $\phi - \theta_{VA} = 3.8^\circ$ and $s_\phi / s_\varphi = 5.1$. The blobs are thus elongated (edge ratio of 5:1) and move orthogonally to their longer edges in a direction which is less than 4° from the VS direction of the component gratings. In Yo and Wilson's experiment, the perceived direction of the plaid motion in the initial period of presentation was observed to be approximately 60°. This is close to the VS direction of 55.6°, and almost exactly equal to the orthogonal direction $\phi = 59.4^\circ$ of motion of the longer edges of the blobs.

The velocity estimate formed by the model during subsequent iterations of the recursive estimation process will also be influenced by the majority of local motion detectors which signal the orthogonal direction ϕ of the longer edges of the blobs, although this influence will gradually decrease with each iteration (see the simulation model description below). Thus for long stimulus presentations the perceived direction of the plaid motion predicted by the model will continue to be biased, but to a lesser extent, in the direction $\phi = (\theta_1 + \theta_2)/2$, the mean of the component gratings' directions. This is precisely what Ferrera and Wilson (1990) observed, i.e. that the perceived direction of the plaid motion has a small residual bias, after approximately 150ms of presentation time, of between 8° and 10° towards the mean of the component gratings'

directions, in this case for plaids with component grating separations of between 22.3° and 51.6°. A similar residual bias was observed by Burke and Wenderoth (1993). They found in addition that as the difference in component grating directions decreased from 40° to 10°, the residual bias increased from 2° to 17°. This dependence of the residual bias on the difference in component grating directions was observed for a constant value of $\phi = (\theta_1 + \theta_2)/2$. Hence they argued that the bias could not be due to the orthogonal direction of motion of the elongated blobs which remained constant in this experiment. In the model however the strength of both the initial and the residual bias is determined by the length of the long edges of the blobs, since this determines the number of local motion detectors which signal the orthogonal 1-D motion of the blob edges in the ϕ direction. Since the elongation of the blobs increases with decreasing difference in the direction of motion of the component gratings, as shown by (35), it follows that the residual bias will always be towards the ϕ direction, but will increase as the difference in component grating directions decreases.

The theoretical predictions of the model, presented above, are largely qualitative in nature, but will be confirmed in a more quantitative form in the next chapter, where we describe the results from using of the computational version of the model, as described in Chapter 2, to simulate the perceptual experiments of Yo and Wilson (1992), Bowns (1996) and Burke and Wenderoth (1993).

7.3. Summary.

In this chapter the basis of the application of the model to plaid perception was presented. First, the geometric analysis of the plaids' blobs was performed by using the product gratings, instead of the component gratings which are usually used to represent the plaids. The goal of this analysis was to define better the distribution of the 1-D and

2-D motion cues represented within the plaid in order to make clear their importance for the model. It was shown that the shape and size of the blobs depends on the angle between the component gratings directions of motion and on the spatial frequencies of the product gratings. The predictions of the model regarding the effect of the 1-D motion signals from the edges of the blobs were shown to correspond to the psychophysical data on the subject.

Chapter 8. Simulation of plaid perception experiments using the motion integration model.

8.1. Description of the model's behavior.

To quantify the predictions of the model and, in particular, to demonstrate the convergence of the estimated direction of the Type II plaid motion towards the true IOC direction, a computational version of the model described in Chapter 2 will be used to simulate the psychophysical experiments of Yo and Wilson(1992), Bowns (1996) and Burke and Wenderoth (1993). In particular the optimal Kalman filter based model will be used, rather than the approximate version of the model. However given the closely related results which were obtained from these two versions of the model when applied to smooth eye pursuit, similarly closely related results might be expected from the two models when applied to plaid motion perception.

As has already been described, the bias in the initial estimate of the plaid direction results from the large number of local motion detectors for which the measured derivatives I_x, I_y, I_t correspond to the 1-D motion of the longer edges of the blobs. For these detectors many solutions to the corresponding gradient-based equations (3) are possible, corresponding to the aperture effect (Wallach, 1935; Wuerger, Shapley and Rubin, 1996, Marr and Ullman, 1981). The zero-valued initial velocity estimate provides a constraint on the estimate formed by the first step of the algorithm, which results in an best-fit solution being selected for which the magnitude of the velocity estimate is smallest. This corresponds to the solutions to (3) for each local motion detector for which the selected velocity is in the direction orthogonal to the longer edges of the blobs. Thus the estimate formed in the first step of the algorithm will be strongly biased in this direction, with the strength of the bias dictated by the number of motion

detectors signalling the direction. As has been shown, the bias will be stronger as the difference between the directions of the component gratings decreases, since this results in a greater elongation of the blobs.

In contrast, measurements of I_x, I_y, I_t from the local motion detectors which signal the 2-D motion of the end-points of the blobs result in a unique (within the noise) solution to the corresponding set of gradient equations (3). This solution corresponds to the VS direction of the product gratings, and thus, equivalently, to the veridical, IOC direction of the plaid. These local motion detectors will therefore influence the estimate of plaid direction towards the IOC direction, both in the initial step of the algorithm and in all further steps. However, lowering the contrast of the plaid stimulus, or equivalently reducing the signal to noise ratio in equation (3), will result in a weaker influence of this solution, and thus allow a greater bias in the estimated direction of the plaid towards the VS direction.

As the number of iterations of the recursive estimation algorithm increases, the effect of the 1-D local motion detectors will decrease relative to that of the 2-D motion detectors, since the velocity estimate formed in each iteration of the algorithm becomes the "prior" estimate for the next iteration. This gradually relaxes the effect of the zero prior constraint on the solution to (3) corresponding to the set of outputs of the 1-D motion detectors, allowing the solution to (3) corresponding to set of outputs of the 2-D motion detectors to increasingly influence the velocity estimate in each iteration.

In the following section simulations of a computational version of the model will show that, in accordance with the above theoretical predictions, the model also yields quantitative predictions of the perceived direction of plaid motion which closely resemble the experimentally obtained data of Yo and Wilson (1992), Bowns (1996) and Burke and Wenderoth (1993). Also a more general set of simulation results will be given for the estimated velocities of plaids corresponding to a range of different

component angular differences and speed ratios. The stimuli used in the simulations were sinusoidal plaids filtered by a circular Gabor filter with a diameter of 80 pixels. The variance of the measurement noise was set equal to 0.1, and the velocity estimation covariance was set equal to 0.04. Both these values were kept constant throughout all the simulations described in the following sections.

8.2. Simulation of the experiments of Yo and Wilson (1992) and Bowns (1996).

Figures 20a - 20f show the results obtained from the computational model when simulating the psychophysical experiments of Yo and Wilson (1992) and Bowns (1996). Yo and Wilson (1992) used as stimulus three different Type II plaids (see their Figure 2), although the results were not given for all three plaids for each of the experiments. The main experiments, which are simulated here, recorded the perceived direction of plaid motion as functions of presentation duration and pattern contrast. Their results on presentation duration are given for the plaid with the following parameters: $\theta_1 = 70.5^\circ$, $\theta_2 = 48.2^\circ$, $r_1 = 1.33$, $r_2 = 2.67$, $s_1 = s_2 = 1.5$, $\theta_1 - \theta_2 = 22.3^\circ$, $\theta_{IOC} = 0.2^\circ$, $r_{IOC} = 4.02$, $\theta_{VS} = 55.6^\circ$, $r_{VS} = 3.93$. For the product plaid representation, these parameters give: $\phi = 59.4^\circ$, $\phi - \theta_{VS} = 3.8^\circ$, $r_\phi = 2.0$, $r_\varphi = 3.5$, $s_\phi = 1.5$, $s_\varphi = 0.3$ and $s_\phi / s_\varphi = 5.0$. Note that both the speed and spatial frequency of the higher spatial frequency product grating are similar to those of the component gratings.

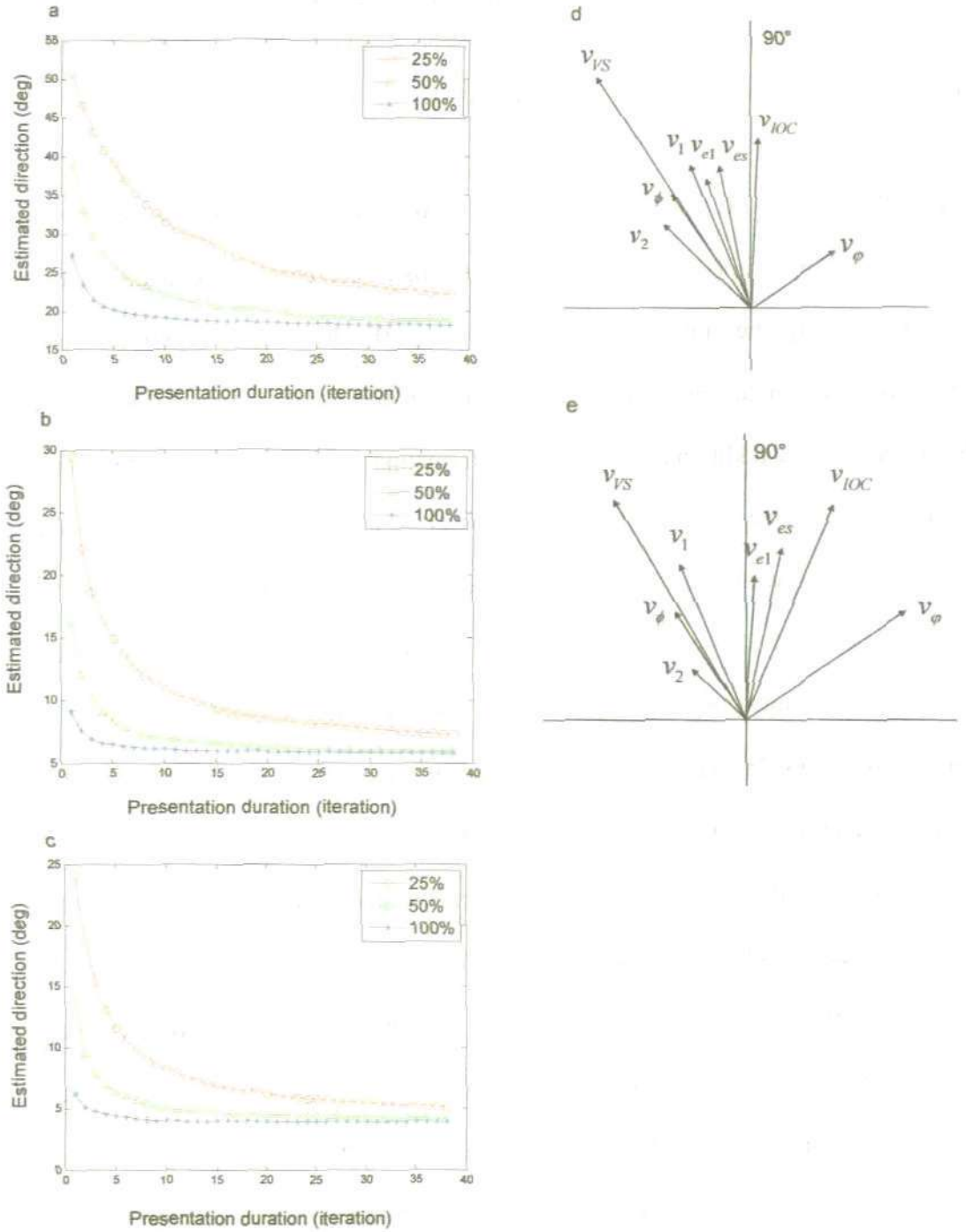


Figure 20. a-c: Simulations of the computational model for three cases of Type 2 plaids used in the experiments of Yo and Wilson (1992):

a. $\theta_1 = 70.5^\circ, \theta_2 = 48.2^\circ, v_1 = 1.33, v_2 = 2.67, \theta_{IOC} = -0.2^\circ, \theta_{VS} = 55.6^\circ$;

b. $\theta_1 = 84.3^\circ, \theta_2 = 36.9^\circ, v_1 = 0.25, v_2 = 2, \theta_{IOC} = 0^\circ, \theta_{VS} = 41.7^\circ$;

c. $\theta_1 = 85.2^\circ, \theta_2 = 33.6^\circ, v_1 = 0.4, v_2 = 4, \theta_{IOC} = 0^\circ, \theta_{VS} = 37.8^\circ$.

and for three different stimulus contrast levels. The results illustrate the dependence on stimulus contrast of the initial and final estimates of plaid direction, and of the convergence rate of the estimate towards the true IOC plaid direction. Presentation duration is represented by the number of iterations of the algorithm.

d-e: Vector space diagrams showing the model simulation results for Experiment 3 of Bowns (1996). The initial v_{e1} and final v_{es} plaid velocity estimates from our model are shown together with the component grating velocity vectors v_1 and v_2, and the vector sum and IOC velocity vectors.

For this plaid (Yo and Wilson, 1992), two observers reported a perceived direction of motion of approximately 60° after 60 ms of presentation, reducing to 15° and 30° respectively after approximately 90 ms, and to 0° (the IOC direction) after 150 ms. The reduction in the bias was gradual rather than occurring discontinuously as might be expected if additional 2-D motion information became available after some fixed time delay, as was suggested in the Wilson *et al.* (1992) model. When the plaid contrast was varied, with values of 5%, 50% and 100%, the observed initial bias at 60 ms was 60°, 40° and 30° respectively, and the length of time for the bias to reduce lengthened considerably with decreasing contrast. For a contrast of 5-10%, a substantial bias of approximately 25° was observed after 1 sec of presentation. These experimental results can be compared with the graph shown in Figure 20a, which shows the results of simulating the experiments with the motion integration model. As the graph shows, the bias in the model's estimated direction at the first iteration, for the three values of contrast, 25%, 50% and 100%, are remarkably similar to the initial perceived bias observed experimentally. We note also that the convergence time decreases substantially with increasing contrast, and that there is a considerable steady-state bias for all contrasts of up to 25° for this plaid, again as observed experimentally. Figures 20b and 20c show the same simulations for the other two plaids used by Yo and Wilson (1992), but for which they did not report the results so fully as for the first plaid. These graphs show similar characteristics of the variation in magnitude and convergence rate of the direction bias with contrast as in Figure 20a, but with the steady-state bias reducing with increasing difference in the directions (47.4° and 51.6° respectively) of the component gratings in Figure 20b and 20c, to between 4° and 10°. In Ferrara and

Wilson (1990), the perceived steady-state bias for similar Type II plaids was approximately 6°.

In Bowns (1996), a number of experiments were carried out which attempted to establish whether or not the misperception of the plaid direction observed by Yo and Wilson (1992) generalises to all Type II plaids and is due to a temporal delay in Fourier and non-Fourier motions processing as proposed in the parallel pathway model of Wilson *et al* (1992). The simulations described here are of their Experiment 3, which used Type II plaids very similar to those used by Yo and Wilson (1992). The component gratings for these plaids had the same spatial frequencies (1.3 cycles/°) and orientations (202° and 225°) but differed in the ratio of their speeds, which ranged from 1:0.45 to 1:0.75, with the speed of one of the component gratings held constant at 3.13 °/sec. The experiments used a simple forced choice response which required subjects to report either a plaid direction to the right or to the left of "the vertical", i.e. 90°. The component grating directions and speeds were such that the vector sum direction remained virtually constant, varying from 29° to 32° to the left of the vertical, for the varying speed ratios, whereas the IOC direction varied from 28° to 2° to the right of the vertical.

The experiments revealed that for the two speed ratios at the extreme ends of the above range, subjects reported a perceived direction of plaid motion which shifted from 100% in the vector sum direction (i.e. left of vertical), for a speed ratio of 1:0.75, to 100% in the IOC direction (i.e. right of vertical), for a speed ratio of 1:0.45. This was interpreted in Bowns (1996) as: "a rather surprising complete reversal of the perceived motion in the direction of the IOC".

The simulations now described are for the cases of the two plaids at the extremes of the ranges of speed ratios referred to above. This experimental data was also simulated by Weiss and Adelson (1998) - see the discussion of their model in section

9.1. The simulation results obtained from the computational model are described in Figure 20d (for a speed ratio of 1:0.75) and 20e (for a speed ratio of 1:0.45) in the form of vector space diagrams. As these Figures show, changing the ratio of the component grating speeds from 1:0.75 to 1:0.45 is sufficient to move the both the estimate formed in the first step of the estimation algorithm, v_{e1}, and the steady- state estimate, v_{es} of the perceived plaid direction from being on the left of the vertical (VS side) to being on the right of the vertical (IOC side).

The difference in the directions of the first step velocity estimate v_{e1} for the two speed ratios is 21° (108° vs. 87°). However the difference in the first step direction bias estimate (relative to the IOC direction) is only 5° (20° from IOC vs. 25°). For the steady-state velocity estimate v_{es}, the estimated direction differs by 24° (102° vs. 78°) for the two speed ratios, but the difference in the estimated bias is only 2° (14° from IOC vs. 16°).

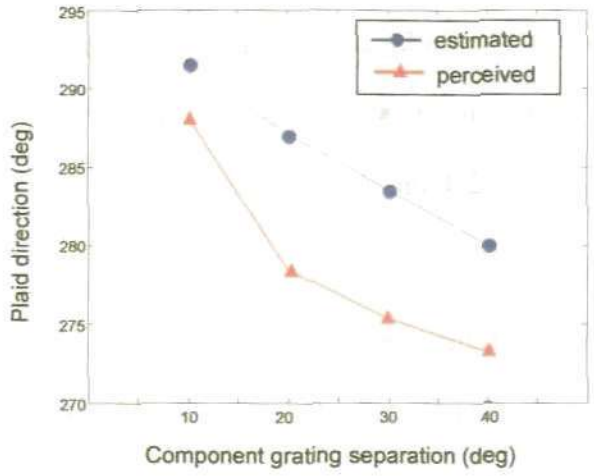
Thus the change in the estimated bias is small with this change in speed ratio, both in the first step of the algorithm and after convergence, and we suspect that the change in the perceived bias is also small. The simple forced choice response of left or right of the vertical appears however to have resulted in an interpretation in Bowns (1996) that there is a large change in bias which leads to a reversal in the perception of the plaid motion direction from the IOC to the vector sum direction.

An alternative interpretation is suggested by the simulation results (see Figures 20d and 20e), i.e. that the value of the perceived bias for the two speed ratios is almost the same, but that the change in speed ratio results in a shift in the IOC direction towards the VS direction, causing the perceived motion direction to switch from right side of the vertical to the left side.

8.3. Simulation of the experiments of Burke and Wenderoth (1993).

Figure 21 shows the results obtained using the computational model to simulate the psychophysical experiments of Burke and Wenderoth (1993), in which they used Type II plaids to study the dependence of the steady-state misperception of plaid direction on the angular difference between the component grating directions.

a



b

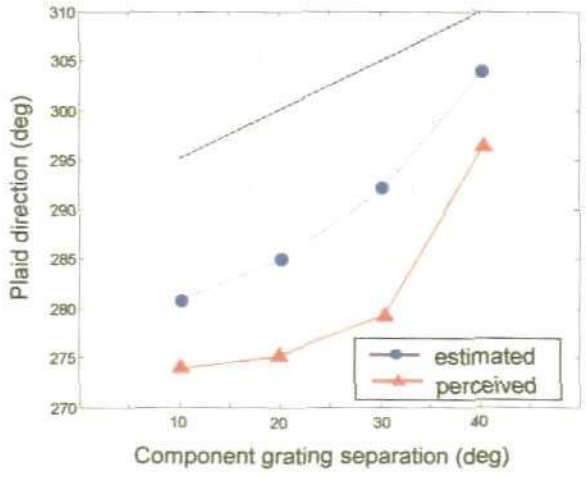


Figure 21. Results from simulations of the computational model for the plaids used in the experiments of Burke and Wenderoth (1993), showing in (a) and (b) both the perceived plaid direction obtained in the experimental study (▲symbols) and the plaid direction estimated by the model (● symbols). In (a) the mean component direction is constant equal to 295°, and in (b) the mean component direction varies and is shown by the black dashed line. The graphs (a) and (b) show that the plaid direction estimated by the model varies with the difference in component grating direction and displays in both cases the same trend

in the variation as observed in the Burke and Wenderoth (1993) study, although with a slightly greater bias towards the mean component direction of up to 9°. From the original diagram for the experimental results in this study, perceived errors were in the region of ±3°.

The plaids were constructed using component gratings with angular differences of 10°, 20°, 30° and 40°. The true plaid direction was 270° and the stimulus was presented for 10 sec. Two experiments were carried out: in the first, the component directions were chosen so that the mean was constant at 295°; in the second, one component direction was kept constant at 315°. Figures 21a and 21b show the results from each experiment, both the perceived plaid direction obtained in the Burke and Wenderoth (1993) study (▲symbols) and the direction estimated by the model (● symbols). In Figure 21a the mean component direction is 295°, and in Figure 21b this direction varies and is shown by the dashed line.

The graphs in Figure 21 show that the estimated plaid direction from the model simulation varies with the difference in component grating direction and displays in both cases the same trend in the variation as observed in the Burke and Wenderoth (1993) study, although with a slightly greater bias towards the mean component direction of up to 9°. Importantly the model shows in Figure 21b the same non-linear variation of the estimated direction with component separation as was observed experimentally for the perceived direction.

8.4 Robustness of the model

In the above cited experiments and those that are described later in Chapter 9, the stimuli were presented in a circular windows with the following diameters: Yo and Wilson (1992) - diameter = 8°; Bowns (1996) - diameter = 3°; Stone, Watson, and

Mulligan (1990) - diameter = 5.4° ; Champion, Hammett,and Thompson (2007) - diameter = 6° ; Alais, Wenderoth, and Burke (1997) - diameter = 3°, 6° and 12°. Information on the size of the stimulus used in the experiments of Burke and Wenderoth (1993) is not given in their paper. In the simulations described in Sections 8.2 and 8.3 and in Chapter 9, the image is displayed in a circular aperture of diameter 200 pixels; thus the size of the 10 x 10 pixel window used in the model corresponds to between 0.15° and 0.6°. This is in close accordance with an average receptive field diameter measurement, for V1 cells in humans, of approximately 0.25° at the fovea, rising linearly to approximately 0.6° at 6° eccentricity (Smith, Singh, Williams and Greenlee, 2001).

Since the simulation results presented here closely match the experimental results in each of these experiments, it can be inferred that the model results are robust if the 10 x 10 pixel window represents a receptive field diameter of between 0.15° and 0.6°, which is the approximate physiological range for V1 cells.

The computational model breaks down when the simple algorithm used to calculate the image intensity derivatives fails to produce acceptably accurate results. This happens when the spatial frequency of the stimulus is sufficiently high that the spatial period falls within a single window, i.e. is less than 10 pixels, corresponding to a frequency of 0.1 cycles/pixel, or between 6.7 cycles/° (corresponding to window size of 0.15° and a stimulus aperture diameter of 3°) and 1.6 cycles/° (corresponding to a window size of 0.6° and a stimulus aperture diameter of 12°). Thus, for the simulations of the Alais *et al* (1997) experiments described in Chapter 9, in which the aperture diameter is 3°, the result for a stimulus of 6 cycles/° was not simulated.

It is important to note that the free parameter of the model σ / σ_p was held constant for all the simulation results described in this Chapter, i.e. for the Yo and

Wilson (1992), the Bowns (1996), and the Burke and Wenderoth (1993) experiments, and the general simulation experiments in the next Section.

8.5. General simulation experiments.

The model was also simulated for a set of different speeds and directions of the components gratings. The range of the directional difference between the components was between 10° and 170° degrees. The first grating was oriented toward 180° and kept constant. The difference between the directions of the two gratings was gradually increased by 10° for each subsequent experiment thus varying the angular difference between 10° and 170° degrees. The speeds of the components varied from 1 to 5 pixels/iteration. The data is presented in Figure 22, which shows the results for four different directions of the components, with angles between them equal to 10°, 60°, 90° and 170°. The symbols on the plot correspond to VS direction (red triangle), IOC (red circle), the orthogonal direction to the long edge of the blob (black dot), initial estimated direction for the low contrast plaid (green square), initial estimated direction for the high contrast plaid (blue star) and the final estimation (40 iterations) for the high contrast stimuli (blue pentagram). The speed ratio is presented in a logarithmic scale. When the direction angle between the components is equal to 10° (Figure 22a) the results show that the bias is toward the VS direction which approximates the direction orthogonal to the long edge of the blob in this case. As the ratio of the component speeds increases, the IOC direction moves away from the VS direction, and hence the predicted bias increases. These results are thoroughly in agreement with the experimental data. Note that for the case of 60° angular difference (Figure 22b) the difference between the VS and IOC is generally smaller and more dependent on the speeds ratio than in the case shown in Figure 22a.

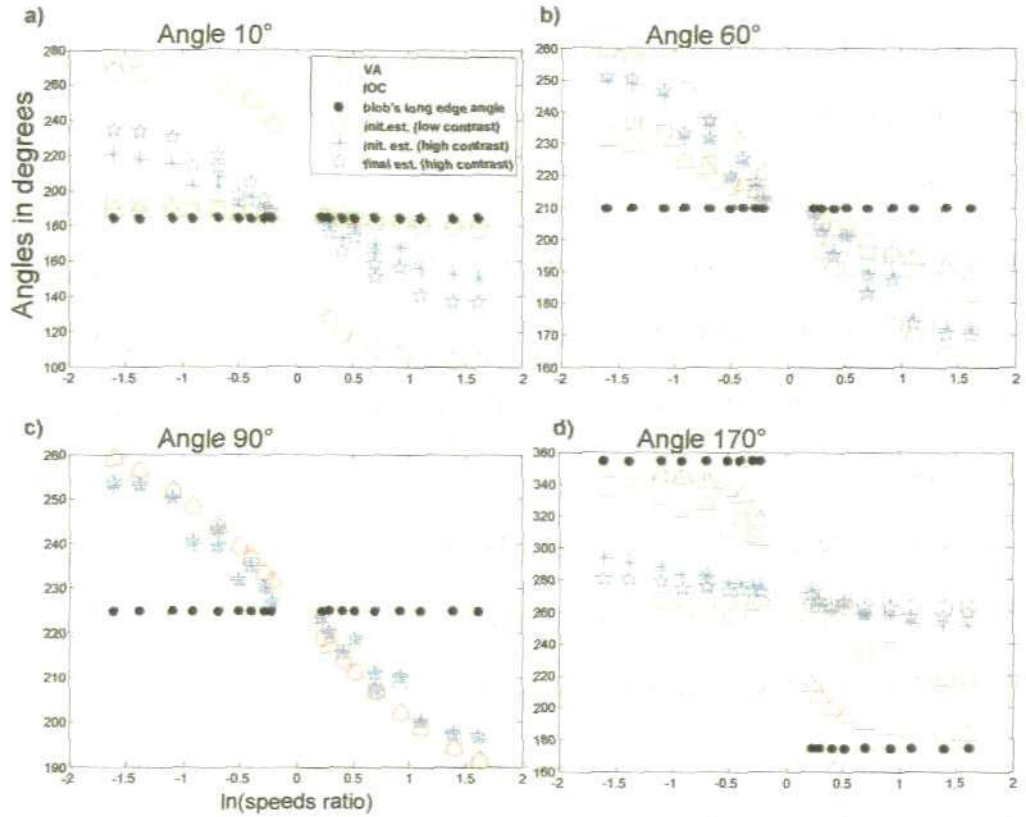


Figure 22. Simulations of the model for 4 different angles (10, 60, 90, 170°) between the directions of the components of the plaid. The first component grating's velocity was oriented at 180° and kept constant. The second grating's velocity direction varied between 190° and 350°. The speed ratio is in logarithmic scale. The absolute values of the speeds varied between 1 and 5 pixels/ step. The symbols are: Δ VS (VA) direction; ○ IOC direction; • Orthogonal direction to the long edge of the blob; □ Initial estimation of the velocity direction for low contrast; * Initial estimation of the direction of the velocity for high contrast; ☆ Final estimation of the velocity direction for high contrast.

The psychophysical experiments are consistent with this result, i.e. that the perceived bias is particularly bigger for angles between components close to 10°. For angles of 90°, as expected the estimation is in the IOC /VS direction, which overlap in this case (Figure 22c). The results are also presented for the angle of 170° (Figure 22d) as it shows the model's predictions for Type I plaids in this most extreme case of the simulations. As far as the author is aware, there is no evidence from psychophysical experiments confirming a bias toward the non-veridical direction, i.e. VA instead of

IOC for the Type I plaids. However, the simulations of the model presented here predict a substantial bias in the initial velocity estimate for the low contrast case, suggesting that for low contrast and short presentation time there should be an experimentally perceived bias for such asymmetric Type I plaids. As a conformation for the requirement of short stimulus presentation in order to detect a non-veridical bias for Type I asymmetric plaids, it can be noticed in Figure 22d that the final estimate is much closer to the IOC direction compared to the results in Figure 22a. This leads to a possible conclusion that the conversion time to the veridical velocity is shorter for Type I plaids as compared to Type II plaids.

There are differences between the case when the angle between the components is equal to 10° (Figure 22a) and 170°, even if the blob's size and shape are the same. The first difference is between the magnitudes of the IOC speeds for both cases. For angles >> 90° and different components speeds, which is the asymmetric Type I plaid, the ratio between the IOC and VS speeds could be much bigger than for the case of 10°. For instance, for speeds of 1 and 2 pixels/iterations of the components, the ratio (IOC speed)/ (VS speed) is equal to 2 for 10° between the components, but for 170° the same ratio is equal to 17. This means that the veridical speed becomes much larger in comparison with the speed of the edge of the blob. In particular, for the angle of 170°, which is presented in Figure 22d, the motion in the IOC direction is almost parallel to the orientation of the blob's long edge and therefore contributes less to the orthogonal motion of this edge. This is a possible explanation of why there was no evidence of a bias in Type I plaids from psychophysical experiments. The bias presented in our simulation results could be also just an artifact resulting from the insensitivity of the simulation set-up to detect higher speeds, which would lead to a smaller effect of the IOC speed on the simulation results as compared to the 1-D motion effect of the blob edge. However, if the higher IOC speed results in a more pronounced 1-D motion of the

edge (as a projection of a bigger vector, even if the angle of projection is relatively small due to the almost parallel orientation of IOC vector toward the edge of the blob) according to the simulation results in Figure 22d it could be speculated that under certain conditions such a bias could be found and they are: 1) low contrast, 2) as short as possible presentation, 3) bigger angle between the components directions of motion (155-175°) and 4) a speed ratio significantly smaller than 1 (which corresponds to zero in the logarithmic scale in Figure 22d). However, as far as the author is aware, such psychophysical data have not been produced so far, i.e. for Type I asymmetric plaids with angles between the components (directions of motion) greater than 90°.

8.6. Summary.

In this chapter a more detailed description was presented of how the model calculates velocity from the pool of 1-D and 2-D motion signals in the case of a plaid stimulus. Furthermore, it was shown that the model simulations of plaid perception, presented in detail in this chapter, showed a close resemblance to the psychophysical data of Yo and Wilson (1992), Bowns (1996) and Burke and Wenderoth (1993). In addition, the model was applied to a wide range of possible velocities of the component gratings of the plaids, thus representing the results for both Type I and Type II plaids. For Type II plaids the simulation results were in agreement with the experimental data. For Type I plaids however, the model predicts a possible bias toward the non-veridical direction for a well-chosen set of conditions.

Chapter 9. Discussion of the application of the motion integration model to plaid perception.

9.1. Comparison with the Adelson and Movshon (1982) model.

The original two-stage model (Adelson and Movshon, 1982; Movshon *et al.*, 1985)) has dominated research in plaid motion perception for almost thirty years, leading to an almost universal view that the first stage of plaid motion analysis involves the detection of the 1-D motion of the component gratings, carried out by "component-direction selective" neurons in V1 (see the review by Pack and Born (2008)). It is important to note however that the available evidence is almost entirely based on using symmetric Type I plaids, in which the component gratings move with equal speeds. For the psychophysical experiments of Movshon *et al.* (1985) the difference in directions of the component gratings was 120°, for their physiological experiments in cat and monkey V1 and in monkey MT (Movshon *et al.*, 1985) the angular difference was 90°, and for Movshon and Newsome's (1996) physiological experiments in monkey V1 the difference was 90° or 45°. For such plaids, a neuron in primary visual cortex (V1) which responds optimally to the motion of a single grating, produces little response to a plaid moving in its optimal direction, as would be predicted from the orientations of the component gratings if the neurons were only responding to the 1-D motion of the gratings (Movshon *et al.*, 1985).

The model proposed in this thesis suggests that neurons in V1 respond both the 1-D and 2-D motion of the blob features of the plaid, and in the case of Type II plaids are driven by the 1-D edges and 2-D end-points of the elongated blobs. Moreover, it is proposed that the 2-D blob motion is detected by end-stopped cells in V1, as observed

by Pack *et al.* (2003). This model leads to theoretical and simulation results which closely mimic the physiological observations of perceived direction for such plaids. So how does the model explain the "component-selective" responses for V1 neurons in the case of Type I plaids, as observed by Movshon *et al.* (1985) and Movshon and Newsome (1996), in particular as the neurons observed by Movshon and Newsome (1996) were apparently mostly of the end-stopped variety ?

For Type I plaids in which the difference in the component grating directions is 90°, the blobs take the form of small square regions of high luminance which are aligned in the same orientations as the component gratings. Therefore, a neuron which is optimally responsive in the direction of the plaid motion, and with a long, narrow receptive field oriented orthogonally to the plaid direction, will respond sub-optimally to the two lines of blobs, each moving at 45° to the optimal direction for the neuron, in exactly the same way as if it were responding to the component gratings themselves, as shown by Tinsley *et al.*, 2003. Little or no 2-D motion signal in the direction of the plaid would be detected due to the absence of well defined end-points in the stimulus, in contrast to the case of Type II plaids with elongated blobs. It is significant however that Movshon and Newsome (1996) observed a degree of "pattern-selective" response in two of the nine neurons they measured. Thus, for such neurons and for symmetric Type I plaids, it is not possible to distinguish whether the neurons are responding to the component gratings or to the lines of small square blobs ("blob-lines") present in the plaid moving in the same directions as the component gratings. The lines formed by the blobs are certainly more perceptually salient to the human observer than the individual component gratings.

In the model proposed here, the outputs of the local motion detectors signalling the two orthogonal 1-D motion of the "blob-lines" described above will be combined in the second stage by the estimation algorithm to yield the VS of the two directions,

which corresponds exactly to the direction of motion in the case of a Type I plaid. Note that no initial or steady-state perceived direction bias was observed for Type I plaids by Ferrera and Wilson (1990) or Yo and Wilson (1992). It is also possible that the 2-D motion of the individual blobs may be signalled by V1 neurons with short, wide receptive fields, as observed by Tinsley *et al.* (2003). The combination of the outputs of the 2-D motion detectors and the 1-D motion detectors in the second stage of the model would reinforce the computation of the velocity estimate in the true plaid direction.

In summary, it is argued here that for symmetric Type I plaids, with a difference between the component grating directions of around 90°, the 1-D motion detectors in V1 will respond in exactly the same way to the blob-lines as to the component gratings. Since the blob-lines and the component gratings are indistinguishable, in terms of their orientation, direction, spatial frequency and speed, it is impossible for any experiment with such Type I plaids to distinguish between a model in which the first stage responds to the 1-D motion of the component gratings and one in which the first stage responds to the 1-D motion of the blob-lines. Since, in addition, the direction of a symmetric Type I plaid is given by the simple averaging (VA) of the 1-D motion directions, it is impossible to distinguish between a model in which the second stage computes the IOC direction from one in which the second stage computes the VS direction. The conclusion is therefore that the psychophysical experiments (Welch, 1989; Derrington & Suero, 1991; Derrington & Badcock, 1992) using symmetric Type I plaids, which have apparently confirmed the two-stage model of Adelson and Movshon (1982), are wholly inadequate in this respect. In contrast, the psychophysical experiments with Type II plaids (Yo and Wilson, 1992) strongly challenge the Adelson and Movshon model.

It is worth noting here that the theoretical analysis of the plaid blobs in Section 7.1 of the thesis indicates, for asymmetric Type I plaids with an angular separation of

component directions of >90° (e.g. the plaid in Figure 19c), that a similar elongation of the blobs occurs, and that the longer edges increase in length as the angular separation increases. Also the orthogonal direction of the longer edges of the blobs approaches the VS direction. The motion integration model predicts a significant bias in the perceived direction of plaid motion towards the VS direction for certain conditions as low contrast and short presentation, which could be of comparable magnitude to that observed for Type II plaids. As far as the author is aware, no psychophysical or physiological experiments have been carried out for such Type I plaids.

9.2. Comparison with the Weiss *et al.* (2002) model.

A recent model of motion integration (Weiss and Adelson, 1998; Weiss *et al*, 2002) aimed at extending the Adelson and Movshon (1982) model to accommodate the Yo and Wilson (1992) results. According to Weiss and Adelson (1998) and Weiss *et al* (2002), their model captures the uncertainty in the 1-D motion of the component gratings in the case of low contrast by using a Bayesian estimation process. The Bayesian formulation of the model results in the identification of a distribution of 1-D and 2-D velocity measurements which correspond to local likelihood functions. The model therefore represents the 1-D motion of each of the component gratings, corresponding to the first stage of the Adelson and Movshon model, as a pair of "fuzzy" (Weiss and Adelson, 1998) constraint lines in velocity space, the degree of fuzziness being dependent on contrast. The estimate of the plaid direction is then given by the mean/maximum of the posterior probability distribution, which is computed from the product of the local likelihoods and the prior distribution for the velocity estimate. The latter is assumed to be Gaussian with zero mean according to a "slow and smooth" (Weiss and Adelson, 1998; Weiss *et al*, 2002) hypothesis based on suggestions that

human observers prefer the slowest motion consistent with the visual input (Ullman, 1979).

In fact, the model described by Weiss and Adelson (1998) and Weiss *et al* (2002) is identical to the first step of the recursive Kalman filter estimation algorithm in the model which has been presented here, and therefore produces an identical, biased first step estimate of plaid direction. There appears therefore to be a contradiction between the explanation in Weiss and Adelson (1998) of the behaviour of the model in predicting plaid motion, which is solely in terms of the 1-D motion of the component gratings, and the explanation which was presented here in Sections 7.2 and 8.2, which is in terms of the 1-D and 2-D motion of the edges and end-points of the blobs. The explanation in Weiss *et al* (2002) is essentially the same as that in Weiss and Adelson (1998) but less detailed and with no supporting diagrams.

To resolve this contradiction, first consider the plaid used to produce the simulation results shown in Figure 20a, and previously discussed in Sections 7.2 and 8.2. This plaid is also used in Weiss and Adelson (1998) and Weiss *et al* (2002) as their main example for demonstrating the misperception of the direction of Type II plaids. The parameters of the component gratings of this plaid are, as given before: $\theta_1 = 70.5^\circ$, $\theta_2 = 48.2^\circ, r_1 = 1.33$, and $r_2 = 2.67$, yielding the following values: $\theta_1 - \theta_2 = 22.3^\circ, \theta_{IOC} = 0.2^\circ$, $r_{IOC} = 3.9$, $\theta_{VS} = 55.6^\circ$, $r_{VS} = 4.0$, $\phi = 59.4^\circ$, $r_\phi = 2.03$,

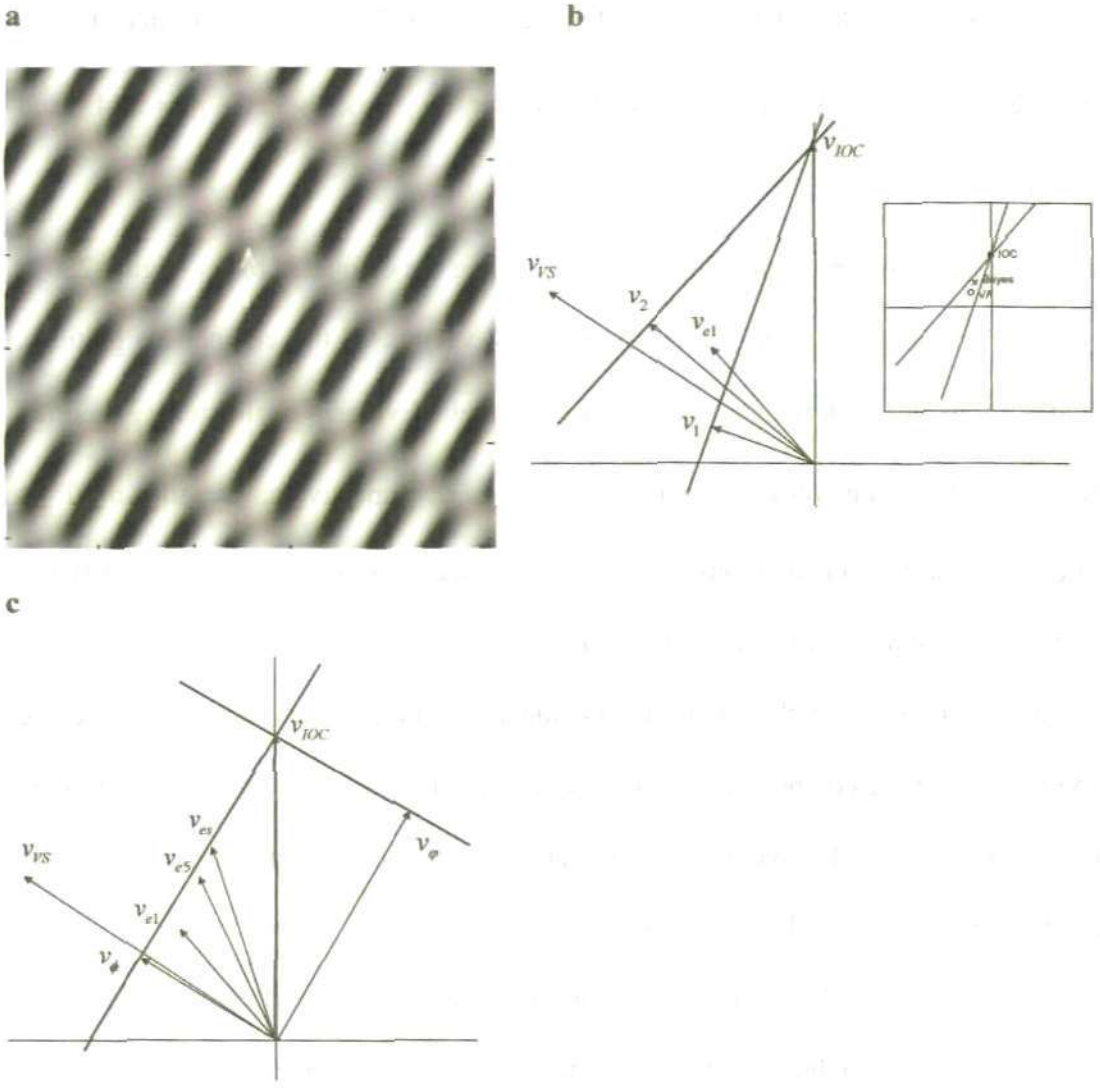


Figure 23. Velocity space diagrams of the plaid used in the experiments of Yo and Wilson (1992) and for which the model simulation results are shown in Figure 20a. a) diagram illustrating the plaid, clearly showing the elongated blobs; b) velocity space diagram on which the velocity vectors, v_1 and v_2, of the component gratings, together with their constraint lines, the IOC and vector sum velocity vectors v_{IOC} and v_{VS}, and the velocity estimate from our model for the first step, v_{e1}, are shown (0° is vertically upward in this diagram). The inset diagram. is redrawn from Figure 15d of Weiss and Adelson (1998); c) velocity space diagram showing the velocity vectors corresponding to the motion of the longer and shorter edges of the blobs in the plaid, v_ϕ and v_φ, together with their constraint lines, the IOC and vector sum velocity vectors, and the velocity estimates from our model for the first step, v_{e1}, the fifth step v_{e5}, and in the steady-state v_{es}. An explanation of the diagrams is given in the text.

$r_\varphi = 3.46$, $\phi - \theta_{VS} = 3.8^\circ$ and $s_\phi / s_\varphi = 5.1$. For a contrast of 50%, the model calculates the initial estimate of the plaid velocity vector v_{e1} as: speed $r_{e1} = 1.84$ and direction $\theta_{e1} = 40^\circ$.

Figure 23a illustrates the plaid, clearly showing the elongated blobs. Figure 23b is a velocity space diagram on which the velocity vectors of the component gratings, v_1 and v_2, together with their constraint lines, the initial velocity estimate, v_{e1}, and the IOC and vector sum velocity vectors, v_{IOC} and v_{VS} respectively, are shown (0° is vertically upward in this diagram). The velocity space diagram in Figure 15d of Weiss and Adelson (1998) is redrawn as an inset in Figure 23b.

In their Figure 15d, the latter authors indicate the magnitude of the vector average (VA) velocity of the component gratings, rather than the vector sum velocity. Although the *direction* of these two velocity vectors are the same, the *magnitude (speed)* of the vector average velocity is half that of the vector sum velocity.

Weiss *et al* (2002) explain the bias in the perceived direction towards the vector sum/average direction by the statement that "the vector average velocity [speed] is much slower than the IOC solution and hence it is favored [by the zero prior] at low contrasts". They assume that the Bayes estimate of plaid velocity is based on "local likelihoods [which] are 'fuzzy' constraint lines" (Weiss and Adelson, 1998) defined by the component grating velocities. Thus any bias in the estimate towards a speed slower than the IOC speed, as a consequence of the zero prior, will automatically result in a shift of the direction of the estimated velocity away from the IOC direction and towards the vector sum/average direction, i.e. the velocity estimate will be constrained to fall along, or close to, the dashed line depicted in Figure 23b.

The explanation of the perceived direction bias in Weiss *et al* (2002) and Weiss and Adelson (1998) is thus based on the Adelson and Movshon (1982) model of plaid perception, in which only the 1-D motion of the component gratings are detected in the

first stage of analysis of the plaid motion, and their model is presented as a Bayesian extension of this model. This is clearly reflected in their explanation since they indicate that their model forms local likelihoods as "fuzzy" constraint lines defined by the 1-D motion of the component gratings. However their model, as is the case with the model presented here, undoubtedly detects both the 1-D and 2-D motion that is present in the stimulus in the form of the motion of the edges and end-points of the blobs, as is clearly demonstrated by their depiction (in Figure 3 of Weiss *et al* (2002)) of the likelihood functions generated by their model for a moving diamond stimulus. It is surprising therefore that no reference is made to the likelihood functions formed from the 2-D motion in the plaid stimulus, and their role in forming the estimate.

An alternative explanation for the perceived plaid motion is offered here, which is based on the 1-D and 2-D motion of the edges and end-points of the blobs. This is illustrated in Figure 23c. Here the velocity vectors are shown corresponding to the orthogonal motion of the longer and shorter edges of the blobs in the plaid,

$$v_{\phi} \triangleq \begin{bmatrix} r_{\phi}\cos\phi \\ r_{\phi}\sin\phi \end{bmatrix}, v_{\varphi} \triangleq \begin{bmatrix} r_{\varphi}\cos\varphi \\ r_{\varphi}\sin\varphi \end{bmatrix}$$

respectively, and their constraint lines, together with the IOC and vector sum velocities, v_{IOC} and v_{VS} respectively, of the component gratings. Also shown are the velocity estimates from the model for the first step ($v_{e1} = 1.83$; $\theta_{e1} = 40^{\circ}$), the fifth step ($v_{e5} = 2.38$; $\theta_{e1} = 25^{\circ}$), and in the steady-state ($v_{es} = 2.62$; $\theta_{es} = 18^{\circ}$).

It is clear that the initial estimate v_{e1} lies very close to the velocity vector v_{ϕ}, corresponding to the orthogonal motion of the long edge of the blob, and to the maximum of the likelihood function (the "fuzzy" constraint line) for v_{ϕ}. Subsequent velocity estimates in further iterations of the recursive algorithm get closer to this maximum, and also to the velocity vector v_{IOC}. Note that the effective prior for each step in the estimation algorithm is given by the velocity estimate in the previous step, which

together with the influence of the likelihood function corresponding to the 2-D velocity of the end-points of the blobs, v_{IOC}, leads to the convergence of the estimate towards the IOC velocity.

9.3. Further support for the motion integration model.

To reinforce the account of the model behaviour given above, a further piece of evidence is now provided that the first stage of plaid motion perception is based on the 1-D and 2-D motion of the blobs rather than the 1-D motion of the component gratings.

Stone *et al.* (1990) investigated the effect on the perceived plaid direction of making the contrasts of the component gratings unequal. They based their investigation on the Adelson and Movshon (1982) model, assuming their first stage in which the 1-D velocities of the component gratings were detected to be correct. They hypothesised that the low contrast grating would be detected at a lower speed than the true value and that if this erroneous value were used in a second stage IOC calculation of plaid direction, a significant contrast-dependent error in the perceived plaid direction would result. They used a Type I plaid with angular separation of the component gratings of 120°, and changes in the ratio of the speeds of the component gratings to vary the true direction of the plaid whilst maintaining a constant plaid speed.

In this way they found that the perceived plaid direction was biased towards the direction of the higher contrast grating and this bias increased for increasing contrast ratio, and also for decreasing total contrast (the sum of the grating contrasts). At 5% total contrast, the average observed bias varied between 0°, at a contrast ratio of 1, to ~16°, at a contrast ratio of 4:1. A maximum bias of 20° was observed for a total contrast of 10% and a contrast ratio of 8:1. The modified Adelson and Movshon (1982) model proposed by Stone *et al.* (1990) using perceived rather than actual component speeds

appeared to give qualitatively similar results to those observed (see their Figure 11, which is presented here as Figure 24).

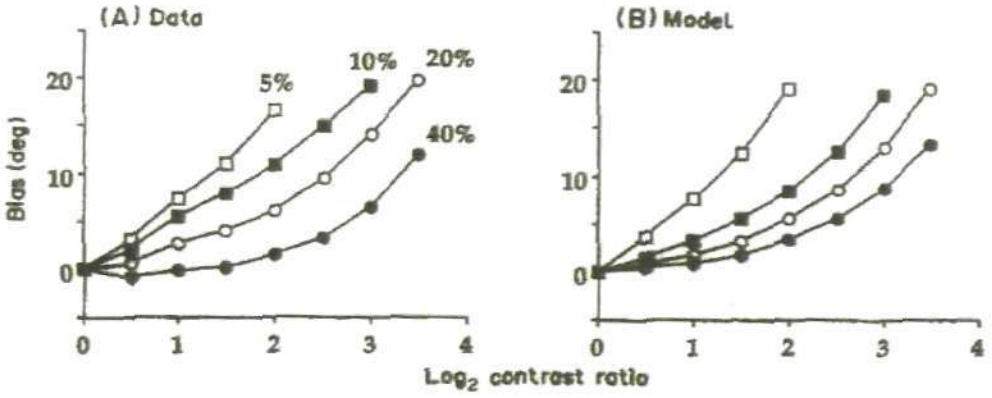
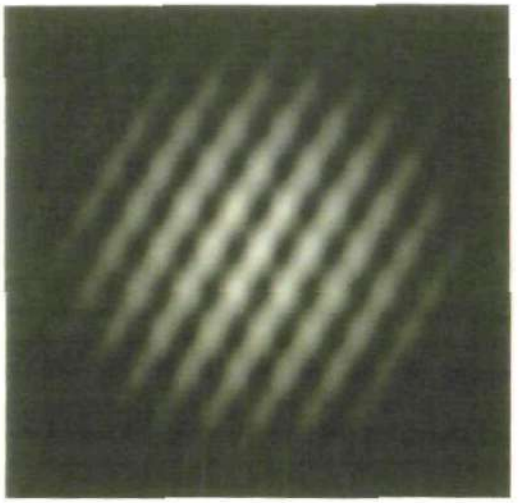


Figure 24 (**Fig. 11** from **Stone et al., 1990**). Simulated vs actual bias. (A) This panel is a plot of the same data as in Fig. 7 from Stone et al., 1990 (for four different contrasts: 5, 10, 20, 40%) averaged over four subjects and over symmetric contrast-ratio pairs. (B) This panel shows simulations of the model in Fig. 9 Stone et al., 1990 under the same conditions as (A).

However, similar experiments by Champion, Hammett and Thompson (2007) appeared to invalidate the modified IOC model of Stone *et al.* (1990), since it would also predict a bias towards the direction of the low contrast component at high component grating speeds due to an increase in the perceived speed of low-contrast gratings for grating speeds above ~12 deg/s (Champion *et al.*, 2007). Champion *et al.*. observed an increasing bias with component speed which was always towards the direction of the high contrast component except for the very lowest component grating speeds, but a *decrease* in the bias at the highest component speeds (above 12 deg/s), consistent with their observed switch in the contrast-related misperception of grating speed for higher speed gratings. It should be noted however that Champion *et al.* used plaids of total contrast equal to 90%, compared to the total contrast values of between 5% and 40%. They also used component gratings with angular separation of 90°, compared with the 120° angular separation used by Stone *et al.*

a



b

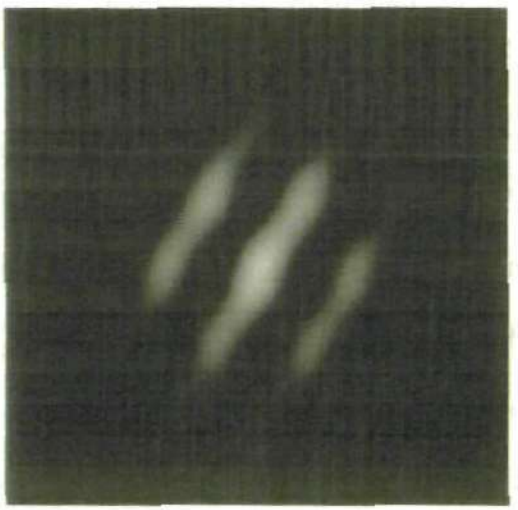


Figure 25. The two plaids used the model simulations of the experiments of Stone *et al* (1990), showing the effect of the spatial frequency of the component gratings on blob size and number. Both plaids correspond to an angular separation of component gratings of 120°, 60° either side of the vertical (0°) and for each the contrasts of the component gratings are in the ratio of 4:1. For the plaid in a. the spatial frequency of the component gratings is twice that for the plaid in b.

Champion *et al.* also suggest that their results are inconsistent with the Bayesian IOC model of Weiss *et al.* (2002), since that model relies upon the perceived speed of the gratings being smaller for lower contrast, and hence higher uncertainty, owing to the

greater influence of the "slow" prior. Champion *et al.* also claim that their results are inconsistent with several other models of plaid perception including the 1-D and 2-D parallel pathways model of Wilson *et al.* (1992), and the blob tracking model of Alais, Wenderoth and Burke (1994).

Applying the model proposed here to this data (Figure 25) shows that it replicates the misperception of the direction of plaid motion towards the direction of the higher contrast grating, but that the magnitude of the bias in the estimated direction is dependent on the spatial frequency of the component gratings. The case of a plaid with a separation of component gratings of 120° , 60° either side of the vertical (0°) and a contrast ratio of 4:1 is shown in Figure 25a. It is clear that the salient feature of this plaid is a set of "blob-lines" which are formed from a joining-up of the plaid blobs. The direction of motion of the blobs is the VS of the component gratings, i.e. the plaid direction of 0°, but the orthogonal direction of motion of the blob-lines is 300°, the direction of the higher contrast component grating. The estimated plaid direction computed by the model is 308°, giving a bias of 52°away from the IOC direction of 0° towards the higher contrast grating direction, much greater than that measured by Stone *et al.* (1990), where the direction error was up to 20° for this contrast ratio (4:1). However, this result was obtained for a grating spatial frequency and a viewing aperture shown for the plaid illustrated in Figure 25a, corresponding to ~14 cycles of the component gratings being present within the viewing aperture.

If the gratings' spatial frequency and the viewing aperture are changed to approximate that used by Stone *et al.* (1990) and Champion *et al.* (2007), approximately 6 cycles of the component gratings are present in the viewing aperture, as illustrated by the plaid in Figure 25b. The blob lines are still clearly visible but the size of the blobs is greater by about a factor of two. In this case, the estimated plaid direction computed by the model is 342°, giving a bias towards the direction of the higher contrast grating of

18°, comparable to that measured by Stone *et al.* (1990) for this contrast ratio (4:1). The bias computed by the model for the contrast ratio of 2:1 was 7°, which is consistent with the Stone *et al.* result of approximately 7° for the 5% contrast case, and with the results of Champion *et al.* (2007) who used a contrast ratio of 2:1 and obtained a maximum bias of approximately 7°.

The above example illustrates the importance, both in psychophysical experiments and in modelling, of the choice of the spatial frequency of the component gratings in relation to the viewing angle/aperture of the stimulus. The model results would suggest that if the psychophysical experiments of Stone *et al.* (1990) or Champion *et al.* (2007) had been carried out using a higher component grating spatial frequency, a far greater bias towards the higher contrast grating would have been obtained, owing to the greater salience of the 1-D motion of the blob-lines in the direction of the higher contrast grating, compared to that of the 2-D motion of the blobs themselves, when viewing the plaid.

Alais *et al.* (1997) investigated the effect of blob size and number on perceived plaid direction, in this case for Type II plaids. They showed, by varying both spatial frequency and viewing aperture size, that there is a large effect of blob size on the perceived direction bias, of up to 14.1°, due to changes in the component spatial frequency, but a small effect of blob number, of about 5°, obtained by changing aperture size whilst spatial frequency is held constant. The motion integration model was used to simulate their experiments, keeping the viewing aperture constant and varying the spatial frequency of the component gratings. Three values of spatial frequency were used: 0.6, 0.3 and 0.2 cycles/pixel. For the sake of comparing the simulation results with the experimental results, it was assumed that these spatial frequencies correspond to the experimental values of 3.0, 1.5, and 1.0 cycles/°.

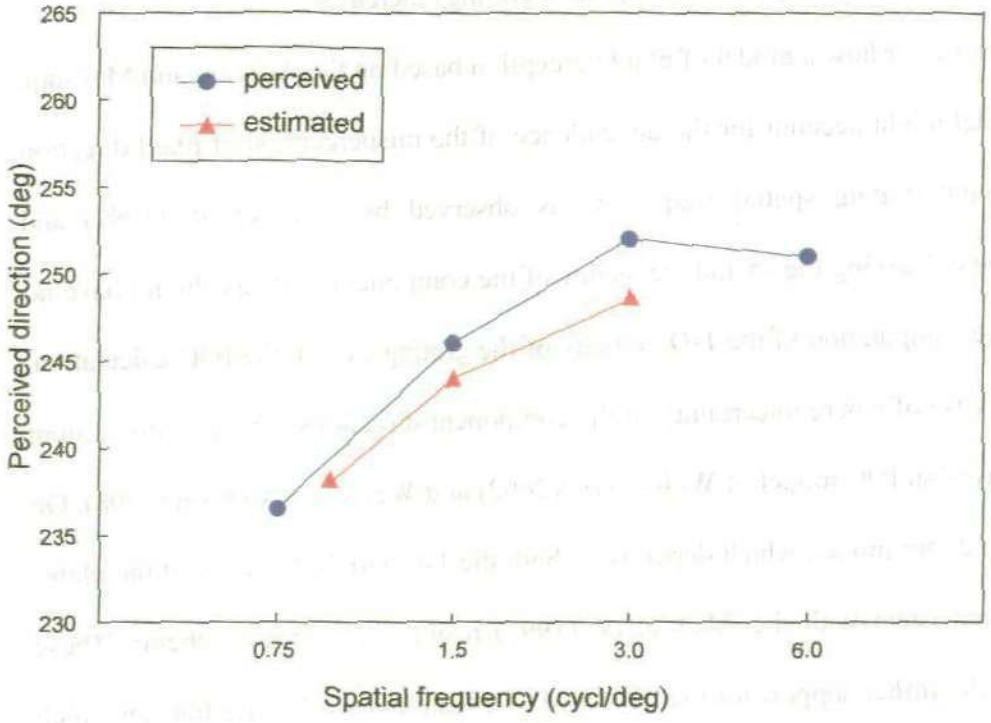


Figure 26. Results from simulations of the computational model for the plaids used in the experiments of Alais *et al* (1997) showing the perceived plaid direction as a function of spatial frequency obtained in the experimental study, for the 3° aperture case (▲symbols) and the plaid direction estimated by the model (• symbols). The estimates of the plaid direction are very similar to the perceived values (within 4°) and show the same trend, with a decrease in the misperceived direction bias as the spatial frequency of the component gratings increases. From the original diagram for the experimental results in this study, perceived errors were in the region of ±2°.

Figure 26 shows the results from Alais *et al.* (1997) (their Figure 5) giving the perceived direction as a function of spatial frequency for the 3° aperture case (• symbols), together with the steady state direction estimates from the model (▲symbols), for each of the component grating spatial frequencies. The 6.0 cycles/° case was not simulated owing to the limitations of the model in dealing with such high frequencies due to the particular choice of window size. As can be seen from Figure 26 , the estimates of the plaid direction are very similar to the perceived experimental values

and, importantly, show the same trend, with a decrease in the misperceived direction bias as the spatial frequency of the component gratings increases.

It is not clear how a model of plaid perception based on the Adelson and Movshon (1982) model might account for the dependence of the misperception of plaid direction on component grating spatial frequency, as observed by Alias *et al.* (1997) and modelled here. Varying the spatial frequency of the component gratings should have no effect on the computation of the 1-D velocity of the gratings, or on the IOC calculation, even in the case of where uncertainty in the component directions is taken into account as in the Bayesian IOC model of Weiss *et al.* (2002) and Weiss and Adelson (1998). On the other hand, our model, which depends on both the 1-D and 2-D motion of the blobs, is entirely consistent with the Alais *et al.* (1997) results. As noted by them: "These results provide further support for the existence of a feature-sensitive mechanism which responds to the motion of plaid features and which is tuned to their various qualities". The model presented in this thesis provides just such a mechanism.

9.4. Other approaches based on a feature tracking mechanism

Other approaches based on a feature tracking mechanism have been proposed which are related to the mechanisms that have been described here. In particular, Bowns (1996) proposed a feature tracking explanation for the misperception of Type II plaids as observed by Yo and Wilson (1992) which is based on specific plaid features, "avgL", "minL" and "maxL" which she introduces, and which clearly relate to the blob features that were defined in Section 7.1.

In Figure 6 of Bowns (1996), here presented as Figure 27, these features and their motion are illustrated for a plaid in which the directions of motion of the two component gratings differ by 10° (directions of 90° and 100°). According to the analysis

presented in Section 7.1 of this thesis, the blobs in this plaid, which appear to correspond approximately in shape to the maxL feature, have an edge ratio of 1:0.09, i.e. the blobs are highly elongated, and the longer edges move in an orthogonal direction of 95°, almost exactly equal to the vector sum direction of 93°. For a component grating speed ratio of 1:0.5, the model simulation gives for this plaid an initial direction estimate of 90°and a final direction estimate of 65°, i.e. 25° to the right hand (IOC = 19°) side of the vertical, implying that in a forced choice decision of left or right of the vertical, as in the Bowns (1996) experiments, a consistent IOC choice would be likely. At this point the reader is referred back to the description and simulations of Bowns' 1996 experiments in Section 8.2. A different explanation is however given in Bowns (1996) for consistent IOC result; namely that, as stated in the legend to her Figure 6 "there are no edges that move in the vector sum direction for this plaid". Hence it is concluded that the choice will always be in the IOC direction.

In Figure 7 of Bowns (1996), presented here as Figure 28, avgL, maxL and minL are again illustrated for a plaid in which the directions of motion of the two component gratings differ by 80° (directions of motion of 90° and 170°). In the legend to the Figure it is stated again that "there are no edges that move in the vector sum direction for this plaid". However, for this plaid the analysis given in Section 7.1 shows that the blobs are not elongated, having an edge ratio of 1:0.84, which would predict a velocity estimate close to the IOC direction. Also for this a plaid, the IOC direction (108°) is close to the vector sum direction (114°), and both are thus to the left of the vertical. The model simulation gives initial and final velocity estimates for this plaid which are both approximately equal to the IOC direction, thus predicting, in a forced choice of left or right of the vertical, a decision of left (vector sum), corresponding to the outcome in the actual experiment, as indicated in the legend to Figure 7 of Bowns (1996).

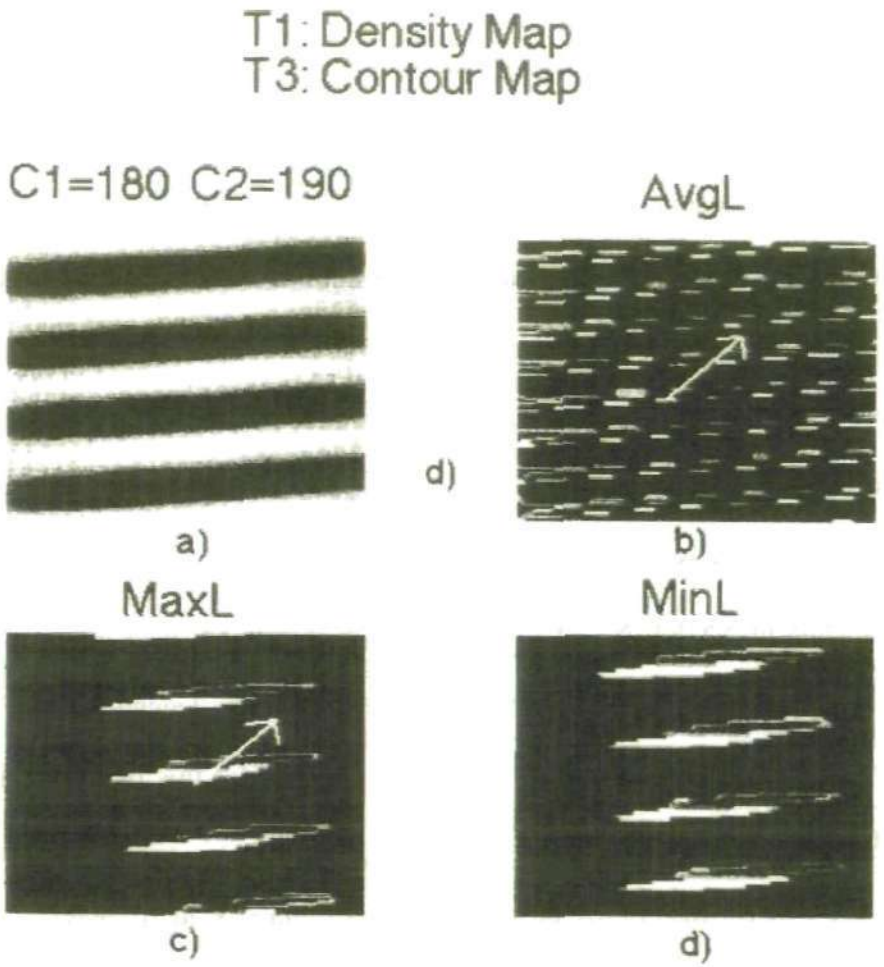


Figure 27 (**Figure 6** from **Bowns (1996)** together with the original legend). (a) A density plot of the plaid with the first component oriented at 180 deg and the second component oriented at 190 deg. The actual stimulus was circular for all stimuli as described in the Methods Section 2. (b) avgL changed over time by showing t1 as a density plot and t3 as a contour map. (c) and (d) show similar plots for maxL and minL, respectively. Possible motion directions are indicated by large white arrows. There are no edges that move in the vector sum direction for this plaid. This is consistent with the results; subjects perceived this plaid moving in the IOC direction for 100% of the presentations. The displacement direction of all regions shown in Figs 6-8 corresponds to the IOC direction.

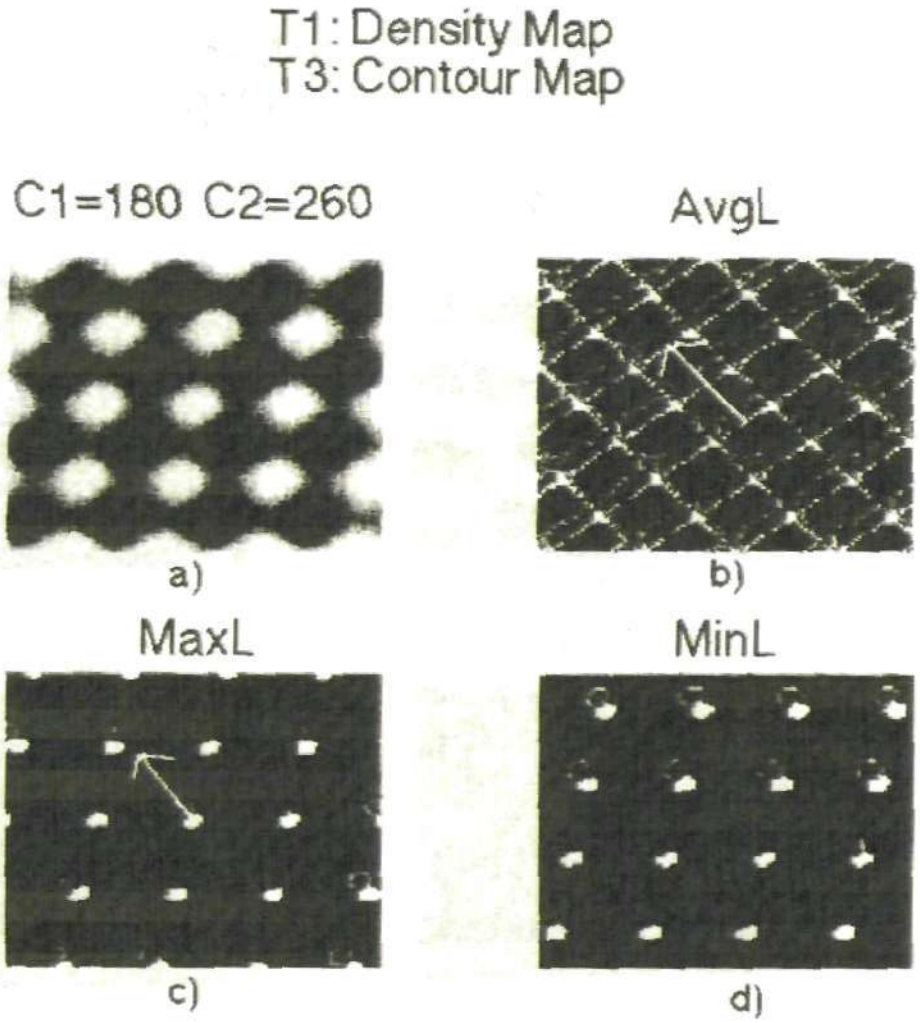


Figure 28 (**Figure 7** from **Bowns (1996)** with the original legend). (a) A density plot of the plaid with the first component oriented at 180 deg and the second component oriented at 260 deg. The plots are similar to those shown in Figure 6. Again there are no edges that move in the vector sum direction for this plaid, and the results are 100% consistent with the IOC predictions.

Finally, in Figure 8 of Bowns (1996), which is shown here as Figure 29, avgL, maxL and minL are illustrated for a plaid in which the directions of motion of the two component gratings differ by 40° (directions of motion of 90° and 130°).

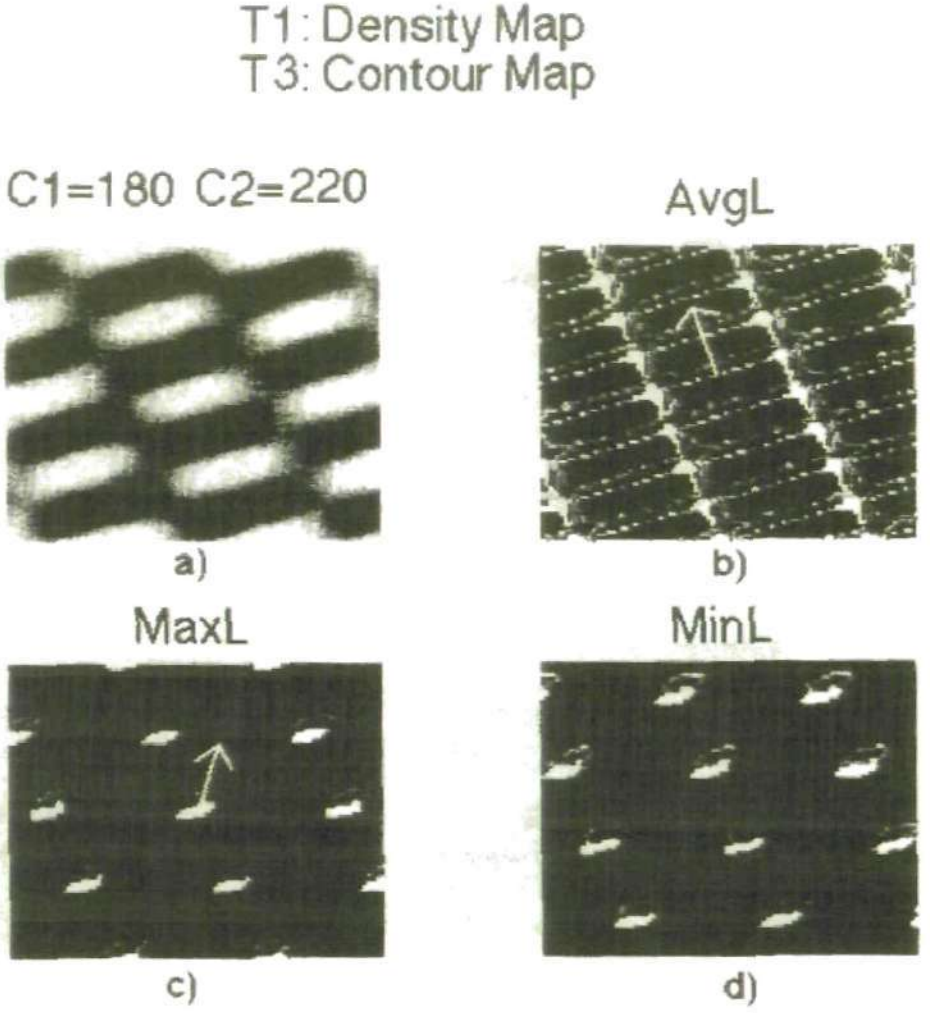


Figure 29 (**Figure 8** from **Bowns (1996)** with the original legend). (a) A density plot of the plaid with the first component oriented at 180 deg and the second component oriented at 220 deg. This is a plaid from the middle of the set of plaids used. Although the regions are all shown to be displaced in the IOC direction, (b) shows a significant edge shift in the vector sum direction (see white arrows). This is consistent with the results, subjects performed variably to this plaid with one subject perceiving it in the IOC direction and the other in the vector sum direction.

In this case, the legend to Figure 29, originally copied from Bowns (1996) indicates that whilst neither of the features maxL or minL have edges moving in the vector sum direction, avgL has an edge which moves in this direction. The inference is made that the presence of this motion resulted in subjects performing variably with this

plaid, one perceiving it in the IOC direction (right of the vertical) and one in the vector sum direction (left of the vertical). The analysis of this plaid, with a speed ratio of 1:0.5, gives a blob edge ratio of 1:0.36, i.e. the blobs are somewhat elongated, and their long edges move in an orthogonal direction of 110°, close to the vector sum direction of 103°. The model estimates a plaid velocity direction in the first step of the estimation algorithm of 95° (5° to the left of the vertical) and a final estimate of 73° (17° to the right of the vertical). The IOC direction is 67°.

The results from the model simulations of the full range of plaids used in Experiment 2 of Bowns (1996), of which those discussed above are a subset, are shown in Figure 30. The plots in Figure 30 show the initial and final estimated plaid directions as a function of the angular difference between the component grating directions for these plaids. The shaded area in the centre of the graph indicates the range of component grating angular differences which resulted in an inconsistent choice by subjects between "vector sum direction" and "IOC direction" for the corresponding plaids. These results, and the discussion above, suggest that the reason for the observed variability between subjects in their choice of IOC or vector sum direction (Bowns, 1996) lies in the variability of subjects in terms of the dependence of their direction perception on the duration of the stimulus. As Yo and Wilson (1992) showed, subjects can display considerable differences in this dependence. In Figure 6 of Yo and Wilson, one subject (HRW) reported a direction bias of 30° after ~90 msec. stimulus duration, from an initial bias of 60° at ~60 msec. Another subject (HJ) reported a direction bias of 15° after ~90 msec., from approximately the same initial bias at 60 msec. Significantly, the stimulus duration used in the Bowns (1996) experiments was 80 msec, which would imply that a significant variation in perceived bias between subjects at this duration was possible. A similar variability to that reported by Yo and Wilson (1992) would therefore probably

be sufficient to cause the difference in direction choice between the two subjects in the Bowns (1996) experiments.

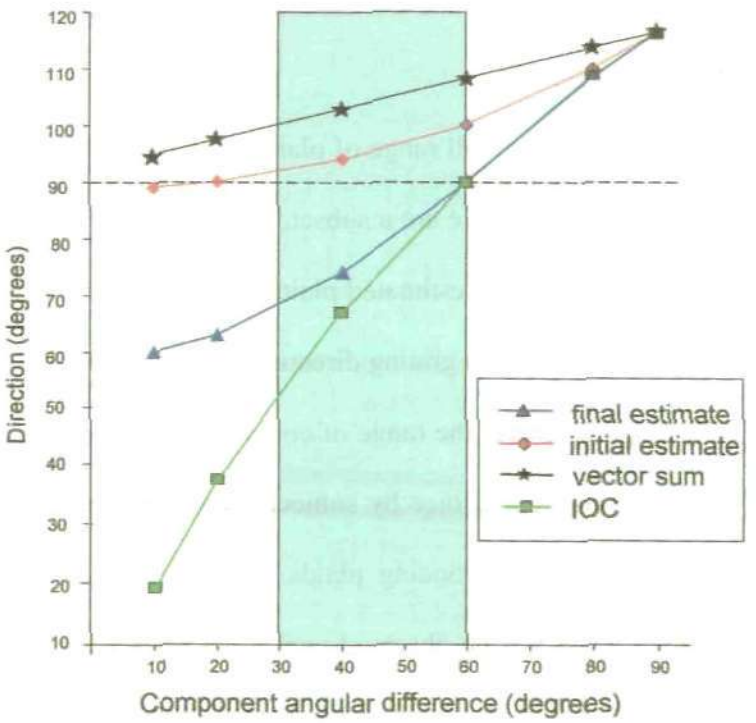


Figure 30. Results from simulations of the computational model for the plaids used in Experiment 2 of Bowns (1996), showing the initial and final estimated plaid directions as a function of the angular difference between the component grating directions for these plaids. The shaded area in the centre of the graph indicates those component grating angular differences which resulted in an inconsistent choice by subjects between "vector sum direction" and "IOC direction" for the corresponding plaids (see the text in the Discussion section for a further discussion of these results).

Whilst the explanation given here contrasts with that of Bowns (1996), her explanation does clearly indicate that there is present in the plaid pattern both motion in the VS direction (in terms of the analysis given here, the orthogonal 1-D motion of the longer edges of the blobs) and in the IOC direction (in terms of the analysis given here, the 2-D motion of the blob end-points). She uses this fact to propose that the variation between subjects may result from a competition between these two sets of motion

information. The analysis given here suggests however that a Bayesian process which using both information sets can also predict this result.

Another analysis of Type II plaid misperception based of the motion of features in the plaid was presented in Bowns (2006). Here a squaring operation is performed on the plaid and two "components" are identified: "sqHF" and "sqLF" which are derived from the squared plaid. The description in Bowns (2006) shows that the "components" are in fact two gratings formed from the squared plaid pattern, a high spatial frequency grating and a low spatial frequency grating, with spatial frequencies and orientations defined in the Appendix. They clearly relate to the product gratings, and have the same orientations and direction of motion as these, as illustrated in Figure 1(d) of Bowns (2006). Examples of the values for the direction of motion of the sqHF and sqLF are also given for three Type II plaids, which were also used in Bowns (1996), showing that the direction of motion of the sqHF "component" is close to the VS direction. This led to the proposal that the direction of motion of the "components" provided a better overall predictor of the misperceived direction of these plaids than either the VS direction, as suggested by Yo and Wilson (1992), or the IOC direction, as suggested by Adelson and Movshon (1982). The analysis and simulation results presented here clearly concur with this conclusion, as the predictions based on the motion of the blob edges show. However, Bowns also suggests that there is no motion energy in the plaids in the IOC direction, so that a full explanation of the misperception would "a model that incorporates both squaring and the IOC". The model presented in this thesis however incorporates both the 1-D motion of the blob edges, which contain motion energy close to the VS direction, and the 2-D motion of the blob end-points, which contain motion energy in the IOC direction. Used together in a recursive Bayes estimation framework, it has been shown that the model closely predicts a wide range of results on perceived direction of plaid motion.

9.5. New research directions suggested by the model.

New directions for experimental investigation are suggested by the motion integration model, including monitoring the response to Type II plaid motion of end-stopped cells (Hubel and Weisel, 1965; Pack *et al.*, 2003) in layer 4b of area V1, the layer which contains the majority of V1 neurons projecting to MT. The model predicts that such experiments for Type II plaids will show that these neurons signal the 2-D motion of the high luminance regions, i.e. the blobs, in the plaid Additionally, studies of the dynamic response of MT neurons to Type II plaids have, as far as the author is aware, not been done, although a stimulus consisting of a field of short bright bars (Lorenceau, Shiffrar, Wells, and Castet, 1993) mimics the high luminance regions in Type II plaids. For this bar-field stimulus, Pack and Born (2001) showed that MT neurons initially respond primarily to the component of motion perpendicular to a contour's orientation, but over a period of approximately 60 ms the responses gradually shift to encode the true stimulus direction, regardless of orientation. Thus the responses of the MT cells closely parallel the psychophysical responses of human observers to the motion of Type II plaids (Yo and Wilson, 1992). Similar studies in which the responses of MT neurons are selectively inhibited, by lesioning or reversibly cooling (Hupé, James, Payne, Lomber, Girard and Bullier, 1998; Supèr and Lamme, 2007) might also be able to test the hypothesis that the local 1-D and 2-D local motion signals are combined to provide the perception of plaid motion via a recursive estimation process, which, it is hypothesised here, is implemented in the recurrent interaction between V1 and MT, an interaction which has been strongly implicated in the perceptual awareness of visual motion (Lamme and Roelfsema, 2000; Sterzer, Haynes and Rees, 2006).

9.6. Summary.

In summary, the discussion in this chapter on the application of the motion integration model of plaid perception covered the comparison with the Adelson and Movshon (1982) model. An alternative explanation of their psychophysical results has been presented as well.

Furthermore, a comparison with the Weiss *et al.* (2002) model was given, where again the concept of a first-stage components' motion processing was analysed in detail together with specific examples. Additional support for the motion integration model has been provided in the last sections of the chapter. The case of a plaid with components of different contrast has been examined. Furthermore, the analysis of the effect of the number and size of the blobs on the plaid perception was presented, together with supporting psychophysical data and model simulations. Finally, new research directions were proposed.

Chapter 10. General conclusions and future research

10.1 Summary of the main conclusions of the thesis.

In this work a Bayesian model of motion integration using a Kalman filter algorithm was presented. The model aimed to describe the dynamics of motion integration which has been observed in the numerous psychophysical experiments regarding both eye smooth pursuit and plaid motion perception. The dynamics of the perception of the object motion starts with a biased or non-veridical value of velocity and it takes some time for the veridical value of velocity to be perceived. The dynamics of the estimation thus proves to be a good source of information about the underlying mechanisms of combining the 1-D and 2-D local motion signals in order to estimate the veridical motion of the object.

The behavior of the model on the smooth pursuit data was very close to the psychophysical experimental results (Lorenceau *et al.*, 1993; Masson and Castet, 2002; Masson *et al.* 2000; Pack and Born, 2001; Lindner and Ilg, 2000; Masson and Stone, 2002, Wallace *et al.*, 2005; Pack *et al.*, 2004).By simulating the smooth pursuit data it was possible to show the effects of different stimulus contrasts and sizes on both the initial biased value of velocity and the temporal dynamics. The model captured well the dynamics of the smooth pursuit eye movements, which reflected the supposed underlying motion integration mechanisms. As discussed, there are algorithmically similar models (Weiss *et al.*, 2002), but they have not applied specifically to the dynamics of the smooth pursuit data.

Based on the idea of similarity of the motion integration mechanisms underlying both the eye smooth pursuit dynamics and the plaid motion perception dynamics, it was possible to propose a new explanation of the motion integration processing in plaid perception. The application of the model to the controversial data of Type II plaid perception showed that the perceptual phenomena could be fully explained

by taking into account the geometrical properties of the high intensity regions of the plaids, the so-called blobs. It was shown that contrary to the well-known two-stage model of Adelson and Movshon (1982), involving the detection of the motion of the component gratings followed by application of the IOC rule, the proposed model was able to estimate the veridical as well the biased plaid velocity without explicitly taking computing the velocities of the component gratings in the first stage. Instead, the model uses the 1-D and 2-D motions of the high intensity features of the plaids, the blobs.

In this sense the presented model and its explanation of plaid motion misperception, although mathematically similar to the model of Weiss *et al.* (2002), can be distinguished from it conceptually. The explanation for the non-veridical bias toward the VA direction, according to the present model, was based on finding the relationship of the blob long edge orientation and the VA direction of motion of the component gratings. It was shown by a detailed geometrical analysis of the blobs' size and orientation that for the cases with the strongest perceived non-veridical direction bias for Type II plaids, the shape of the blobs were substantially elongated and the orthogonal direction of motion of the longer edge of the blob was close to the VA direction. This is the case when the directions of motion of the component gratings are oriented relative to each other with a small angle (up to about 20°).

The main conclusion of this analysis was that the motion of the edges of the blobs played a more important perceptual role in plaid motion detection, as 1-D motion signals, than the motion of the component gratings. It was shown that the relative length of the blob edges is important not only in relation to the 1-D motion cues from the blobs themselves, but also in relation to the visual angle of viewing the stimulus. This led to a further conclusion that depending on the visual angle and spatial frequency of the gratings forming the plaid, the most dominant perceptual feature in some cases would be not the separate blobs, but the lines of blobs which they form, which is in complete

agreement with the concept used in the model of analyzing the high intensity features of an image.

The model also gave a new explanation of the psychophysical results for some specific experimental conditions, which seemed very inconsistent within the participants, who were reporting very different directions of motion of specific Type II plaids (Bowns 1996). One of the main contributions of the thesis is therefore the conclusion that plaid motion perception can be treated generally as dependant on the visible features of the plaids, i.e. blobs and lines of blobs, and not on the component gratings themselves. This departs substantially from the apparently well-accepted idea that the motion of the component gratings, computed in the first stage of an integration mechanism is essential to the motion perception process (Adelson and Movshon, 1982), as distinct from the idea presented in the thesis that the dominant visible features of the plaid play the most significant role, for all plaid types.

Two versions of the Bayesian model of the dynamics of motion integration were presented in the thesis: a standard form of the Kalman filter algorithm (Section 2.1) and an approximate version (Section 2.2). The latter version of the model was created in order to make the algorithm computationally closer to a neural network implementation. The main simplification of the algorithm aimed to avoid matrix inversions when calculating the Kalman gain, and for that purpose the covariance matrix of the velocity estimate was approximated by a diagonal matrix. Another simplification involved the assumption that the spatial intensity derivatives in the x and y directions were independent. This led to the approximation of the matrix computing the covariance of the local measurements of the intensity spatial derivatives by a diagonal form. Together these approximations led to a simpler calculation of the Kalman gain by avoiding the matrix inversion.

A further modification to the algorithm, by which the covariance matrix of the velocity estimate P_k^k was not updated in each step of the algorithm, but was maintained at its initial value P_0^0. This version of the algorithm allows it to be represented in a distributed, recurrent neural network form.

Both the standard and approximate versions of the algorithm gave very similar simulation results, which were presented in the simulation experiments for the data on smooth pursuit eye movements. The same similarity held true for the simulation results for plaid motion perception, however in the thesis only the results for the standard form of the algorithm were presented.

10.2. Critical reflection on the presented work.

The very simple form of the local motion detectors, i.e. non-overlapping square windows, used in the simulations, and also of the method for calculating the spatial and temporal derivatives of the image intensity, although successful for the limited type of stimuli considered in this study, would be candidates for improvement in order to develop the model further. The simple form of the motion detector windows and the simplicity of the spatial and temporal intensity derivative calculations led, in the case of the smooth pursuit eye movement simulations, to a variable estimation of the velocity, in particular in the approximate version of the algorithm, which depended on the spatial shift (speed of stimulus), the size of the window and size of the stimulus. This appeared as a periodic, approximately constant amplitude variation of the velocity estimate (Figure 13). This problem was solved simply by averaging the velocity estimate for several steps of the algorithm, the number of the steps used for the averaging depending on the ratio of window size and the speed of stimulus. A more sophisticated solution would have been to use overlapping. circular windows, with spatial Gaussian filters for the local motion detector windows, and Gaussian derivative filters (

or other methods (e.g. Johnston et al, 2003) for the calculation of the spatial derivatives. However, this issue did not seem to cause a major problem as the general results of the algorithm proved to be robust and consistent both with regard to the psychophysical data and also between the simulations themselves in eye smooth pursuit and plaid motion perception.

The approximate, more neurally plausible version of the algorithm gave results which were very similar to the standard version, and in some cases appeared to give results which were closer to the psychophysical data. However, very little work was done specifically on how the neural network circuits suggested by the approximate version of the model (Figure 9) would carry out the calculation and integration of the local motion signals, which was treated as a subject of future investigation. The modifications made to the standard version of the algorithm which led to the approximate version are also somewhat crude, with assumptions which need further investigation and depth of understanding. A more sophisticated approach to a neural network implementation of the Kalman filter algorithm might use the recent work of Linsker (2008).

Another desirable aspect of the present work would have been the possibility of conducting some psychophysical experiments, even if relatively simple, in order to confirm or not some of the predictions made from the model, in particular the predictions about the asymmetric Type I plaids, described in Section 8.5. Unfortunately this was beyond the scope of the facilities and of the time available to the project.

10.3. Future development of the model and predictions.

New directions for further experimental investigation have been suggested in previous Chapters of the thesis. The validity of the motion integration model in terms of its use of both 1-D and 2-D motion signals derived from the plaid blob features could be tested

physiologically by monitoring the response of end-stopped cells (Hubel and Weisel, 1965; Pack *et al.*, 2003) in monkey subjects to the coherent motion of Type II plaids, in particular in layer 4b of area V1, which is known to contain the majority of V1 neurons projecting to MT. The prediction would be that such experiments for Type II plaids will show that the end-stopped cells would signal the 2-D motion of the high luminance regions, i.e. the blobs, in the plaid. Additionally, it would be interesting to investigate the dynamic response of MT neurons to Type II plaids, which have, as far as the author is aware, not been done. The study closest to this was performed with a stimulus consisting of a field of short bright bars (Lorenceau, Shiffrar, Wells, and Castet, 1993), which could mimic the response of the high luminance regions in Type II plaids. Also Pack and Born (2001) studied the response of MT neurons to such a bar-field stimulus and demonstrated a dynamic response, initially responding to the component of motion perpendicular to a contour's orientation, but over a period of approximately 60 ms the responses gradually shifting to encode the true stimulus direction, thus closely paralleling the psychophysical responses of human observers to the motion of Type II plaids (Yo and Wilson, 1992).

The hypothesis put forward here that the local 1-D and 2-D local motion signals are combined via a recursive estimation process involving the recurrent interaction between V1 and MT might also be tested by studies of the responses of monkey subjects in either smooth pursuit or plaid motion perception experiments, in which the responses of MT neurons are selectively inhibited, by lesioning or reversibly cooling (Hupé, James, Payne, Lomber, Girard and Bullier, 1998; Supèr and Lamme, 2007)

Further predictions from the model relate to Type I asymmetric plaids. It was possible to explain why there would be no significant bias for the majority of Type I asymmetric plaids, and at the same time to predict such a bias for some specific cases of Type I plaids which have low contrast, a specific range of angles between component

gratings and specific ratio of component grating speeds, presented for short durations as in the case of Type II plaids. Psychophysical experiments with such Type I plaids have not, to the author's knowledge, been carried out to date.

Further predictions were made regarding the choice of the spatial frequency of the component gratings in relation to the viewing angle/aperture of the stimulus, used in both psychophysical and physiological experiments. The model results would suggest that if the psychophysical experiments of Stone *et al.* (1990) or Champion *et al.* (2007) had been carried out using a higher component grating spatial frequency, a far greater bias towards the higher contrast grating would have been obtained, owing to the greater salience of the 1-D motion of the blob-lines in the direction of the higher contrast grating, compared to that of the 2-D motion of the blobs themselves, when viewing the plaid. This would be interesting and a valuable test of the model and its conceptual use of plaid blob features in plaid motion perception.

It was also suggested that physiological experiments with varying sizes of blobs and orientations of Type II plaids, would show specific dynamics in the response of neurons in V1 and MT areas. More specifically, the stimuli for experiments which measure neural responses in V1 should have relatively bigger blob size in order to avoid the effect of the merging of the blobs into lines, leading to an apparent response to the component gratings' velocities rather than to the velocities of the blob edges.

The model could be extended in at least two possible directions. One is to make it closer to the known neural properties and circuits, e.g. more close to the biological reality. In particular the distributed version of the model presented in Section 2.3 could be examined in relation to the known recurrent circuitry between motion sensitive neurons in V1 and MT.

Another approach for future development of the model could be to keep it on a similar abstract level, but develop the algorithm and apply it to the related but more

difficult problem of image segmentation. A possible approach might be to create a model which is capable of simultaneously building multiple Kalman filter based motion integration sub-models, by selectively combining local motion signals according to the similarity of their measurements with a global velocity vector. Where the velocity vector of a local motion signal differs substantially in speed and direction with the globally estimated velocity vector, this local motion signal can be assigned to a sub-model corresponding to a separate segregated object. Spatial information on the form of potential objects would be necessary in order to make such an assignment in an effective way. Substantial theoretical and simulation work would need to be carried out in order to investigate and develop such a model.

References

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.* **A 2**, 284-299.

Alais, D.M, Wenderoth, P.M, & Burke, D.C. (1994). The contribution of 1-D motion mechanisms to the perceived direction of drifting plaids and their aftereffects. *Vision Res.* **34**, 1823-1834.

Alais, D.M, Wenderoth, P.M, & Burke, D.C. (1997). The size and number of plaid blobs mediate the misperception of Type-II plaid direction. *Vision Res.* **37**, 143-150.

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiology*, **52**, 1106-1130.

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

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

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *J. Physiol.* **119**, 69-88.

Barlow, H. B. & Hill, R. M. (1963). Selective sensitivity to direction of motion in ganglion cells in the rabbits retina. *Science* **139**, 412-414.

Barlow, H. B. & Levick, W. R. (1965). The mechanism of directionally selective units in the rabbit's retina. *J. Physiol.* **178**, 477-504.

Barthelemy, F. V., Perrinet, L.U., Castet, E. & Masson, G. S. (2008) Dynamics of distributed 1D and 2D motion representations for short latency ocular following. *Vision Res* **48**: 501-522.

Bayerl, P. & Neumann, H. (2004). Disambiguating visual motion through contextual feedback modulation. *Neural Comp* **16**: 2041–2066.

Benton, C. & Johnston, A. (1997). First-order motion from contrast modulated noise? *Vis. Res.* **37**, 3073-3078.

Benton, C. P., Johnston, A., McOwan, P. W. & Victor, J. D. (2001). Computational modelling of non-Fourier motion: further evidence for a single luminance-based mechanism. *J. Opt. Soc. Am.* A**18**, 2204-2208.

Born, R. T., Groh, J. M., Zhao, R. & Lukasewycz, S. J. (2000). Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements. *Neuron* **26**: 725–734.

Born, R. T., Pack, C. C., Ponce, C. R. & Si Yi (2006). Temporal Evolution of 2-Dimensional Direction Signals Used to Guide Eye Movements. *J Neurophysiol* **95**:284-300.

Born, R. T., Pack, C. C. & Zhao, R. (2002). Integration of motion cues for the initiation of smooth pursuit eye movements. *Prog Brain Res* **140**:225–237.

Bowns, L. (1996). Evidence for a feature tracking explanation of why type II plaids move in the vector sum directions at short durations. *Vision Res* **36**: 3685–3694.

Bowns, L. (2006). `Squaring' is better at predicting plaid motion than the vector average or intersection of constraints. *Perception* **35**, 469-481

Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745- 4765.

Burke, D. & Wenderoth, P. (1993). The effect of interactions between one dimensional component gratings on two dimensional motion perception. *Vision Res.* **33**: 343–350.

Carandini M. & Heeger D. J. (1994). Summation and division by neurons in visual cortex. Science 264:1333–1336.

Carandini, M., Heeger, D. J. & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* **17**:8621-44.

Cavanagh, P. & Mather, G. (1989). Motion: the long and the short of it. *Spat.Vis.* **4**, 103-129.

Champion, R. A., Hammett, S. T., & Thompson, P. G. (2007). Perceived direction of plaid motion is not predicted by component speeds. *Vision Res.* **47**, 375-383.

Chubb, C. & Sperling, G. (1988). Drift-balanced random dot stimuli: a general basis for studying non-Fourier motion perception. *J. Opt. Soc. Am.* **A5,** 1986-2007.

Chubb, C. & Sperling, G. (1989). Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. *Proc. Natl Acad. Sci. USA* **86**, 2985-2989.

Churchland, M. M, Priebe, N. J. & Lisberger, S. G. (2005). Comparison of the spatial limits on direction selectivity in visual areas MT and V1. *J. Neurophysiol.* **93**, 1235-1245.

Derrington, A. M. & Badcock, D. R. (1992). Two-stage analysis of the motion of 2-dimensional patterns: what is the first stage? *Vision Res.* **32,** 691–698.

Derrington, A. M. & Suero, M. (1991). Motion of complex patterns is computed from the perceived motions of their components. *Vision Res.* **31**, 139-149.

Derrington, A. M., & Ukkonen, O.I. (1999). Second-order motion discrimination by feature-tracking. *Vision Res.* **39**, 1465-1475.

Dimova, K. D. & Denham, M. J. (2009). A neurally plausible model of motion integration in smooth eye pursuit based on recursive Bayesian estimation. *Biol. Cybern.* **100**, 185—201.

Donnelly, M., Bowd, C. & Patterson, R. (1997). Direction discrimination of cyclopean (stereoscopic) and luminance motion. *Vision Res* **37(15)**:2041-2046.

Fennema, C. L. & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, **9**: 301-315.

Ferrera, V. & Wilson, H. (1990). Perceived direction of moving two-dimensional patterns. *Vision Res.* **30**, 273–287.

Ferrera, V, & Wilson, H. (1991). Perceived speed of moving two-dimensional patterns. *Vision Res.* **31**, 877–893.

Geisler, W. S., Albrecht, D. G., Crane, A. M. & Stern, L. (2001). Motion direction signals in the primary visual cortex of cat and monkey. *Visual Neurosci.* **18**, 501-516.

Groh, J.H., 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. *J Neurosci* **17**:4312–4330.

Grossberg, S., Mingolla, E. & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research* **41**: 2521-2553.

Heeger, D. J. (1992) Normalization of cell responses in cat striate cortex. Visual Neurosci. 9, 181-197.

Heeger, D. J. & Simoncelli, E. P. (1993). Model of visual motion sensing. In: Harris L, Jenkin M (eds) Spatial Vision in Humans and Robots, Cambridge University Press, pp 367-392.

Heeger, D. J, Simoncelli, E. P. & Movshon, J. A. (1996). Computational models of cortical visual processing. *Proc. Nat. Acad. Sci. U. S. A.* **93**, 623-627.

Hubel, H. D. & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *J Physiol* (Lond) **148**:574-591.

Hubel, H. D. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* (Lond) **160**: 106-154.

Hubel, D. H. & Weisel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J Neurophysiol.* **28**, 229-289.

Hupé, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* **394**, 784-787.

Hupé, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R. & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *J Neurophysiol* **85**: 134–145.

Johnston, A., McOwan, P.W., Benton, C.P. (2003). Biological computation of image motion from flows over boundaries. *Journal of Physiology - Paris* **97**, 325–334.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng **82**: 35-45.

Koechlin, E., Anton, J. L., Burnod, Y. (1999). Bayesian inference in populations of cortical neurons: a model of motion integration and segmentation in area MT. *Biol. Cybern* **80**: 25–44.

Koenderink, J.J. & van Doorn, A.J. (1987). Representation of local geometry in the visual system, *Biol. Cybernetics* **55**, 367–375.

Komatsu, H. & Wurtz, R. H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *J Neurophysiol* **60**: 580–603.

Komatsu, H. & Wurtz, R. H. (1989). Modulation of pursuit eye movements by stimulation of cortical areas MT and MST. *J Neurophysiol* 62: 31–47.

Lamme, V. A. & Roelfsema P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* **23**, 571-579.

Lettvin, J. Y., Maturana, H., McCulloch, W. S. & Pitts, W.H. (1959). What the frog's eye tells the frog's brain. *Proc. I.R.E.* **47**, 1940-1959.

Lidén, L. & Pack, C. (1999). The role of terminators and occlusion cues in motion integration and segmentation: a neural network model. *Vision Res* **39**: 3301–3320.

Lindner, A. & Ilg, U. W. (2000). Initiation of smooth-pursuit eye movements to first-order and second-order motion stimuli. *Exp Brain Res* **133**: 450–456.

Linsker, R. (2008). Neural network learning of optimal Kalman prediction and control. *Neural Networks* **21**, 1328-1343.

Lisberger, S. G. & Westbrook, L. E. (1985). Properties of visual inputs that initiate horizontal smooth pursuit eye movements in monkeys. *J. Neurosci.* **5**: 1662-1673.

Livingstone, M. S., Pack, C. C. & Born, R. T. (2001). Two-dimensional substructure of MT receptive fields. *Neuron* **30**: 781–793.

Löffler, G., Orbach, H. S. (1999). Computing feature motion without feature detectors: a model for terminator motion without end-stopped cells. *Vision Res* **39**: 859–871, 1999.

Lorenceau, J., Shiffrar, M., Wells, N. & Castet, E. (1993). Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Res* **33**: 1207–1217.

Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proc R Soc Lond B Biol Sci* **211**: 151–180.

Masson, G. S. & Castet, E. (2002). Parallel motion processing for the initiation of short-latency ocular following in humans. *J Neurosci* **22**: 5149–5163.

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

Masson, G.S. & Stone, L.S. (2002). From following edges to pursuing objects. *J Neurophysiol* **88**: 2869–2873.

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

Mikami, A, Newsome, W. T. & Wurtz, R. (1986). Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. *J. Neurophysiol.* **55**, 1328-1339.

Mingolla, E, Todd, J. T. & Norman, J. F. (1992). The perception of globally coherent motion. *Vision Res.* **32**, 1015-1031.

Montagnini, A., Mamassian, P., Perrinet, L., Castet, E. & Masson, G.S. (2007). Bayesian modeling of dynamic motion integration. *Journal of Physiology* - Paris **101**: 64-77.

Morrone, M. C., Burr, D. & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. Proc. R. Soc. Lond. B 216, 335-354.

Movshon, J.A., Adelson, E.H., Gizzi, M.S. & Newsome, W.T. (1985). The analysis of moving visual patterns. In Pattern Recognition Mechanisms, ed. C. Chagas, R. Gattass, C.Gross (Pontificiae Academiae Scientiarum Scripta Varia **54**: 117-151). Rome: Vatican Press. (Reprinted in Experimental Brain Research, Supplementum **11**: 117-151, 1986)

Movshon, J.A. & Newsome, W.T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J. Neurosci.* **16**, 7733-7741.

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

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**: 1040-1042.

Pack, C. C., & Born, R. T. (2008). Cortical mechanisms for the integration of visual motion. In: A. I. Basbaum, Akimichi Kaneko, G. M. Shepherd, & G. Westheimer, (Eds.). The Senses: A Comprehensive Reference, Vol **2**, Vision **II** (pp. 189-218), Thomas D. Albright and Richard Masland: Academic Press.

Pack, C. C., Gartland, A. J., Born, R.T. (2004). Integration of contour and terminator signals in visual area MT. *J Neurosci* **24**: 3268–3280.

Pack, C. C., Born, R. T. & Livingstone, M. S. (2003). Two-dimensional substructure of motion and stereo interactions in primary visual cortex of alert monkey. *Neuron* **37**, 525-535.

Pack, C. C., Conway, B. R., Born, R. T. & Livingstone, M. S. (2006). Spatiotemporal structureof nonlinear subunits in macaque visual cortex. *J. Neurosci*. **26**, 893-907.

Pack, C. C., Livingstone, M. S., Duffy, K. R. & Born, R.T. (2003). End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. *Neuron* **39**: 671–680.

Patterson, R., Bowd, C. & Donnelly, M. (1998). The cyclopean (stereoscopic) barber pole illusion. *Vision Res* **38**:2119-25.

Polat, U., Mizobe, K., Pettet, M.W., Kasamatsu, T. & Norcia, A.M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature* **391**: 580–584.

Ullman, S. (1979). The Interpretation of Visual Motion. Cambridge: MIT Press.

Rashbass, C. (1961). The relationship between saccadic and smooth tracking eye movements. *J. Physiol*. **159**, 326-338.

Raiguel, S. E., Lagae, L., Gulya, S. B. & Orban, G. A. (1989). Response latencies of visual cells in macaque areas V1, V2 and V5. *Brain Res* **493**: 155–159.

Rao, R. P. N. (2004). Bayesian computation in recurrent neural circuits. *Neural Comp* **16**:1–38.

Reichard, W. (1961). *Autocorrelation, a principle for the evaluation of sensory* informationby the central nervous system. In *Sensory Communication* (ed. Rosenblith, W. A.), 303-318, Wiley, New York.

Rockland, A. & Knutson, T. (2000). Feedback connections from area MT of the squirrel monkey to areas V1 and V2. *J Comp Neurol.* **425**:345-368.

Rodman, H. R. & Albright, T. D. (1989). Single unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Exp. Brain Res.* **75**, 53-64.

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

Rubin, N. & Hochstein, S. (1993). Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional object. *Vision Res.* **33**, 1385-1396.

Rust, N. C., Mante, V., Simoncelli, E. P., Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nat Neurosci* **9**:1421-1431.

Sceniak, M. P., Ringach, D. L., Hawken, M. J. & Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nat Neurosci* **2**: 733–739.

Sceniak, M. P., Hawken, M. J. & Shapley, R. (2001) Visual spatial characterization of macaque V1 neurons. *J Neurophysiol* **85**: 1873–1887.

Schmolesky, M. T., Youngchang, W., Hanes, D. P., Thompson, K. G., Leutgeb, S. & Schall, J. D. (1998). Signal timing across the macaque visual system. *J Neurophysiol* **79**: 3272–3278.

Shipp, S. & Zeki, S. (1989). The organization of connections between areas V5 and V1 in macaque monkey visual cortex. *Eur J Neurosci* **1**: 309–332.

Simoncelli, E.P. (2003). Local analysis of visual motion. In: Chalupa LM, Werner JS, The Visual Neurosciences, MIT Press, pp 1616-1623.

Simoncelli, E., Adelson, E. & Heeger, D. (1991). Probability distributions of optical flow. In: Proc. IEEE Conf Comput Vision Pattern Recog, IEEE, Washington DC, pp 310-315.

Simoncelli, E. P. & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Res* **38**: 743-761.

Smith, A.T., Singh, K.D., Williams, A.L. & Greenlee, M.W. (2001). Estimating receptive field Size from fMRI Data in human striate and extrastriate visual cortex. *Cerebral Cortex*, **11**, 1182-1190.

Smith, M. A., Majaj, N. J. & Movshon JA (2005). Dynamics of motion signaling by neurons in macaque area MT. *Nature Neuroscience* **8**: 220-228.

Sterzer, P. , Haynes, J. D., & Rees, G. (2006). Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT+/V5.*Neuroimage* **32**, 1308-1316.

Stocker, A. A., Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nat Neurosci* **9**: 578-585.

Stone, L., Thompson, P. (1990). Human speed perception is contrast dependent. *Vision Res* **32**: 1535-1549.

Stone, L., Watson, A. & Mulligan, J. (1990). Effect of contrast on the perceived direction of a moving plaid. *Vision Res* **30**: 1049-1067.

Stoner, G. R. & Albright, T. D. (1992). Neural correlates of perceptual motion coherence. *Nature* **358**, 412-414.

Supèr, H, & Lamme, V.A. (2007). Altered figure-ground perception in monkeys with an extra-striate lesion. *Neuropsychologia* **45**, 3329-3334.

Tinsley, C. J., Webb, B. S., Barraclough, N. E., Vincent C. J., Parker, A. & Derrington, A. M. (2003). The nature of V1 neural responses to 2D moving patterns

depends on receptive field structure in the marmoset monkey. *J. Neurophysiol.* **90**, 930–937.

Van Essen, D., Maunsell, J. & Bixby, J (1981). The middle temporal visual area in macaque: myeloarchitecture, connections, functional properties and topographic organization. *Journal of Comparative Neurology* **199**: 292–326.

Wallace, J. M., Stone, L. S. & Masson, G. S. (2005). Object motion computation for the initiation of smooth pursuit eye movements in humans. *J Neurophysiol* **93**:2279-2293.

Wallach, H. (1935). Über visuell wahrgenommene Bewegungsrichtung. *Psychologische Forschung* **20**: 325-380.

Weiss, Y, & Adelson, E. H. (1998). Slow and Smooth: a Bayesian theory for the combination of local motion signals in human vision. Massachusetts Institute of Technology, Dept. of Brain and Cognitive Sciences, A.I. Memo. No. 1624.

Weiss, E., Simoncelli, E. P. & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat Neurosci* **5**:598-604.

Welch, L. (1989). The perception of moving plaids reveals two processing stages. *Nature* **337**, 734–736.

Wenderoth, P. M., Alais, D. M., Burke, D. C., & van der Zwan, R. (1994). The role of "blobs" in determining the perception of drifting plaids and motion aftereffects. *Perception* **23**, 1163-1169.

Wilson, H. R., Ferrera, V. P., Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neurosci* **9**: 79–97.

Wilson, H. R, & Kim, J. (1994). A model for motion coherence and transparency. *Visual Neuroscience* **11**, 1205-1220.

Wuerger, S., Shapley, R. & Rubin, N. (1996). On the visually perceived direction of motion by Hans Wallach: 60 years later. *Perception*, **25**: 1317-1367.

Yasui S. & Young, L. R. (1975). Perceived visual motion as effective stimulus to pursuit eye movement system. *Science* **190**, 906-908.

Yo, C. & Wilson, H. (1992).Perceived direction of moving two-dimensional patterns depends on duration, contrast, and eccentricity. *Vision Res.* **32**, 135–147.

Appendices

Appendix A

Intuitive explanation of the Kalman filter algorithm

The Kalman filter is a form of recursive least squares estimation. It is based an algorithm which recursively forms the posterior probability distribution of an unknown variable, conditional on a sequence of noisy observations of the variable. As each observation is made, it is used to update the current conditional probability distribution, which is generally assumed to be of Gaussian form. The best estimate of the unknown variable is then the mean of the distribution, since this value maximises the conditional posterior probability. The variance of the estimate corresponds to the variance of the posterior distribution.

As a simple example, assume that a sequence of observations/measurements h_k are made of an unknown variable v_k. Each observation of v_k is corrupted by measurement noise, i.e.

$$h_k = v_k + \eta_k \tag{A.1}$$

where η_k is assumed to be a Gaussian distributed random noise sequence with zero mean and variance σ^2. The objective is to obtain the best estimate of v_k from the sequence of observations, in the case where a new estimate is made after each observation, based on the previous estimate. Thus the estimation is made recursively. In this example, the unknown variable v_k is assumed to remain constant, i.e.

$$v_{k+1} = v_k = v \tag{A.2}$$

However, in the general case, the variable can change at each step, according to a dynamic equation:

$$v_{k+1} = \phi_k v_k \tag{A.3}$$

The estimation algorithm begins $(k = 0)$ by making a prior assumption of the probability distribution of v_0. Assume that this is a Gaussian distribution with mean value $\overline{v}_0 = 0$ and variance $\sigma^2_{p,0}$, the so-called **prior distribution**. The best prior estimate of v_0 is the mean (maximum) of the prior distribution, $\overline{v}_0 = 0$. The variance of the best estimate of v_0 is thus the variance of the prior distribution, $\sigma^2_{p,0}$.

In the following description of the Kalman filter algorithm, we denote

v_k^j - the estimate of v_k, based on the observations up to and including the j th one

P_k^j - the variance of the estimate of v_k, based on the observations up to and including the j th one.

The algorithm is thus **initialised** with the mean and variance values of the prior distribution:

$$v_0^0 = \overline{v}_0 = 0$$
$$P_0^0 = \sigma_{p,0}^2$$

and the first **estimate update step** is then made:

$$v_1^0 = v_0^0 = 0$$
$$P_1^0 = P_0^0$$

Next the first observation h_1 is made. The conditional probability distribution $p(v_1 \mid h_1)$ has a mean value, based on the value h_1 of the observation, of:

$$\overline{v}_1 = \frac{\sigma_{p,0}^2}{\sigma_{p,0}^2 + \sigma^2} \cdot h_1 = K_1 h_1 \tag{A.4}$$

where

$$K_1 \triangleq \frac{\sigma_{p,0}^2}{\sigma_{p,0}^2 + \sigma^2} \tag{A.5}$$

whilst the variance $\sigma_{p,1}^2$ of $p(v_1 \mid h_1)$ is given by:

$$\frac{1}{\sigma_{p,1}^2} = \frac{1}{\sigma_{p,0}^2} + \frac{1}{\sigma^2} \tag{A.6}$$

This latter equation can be rewritten as:

$$\sigma_{p,1}^2 = \sigma_{p,0}^2 - K_1 \sigma_{p,0}^2 \tag{A.7}$$

The best estimate v_1^1 of v_1 is given by the mean of the distribution $p(v_1 \mid h_1)$, i.e. by $\overline{v}_1$, and its variance is given by the variance of $p(v_1 \mid h_1)$.

Thus we can write the **estimate update step at first observation**, using the above equations, as:

$$v_1^1 = K_1 h_1$$
$$P_1^1 = P_1^0 - K_1 P_1^0$$
$$K_1 = P_1^0 / (P_1^0 + \sigma^2)$$

Next the estimate and its variance are **updated between the first and second observations,** based on the assumption that the unknown variable remains constant, i.e. $v_{k+1} = v_k = v$. Thus

$$v_2^1 = v_1^1$$
$$P_2^1 = P_1^1$$

Next a second observation h_2 is made. The aim now is to obtain the parameters of the conditional probability distribution $p(v_2 \mid h_1, h_2)$. The mean of this distribution is based on the mean of $p(v_1 \mid h_1)$ and the new observation value h_2, and is given by:

$$\overline{v}_2 = \frac{\sigma^2}{\sigma_{p,1}^2 + \sigma^2} \cdot v_2^1 + \frac{\sigma_{p,1}^2}{\sigma_{p,1}^2 + \sigma^2} \cdot h_2 \qquad \text{(A.8)}$$

whilst its variance $\sigma_{p,2}^2$ is given by:

$$\frac{1}{\sigma_{p,2}^2} = \frac{1}{\sigma_{p,1}^2} + \frac{1}{\sigma^2} \qquad \text{(A.9)}$$

The best estimate of v_2^2 of v_2, that maximises $p(v_2 \mid h_1, h_2)$ is thus the mean, $\overline{v}_2$, the expression for which can be rewritten from (A.8) as:

$$\overline{v}_2 = v_2^1 + \frac{\sigma_{p,1}^2}{\sigma_{p,1}^2 + \sigma^2} \cdot (h_2 - v_2^1) = K_2 (h_2 - v_2^1) \qquad \text{(A.10)}$$

where

$$K_2 = \frac{\sigma_{p,1}^2}{\sigma_{p,1}^2 + \sigma^2} \tag{A.11}$$

Similarly the variance of this estimate can be rewritten from (A.9) and (A.11) as:

$$\sigma_{p,2}^2 = \sigma_{p,1}^2 - K_2 \sigma_{p,1}^2 \tag{A.12}$$

Thus we can write the **estimate update step at second observation**, using the above equations (A.10) to (A.12), as

$$v_2^2 = v_2^1 + K_2(h_1 - v_2^1)$$
$$P_2^2 = P_2^1 - K_2 P_2^1$$
$$K_2 = P_2^1 / (P_2^1 + \sigma^2)$$

As before, this estimate and its variance are **updated between the second and third observations,** based on the constancy assumption for the unknown variable. Thus

$$v_3^2 = v_2^2$$
$$P_3^2 = P_2^2$$

This set of update equations are then repeated for the third observation h_3 and for all subsequent observations. Reviewing the update equations (A.10) to (A.12), it can readily be seen that the algorithm can be written in a general form as:

1. Initialisation

$$v_0^0 = \overline{v}_0$$
$$P_0^0 = \sigma_{p,0}^2$$

2. Update of estimate between observations

$$v_{k+1}^k = v_k^k$$
$$P_{k+1}^k = P_k^k$$

3. Update of estimate based on k th observation

$$v_k^k = v_k^{k-1} + K_k(h_k - v_k^{k-1})$$
$$P_k^k = P_k^{k-1} - K_k P_k^{k-1}$$
$$K_k = P_k^{k-1} / (P_k^{k-1} + \sigma^2)$$

A diagram of the general form of the algorithm, in which the unknown variable v_k is assumed to change dynamically between observations, is shown Figure A.1, illustrating the algorithm's recursive form.

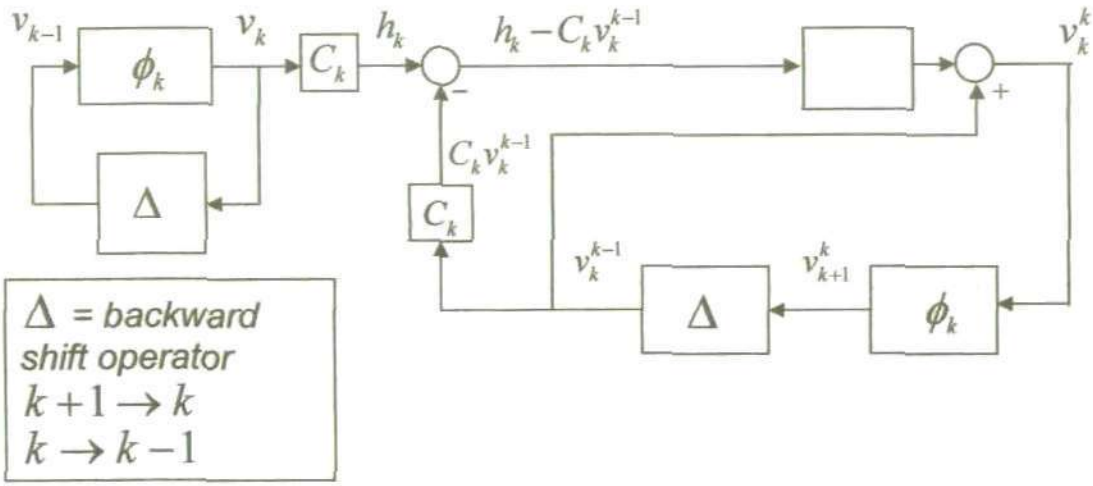


Figure A.1 Diagram of the Kalman filter algorithm for the general case, where the dynamics and observation of the unknown variable are described by $v_{k+1} = \phi_k v_k$, $h_k = C_k v_k + \eta_k$. In the case described in the text, the unknown variable is constant, i.e. $v_{k+1} \approx v_k = v$.

Dimova, K. D. & Denham, M. J. (2009). **A neurally plausible model of the dynamics of motion integration in smooth eye pursuit based on recursive Bayesian estimation.** ***Biological Cybernetics***, 100(3):185-201

A neurally plausible model of the dynamics of motion integration in smooth eye pursuit based on recursive Bayesian estimation

Kameliya Dimova* and Michael Denham

Centre for Computational and Theoretical Neuroscience, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK

* Corresponding author:

tel.: +44 1752 233359

fax.: +44 1752 233349

email: kameliya.dimova@plymouth.ac.uk

Dimova, K. D. & Denham, M. J. (2009). **A neurally plausible model of the dynamics of motion integration in smooth eye pursuit based on recursive Bayesian estimation.** ***Biological Cybernetics*****,** 100(3):185-201

Abstract

In this study, we describe a model of motion integration in smooth eye pursuit based on a recursive Bayesian estimation process, which displays a dynamic behaviour qualitatively similar to the dynamics of the motion integration process observed experimentally, both psychophysically in humans and monkeys, and physiologically in monkeys. By formulating the model as an approximate version of a Kalman filter algorithm, we have been able to show that it can be put into a neurally plausible, distributed recurrent form which coarsely corresponds to the recurrent circuitry of visual cortical areas V1 and MT. The model thus provides further support for the notion that the motion integration process is based on a form of Bayesian estimation, as has been suggested by many psychophysical studies, and moreover suggests that the observed dynamic properties of this process are the result of the recursive nature of the motion estimation.

1 Introduction

Smooth pursuit eye movements are the rotations of the eye which an observer performs in order that a moving target object is held in an approximately stationary position on the foveal region of the retina so that it can be subjected to visual processing at a high level of acuity. This is a valuable phenomenon for theoretical brain studies as it is relatively easy to access experimentally, in terms of the ease of measurement and manipulation of both the inputs to the process, the motion of the visual target, and the outputs, the eye movements. It is likely that the neural *mechanism involved in smooth pursuit eye motion is an oculomotor feedback control* mechanism in which the motion of the image of the target on the retina is used as a driving signal for the generation of eye movements which aim at minimising this motion signal. As with any such feedback control mechanism, under steady state conditions the response properties of the system tend to be dominated by the feedback gain, and any knowledge about the neural circuits in the feed-forward pathway is impossible to obtain without breaking the feedback loop in some way. Since this is experimentally infeasible, other ways have been sought by many researchers of revealing the successive stages of neural processing involved in the system and their interactions.

In particular, there has been a substantial effort to understand the neural mechanisms which underlie one of the earliest stages of processing in the oculomotor system, in which the local retinal motion information is integrated to provide accurate global information on motion of the object, the image of which must be held steady on the retina. It is well known that owing to the small receptive field size of early visual neurons in the retina, thalamus and primary visual cortex, some local information about motion of the object is potentially inaccurate owing to the ambiguity introduced by the so-called "aperture effect" (Wallach, 1935 (English translation in Wuerger *et al,* 1996); Fennema and Thompson, 1979; Marr and Ullman, 1981). The ambiguity results not from the "aperture" *per se*, but is owing to the fact that the small receptive field may only give access to a one-dimensional motion stimulus, eg a single contour edge, which will behave identically for a range of stimulus motion directions,

corresponding to a constraint line in velocity space. If the stimulus within the aperture has two-dimensional information, e.g. two contour edges forming a corner, then unambiguous directional information is available for the stimulus, since the intersection of the corresponding two constraint lines uniquely defines the motion direction. To ensure that object motion is accurately tracked by the oculomotor system, all the available local information, both unambiguous (2-D) and ambiguous (1-D) must be integrated, in order to provide an estimate of the object motion which will drive the eye movement in the correct target direction.

In order to study the neural mechanisms involved in this integration process, several researchers have investigated the short period at the initiation of pursuit eye movement in which the ocular-motor feedback control system effectively operates in "open-loop" mode, owing to the time delay involved in the transmission of information to the ocular motor plant and its inherent response dynamics (Lisberger and Westbrook, 1985). From the time of onset motion of the visual target stimulus which is to be tracked, there is a time delay of about 100 ms before the visually driven eye pursuit motor response begins. As already mentioned, for certain visual stimuli, some of the signals which convey local motion of the target image do not accurately reflect the target motion, owing to the aperture effect. Experimental evidence from human and non-human primates indicates that, as a result, the initial eye pursuit movement has both an on-axis component, ie in the direction of the object motion, and an off-axis component which reflects the inaccuracies in the local motion signals (Masson & Stone, 2002; Pack and Born, 2001).

The experimental evidence suggests that the integration process begins as soon after the onset of the target motion as a visual motion signal is available, ie at the latency of transfer of the retinal motion signal to the site of the neural integration process. In particular, if the neural mechanisms of the integration process are located in visual area MT, physiological experiments (Pack and Born, 2001) suggest that this latency is about 70-80 ms in monkeys. At this point in time the eye movement response has yet to start owing to its ~100ms latency, so the error between the target motion and the eye motion increases without compensation from any eye movement (Lisberger and

Westbrook, 1985). However, the experimental data of Wallace *et al* (2005) and of Born *et al* (2006), suggest that the integration process is already starting to correct for the inaccurate off-axis motion signals, and this can be observed in the eye movement within ~70 ms after it begins. This is before the error signal induced by the inaccurate off-axis eye movement can be compensated for by the oculomotor feedback system, owing to the ~100 ms delay in this system. Their results indicate that for the first 70 ms or so the eye movement is driven by a simple uncorrected pooling of both the unambiguous (2-D) and ambiguous (1-D) local motion measurements, with the 2-D measurements of the target object velocity only starting to dominate after this period. The off-axis direction error in the eye movement then decays to zero over a further period of 200 - 300 ms (see Figure 3 of Wallace *et al*, 2005). This temporal evolution toward an accurate representation of target object motion is consistent with several experimental results: from human direction judgments (Lorenceau *et al*, 1993); the human ocular following response (Masson and Castet, 2002; Masson *et al* 2000), smooth pursuit in monkeys (Pack and Born, 2001) and humans (Lindner and Ilg, 2000; Masson and Stone, 2002, Wallace *et al*, 2005), and the responses of MT neurons in alert monkeys (Pack and Born, 2001; Pack *et al*, 2004).

The neural process of integration of the local motion signals is generally supposed to involve two stages. The first stage, which is usually attributed to neural mechanisms in primary visual cortex (V1), involves the extraction of directionally selective motion information. Since V1 neurons are subject to the ambiguities introduced by 1-D motion stimulus signals, a second stage, attributed to the medial temporal area of cortex (MT), is perceived to be the location of the neural mechanisms for integration of the local motion signals and the resolution of the ambiguities introduced by the 1-D signals. One model based on this two stage process supposes a feed-forward mechanism which achieves the integration process by differential weighting of the feed-forward projections of the 1-D and 2-D signals (Simoncelli and Heeger, 1998). This model does not apparently account however for the temporal dynamics of the motion integration process observed in the above cited experimental data.

An alternative model (Born *et al*, 2006) suggests that the dynamics of the integration process might be determined by the properties of visual neurons early in the pursuit pathway, in particular a temporal delay in the emergence of end-stopping in direction-selective neurons (Pack *et al*, 2003). In this formulation, the temporal evolution of 2D motion signals for pursuit and those that have been observed in MT neurons (Pack and Born 2001; Pack *et al*, 2004) reflect a change in the "weighting" applied to the outputs of V1 direction-selective neurons, whereby the activity of 2-D related end-stopping neurons eventually suppresses that of the 1-D contour-related neurons, and the motion signals from the 2-D terminators becomes dominant. It is suggested that this mechanism might also explain the perceptual dominance of a contour-vector average for stimuli of *low contrast* (Weiss *et al*, 2002), since end-stopping is weak or absent for such stimuli (Polat *et al*, 1998; Sceniak *et al*, 1999). This is also consistent with the experimental observations of Wallace *et al* (2005) on pursuit initiation in humans, in which the effect of lowering the stimulus contrast is an increase in the off-axis bias in the initial transient eye movement and a lengthening of the time taken for this bias to reduce to zero. Note however that for the low contrast stimulus, the off-axis motion is eventually eliminated, albeit with a time constant up to nearly three times that for the high contrast stimulus, indicating that the suppressive mechanism, if this is indeed the case, does still operate although more weakly. As pointed out in Born *et al* (2006), the suggested role of the end-stopping cells is also physiologically very plausible given the fact that neurons in layer 4B of V1, which is the source of the main projection from V1 to MT (Maunsell and Van Essen, 1983; Shipp and Zeki, 1989), are strongly end-stopped (Sceniak *et al*, 2001).

An alternative model, which also captures the dynamics of the integration process, is proposed here. Our model is based on the concept that the integration process computes a Bayesian estimate of the target object velocity using the local observations of target motion provided by the directionally selective neurons in V1. We implement the estimation process recursively, in the form of a Kalman filter (Kalman, 1960), and show that the dynamics of this recursive estimation process closely replicates the dynamics of the motion integration process, as measured

experimentally, under a variety of conditions involving changes in shape and contrast of the stimulus.

The application of Kalman filtering in modeling smooth pursuit eye movements is not new. Shibata *et al* (2005) address how the remarkably high level of accuracy (zero-lag) of the primate smooth pursuit system, as identified experimentally, might be modeled by a predictive controller together with fast learning of the target motion dynamics, using a form of Kalman filter for target-state estimation and prediction. Whilst their model thus addresses a different problem from ours, it is however of relevance to our model in the following sense. The velocity estimation process in our model depends on an internal generative model of target velocity. In our simulations, and the experiments they seek to replicate, the target motion is constant, and therefore the internal model reflects this. When the target motion is changing dynamically, our model would need to incorporate a process for fast learning of this dynamics, in order to produce an accurate estimate of the target velocity, and the Shibata et al (2005) model would provide this.

Bayesian models of motion integration and estimation have also been previously proposed (Koechlin *et al*, 1999; Weiss *et al*, 2002; Rao, 2004). In particular, Weiss *et al* (2002) proposed an optimal Bayesian probabilistic model of motion integration in which an optimal estimate of the target object velocity is computed using the likelihoods of the local velocity measurements distributed over the visual space (Simoncelli *et al*, 1991; Simoncelli, 2003). They used a priori distribution for the target object velocity which assumes a human preference for the assumption of slow speeds, and formulated this as a Gaussian prior centred on zero. They also assumed an additive measurement noise which was independently Gaussian distributed, with zero mean and known variance. This had the effect of making the local likelihood functions dependent on stimulus contrast. They computed the posterior distribution of the velocity as the product of the likelihoods over all the spatial locations (assuming that the likelihoods are independent) multiplied by the prior.

The authors show that their model reflects several of the main characteristics of human motion perception, as observed in a range of psychophysical studies, both their own experiments which use "fat" and "thin" rhombus figures at high and low contrast, and those of others (Burke and Wenderoth, 1993; Bowns, 1996; Stone *et al*, 1990; Stone and Thompson, 1990; Lorenceau *et al*, 1993) using mainly moving plaid patterns. For fixed values of measurement noise variance and prior variance (or more specifically, a fixed ratio of these quantities), and for a moving "thin" rhombus stimulus, at low contrast the posterior velocity distribution has a maximum (mean) corresponding to the vector average of the local velocities. At high contrast, this maximum occurs at, or near, the veridical target object velocity given the intersection of the constraint lines provided by the local velocity measurements.

The model of Weiss *et al* (2002) was not intended to replicate the experimental data from the initial eye pursuit experiments cited above, and indeed is not capable of reproducing the dynamics of the motion integration process as observed in these experiments. Moreover, their model, tuned to replicate their human perceptual experiments, would predict that a stimulus consisting of a horizontally moving high contrast "thin" rhombus would not result in any offset bias in the pursuit eye motion, whereas the experimental data of Wallace *et al* (2005) would indicate that there is always an offset bias in the initial eye motion of ~30° even in the case of a "thin" rhombus of high contrast and slow speed.

2 A model of the motion integration process based on the Kalman filter estimation algorithm

We show that the motion integration process can be modelled as a recursive Bayesian estimation algorithm, based upon a spatially distributed set of local observations of spatial and temporal changes in the image intensity. Thus, at each local position *(x,y)* in the visual space, identified with the receptive field positions of V1 neurons, we assume that observations are made of the change in the intensity of the image $I(x,y,t)$ over a small change in time Δt, and over a small spatial change in the two cardinal directions (x,y), i.e. $I_t(x,y,t)$, $I_x(x,y,t)$ and $I_y(x,y,t)$, where we define

$$I_t(x,y,t) \triangleq \left[I(x,y,t+\Delta t) - I(x,y,t)\right]/\Delta t$$
$$I_x(x,y,t) \triangleq \left[I(x+\Delta x,y,t) - I(x,y,t)\right]/\Delta x$$
$$I_y(x,y,t) \triangleq \left[I(x,y+\Delta y,t) - I(x,y,t)\right]/\Delta y$$

As is standard in the formulation of the motion estimation problem, we assume that image intensity changes with time at a location *(x,y)* are only the result of motion of the image. Then it follows that, if v_x and v_y are the x and y components of the image velocity vector v,

$$\begin{aligned} I(x,y,t) &= I(x+\Delta x, y+\Delta y, t+\Delta t) \\ &= I(x+\Delta t.v_x, y+\Delta t.v_y, t+\Delta t) \end{aligned} \tag{1}$$

If we approximate the right hand side of (1) by its first order Taylor series expansion, we get

$$I(x,y,t) \approx I(x,y,t) + v_x I_x(x,y,t) + v_y I_y(x,y,t) + I_t(x,y,t) \tag{2}$$

We can then rewrite this equation as

$$I_t(x,y,t) = -\begin{bmatrix} I_x(x,y,t) & I_y(x,y,t) \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} + \eta(x,y,t) \tag{3}$$

where $\eta(x,y,t)$ is a zero mean, Gaussian distributed noise process with variance σ^2, representing the measurement error. (Weiss *et al*, 2002; Heeger and Simoncelli, 1992; Fennema and Thompson, 1979).

We can think of equation (3) as an observation equation for the unknown image velocity vector v, in which I_t, I_x, and I_y are all measured (observed) values, and rewrite it in the discrete time form as a sequence of observations of the unknown vector v_k, ie

$$h_k = C_k v_k + \eta_k \tag{4}$$

where we define

$$h_k \triangleq I_t(x,y,t)$$
$$C_k \triangleq -\begin{bmatrix} I_x(x,y,t) & I_y(x,y,t) \end{bmatrix}$$
$$t \triangleq k.\Delta t$$

and where η_k is a zero mean, Gaussian distributed noise sequence with variance σ^2, representing the measurement error in the observation equation.

The process of estimating the velocity vector v becomes one of maximising the posterior probability density function $p(v_0,...,v_N \mid h_1,...,h_N)$ with respect to $\{v_0,...,v_N\}$. Bayes' rule gives us that

$$p(v_0,...,v_N \mid h_1,...,h_N) = \frac{p(h_1,...,h_N \mid v_0,...,v_N)\,p(v_0,...,v_N)}{p(h_1,...,h_N)} \tag{5}$$

We can write

$$p(h_1,...,h_N \mid v_0,...,v_N) = \prod_{k=1}^{N} p_{h_k}(h_k - C_k v_k) \tag{6}$$

and if we assume that the image velocity is constant

$$p(v_0,...,v_N) = p(v_0) \tag{7}$$

Thus Bayes' rule becomes

$$\begin{aligned} p(v_0,...,v_N \mid h_1,...,h_N) &\propto p(v_0)\prod_{k=1}^{N} p_{h_k}(h_k - C_k v_k) \\ &\propto \exp\left\{-\frac{1}{2}(v_0-\overline{v_0})^T P_0^{-1}(v_0-\overline{v_0}) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(h_k - C_k v_k)^T(h_k - C_k v_k)\right\} \end{aligned} \tag{8}$$

where $\overline{v_0}$ and P_0 are, respectively, prior estimates of the mean and covariance matrix of the probability distribution of the image velocity vector v_k.

Rather than find the maximum of this posterior distribution directly, we formulate the process as a recursive estimation procedure, ie in the form of a Kalman filter estimation algorithm:

1. initialization:

$$v_0^0 = \overline{v_0}$$

$$P_0^0 = P_0$$

2. update of the estimate between observations (this captures the dynamics of the stimulus – in this case the stimulus velocity is constant and so the estimate is unchanged between observations):

$$v_{k+1}^k = v_k^k$$

$$P_{k+1}^k = P_k^k$$

3. update of the estimate at observation *k*

$$v_k^k = v_k^{k-1} + K_k(h_k - C_k v_k^{k-1})$$
$$P_k^k = P_k^{k-1} - K_k C_k P_k^{k-1}$$
$$K_k = P_k^{k-1} C_k^T (C_k P_k^{k-1} C_k^T + \sigma^2 I_2)^{-1}$$

where I_2 denotes the 2x2 identity matrix.

Alternatively we can write the so-called *Kalman gain* term K_k as

$$K_k = [(P_k^{k-1})^{-1} + C_k^T C_k \frac{1}{\sigma^2}]^{-1} C_k^T \frac{1}{\sigma^2} \qquad (9)$$

If we assume that $P_k^{k-1} = \sigma_p^2 I_2$, then

$$\begin{aligned} K_k &= [\frac{1}{\sigma_p^{\ 2}} I_2 + C_k^{\ T} C_k \frac{1}{\sigma^2}]^{-1} C_k^{\ T} \frac{1}{\sigma^2} \\ &= [\frac{\sigma^2}{\sigma_p^2} I_2 + C_k^{\ T} C_k]^{-1} C_k^{\ T} \end{aligned} \tag{10}$$

Substituting for C_k in (10), we get

$$K_k = \begin{bmatrix} I_x^2 + \frac{\sigma^2}{\sigma_p^2} & I_x I_y \\ I_x I_y & I_y^2 + \frac{\sigma^2}{\sigma_p^2} \end{bmatrix}^{-1} \begin{bmatrix} I_x \\ I_y \end{bmatrix} \tag{11}$$

If we chose $\overline{v_0}$, the prior estimate of the mean of the probability distribution of the image velocity vector v_k, to be zero in the initialisation step in the Kalman filter estimation algorithm, then the first estimate update step of the algorithm gives

$$v_1^1 = K_1 h_1 \tag{12}$$

Substituting for K_1 and h_1 in this equation yields

$$v_1^1 = \begin{bmatrix} I_x^2 + \frac{\sigma^2}{\sigma_p^2} & I_x I_y \\ I_x I_y & I_y^2 + \frac{\sigma^2}{\sigma_p^2} \end{bmatrix}^{-1} \begin{bmatrix} I_x \\ I_y \end{bmatrix} I_t \tag{13}$$

If we generalize the estimation process described above to the case where we have n observation equations for the unknown image velocity vector v at n spatial locations, ie in the definitions

$$h_k \triangleq I_t(x,y,t+\Delta t)$$
$$C_k \triangleq \begin{bmatrix} I_x(x,y,t) & I_y(x,y,t) \end{bmatrix}$$

h_k is now an n vector representing the local observations of the temporal derivatives of the image intensity at n spatial locations (x,y), and C_k is correspondingly an $n \times 2$ matrix of the local observations of the spatial derivatives of the image intensity at the n spatial locations (x,y), then the equation for the initial estimate of the velocity vector becomes

$$v_1^1 = \begin{bmatrix} \sum(I_x^2) + \frac{\sigma^2}{\sigma_p^2} & \sum(I_x I_y) \\ \sum(I_x I_y) & \sum(I_y^2) + \frac{\sigma^2}{\sigma_p^2} \end{bmatrix}^{-1} \begin{bmatrix} \sum(I_x I_t) \\ \sum(I_y I_t) \end{bmatrix} \qquad (14)$$

in which the summations are over the local spatial and temporal derivatives of the image intensity at the n locations, and we assume that the measurement noise variance σ^2 is the same for all n locations.

The above equation (14) is precisely that derived by Weiss *et al* (2002) for the velocity estimate v^* of their "ideal observer", as the mean of the posterior distribution (equation (1) of Weiss *et al*, 2002). In other words, if the prior estimate of the covariance matrix of the probability distribution of the velocity vector v_0 is given as $P_0^0 = \sigma_p^2 I_2$ then the initial step of the Kalman filter estimation algorithm will compute a velocity estimate v_1^1 which corresponds to the velocity estimate v^* of the "ideal observer" of Weiss *et al* (2002). However, in our model of motion integration for smooth pursuit, we assume that the integration process continues to use successive observations of the temporal and spatial changes in image intensity in order to recursively update the estimates of the velocity vector v_k^k and its covariance matrix P_k^k, as expressed in the Kalman filter algorithm. This implies that the Kalman filter algorithm will take some time to converge to an optimal (least squares) estimate of the mean and covariance of the image velocity vector v, based on this sequence of

observations. We postulate that the dynamical behaviour of this recursive estimation process corresponds to the dynamics of the integration process in the initial period of motion pursuit, and show below that the recursive estimation process dynamics reflects several key characteristics of the experimentally observed integration process dynamics.

It is clear however from the above analysis that the velocity estimate given by the first step of the Kalman filter algorithm coincides with the optimal velocity estimate of Weiss et al's ideal observer. Therefore, if the "free parameter" (Weiss *et al*, 2002) of the optimal estimate (14), the σ / σ_p ratio, is set in the algorithm to the same value as in the "ideal observer of Weiss *et al*, our model of the motion integration process will also suffer from the fact that it will predict that a stimulus consisting of a horizontally moving high contrast "thin" rhombus will not result in any initial offset bias in the pursuit eye motion, contrary to the experimental observations of Wallace et al (2005). However, in the following section, where we describe our computer simulations of the Kalman filter estimation algorithm as a model of the motion integration process, we show that for certain values of the σ / σ_p ratio, the high contrast "thin" rhombus will result in an offset bias.

In the later sections of the paper, we show how the Kalman filter based model can be formulated so that it is amenable to implementation as a neural computation process. This results in an approximate, sub-optimal recursive estimation process which nevertheless retains its close relationship to the experimentally observed integration process. In fact in the approximate, neural computation based form, our simulations show that the closeness of fit of the behaviour of the model to the experimental data improves over the optimal Kalman filter based form of the model. We also discuss how this neural computation based model might be mapped onto the neural circuitry involved in the interaction between the V1 and MT areas of cortex, in a way that mimics of the distributed, recurrent nature of the V1-MT circuitry. This is important since the V1-MT circuitry has been identified by many researchers as the location for the motion integration processing stage of the oculomotor system (Groh *et al*, 1997;

Simoncelli, 2003; Simoncelli and Heeger, 1998; Pack and Born, 2001; Pack *et al*, 2003; Pack *et al*, 2004)

3 Simulation experiments using the Kalman filter based motion integration model

As in Wallace *et al* (2005), two visual stimuli are used in the computer simulation experiments of our motion integration model, in our case as solid figures: (i) a square diamond with main axes at 0° and 90°, and (ii) an elongated tilted diamond |(or "thin rhombus") with the main axis at 45° and internal angles of 10° and 170°. These stimuli are presented on a visual space which is represented by a set of 200 by 200 pixels. Each of the stimuli is contained in a square area of 50 by 50 pixels.

The visual space is further divided into 400 uniformly spaced and sized, non-overlapping square windows each of size 10 by 10 pixels. Thus the stimulus is contained within an area covered by 25 windows. In each window, the observations of temporal and spatial derivatives of image intensity are calculated as the image moves at a constant velocity in the horizontal direction from the left to the right of the visual space. Temporal and spatial derivatives $I_t(x,y,t)$, $I_x(x,y,t)$and $I_y(x,y,t)$ are computed using small spatial and temporal shifts $\Delta x, \Delta y,$and Δt of the stimulus.

Figure 1 shows the results from the computer simulation of the model in estimating the angular direction of the rhombus (0° = horizontal motion) of both the square diamond and thin rhombus stimuli, for three values of contrast: max (image intensity = 1), half (intensity = 0.5) and quarter (intensity = 0.25), and for four different values of the ratio σ / σ_p(shown in Table 1), as a function of the iterations of the Kalman filter algorithm. Each iteration shown corresponds to five iterations of the Kalman filter algorithm, and the estimated angular direction shown is the average of the estimated direction over these five iterations (thus in each plot the total number of iterations of the algorithm is thirty five).

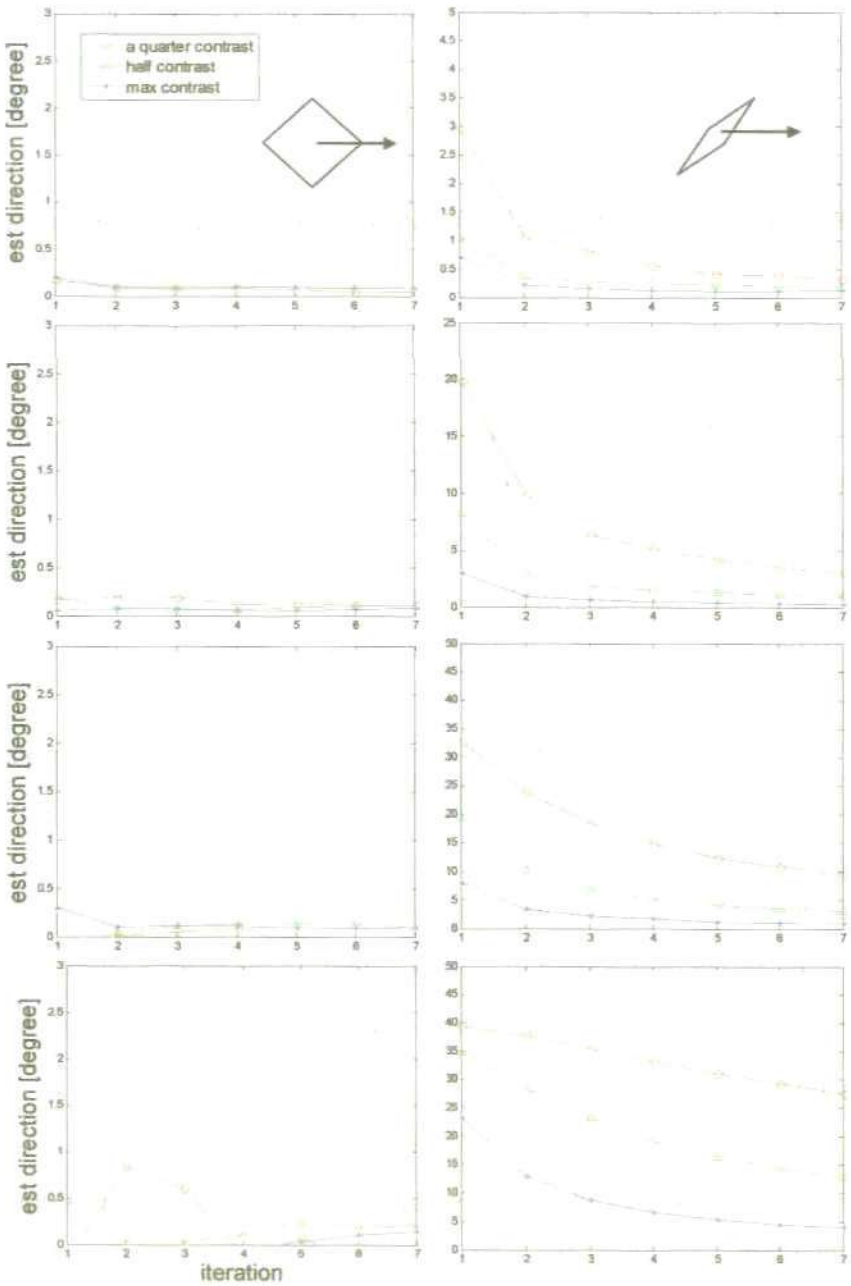


Figure 1. Results from the computer simulations of the motion integration model based on the optimal Kalman filter algorithm, showing the estimated angular direction of the stimulus (0° = horizontal motion), for horizontally moving diamond and rhombus stimuli, for three different values of contrast, and for four different values of the ratio σ / σ_p increasing from the top row to the bottom row of plots. Note the change in the vertical scale in the plots in the first two rows. The values of σ / σ_p used are given in Table 1.

σ	σ_p	$\frac{\sigma}{\sigma_p}$
0.04	0.07	0.6
0.1	0.04	2.5
0.2	0.04	5.0
0.5	0.04	12.5

Table 1. The values of σ / σ_p used in the model simulations

There is a small, but non-zero peak offset bias for all values of σ / σ_p (note the change in the vertical scale in the plots in the first two rows). The value of σ / σ_p is increased from the top to the bottom row in Figure 1, and results in a corresponding increase in the offset bias in the estimated direction. This offset bias is however very small in the case of the lowest value of the σ / σ_p ratio (top row of plots) for both the diamond and thin rhombus shapes (< 0.25° for the diamond and < 3° for the thin rhombus). In each case, the peak offset bias is followed by a temporal decay of this bias to an asymptotic value of close to 0°.

These characteristics of the model dynamics, as shown in Figure 1, ie an initial offset bias followed by a regular temporal decay of this bias to zero, are also the main characteristics of the dynamics of the motion integration process, as observed experimentally by Wallace *et al* (2005). Their data show (see their Figures 5 and 6) a peak tracking direction error followed by a temporal decay toward an asymptotic value corresponding to the true target motion direction (0°). Moreover, the model simulations show (Figure 1) that decreasing the contrast of the stimulus, for any fixed value of the ratio σ / σ_p, results in an increase in both the peak offset bias in the estimated velocity and the time constant for the decay of the bias to zero. This is again in close qualitative agreement with the data of Wallace *et al* (2005), in which the initial tracking bias was reduced from ~44° to ~30° when the contrast was increased from 10% to 40%, after which there was little further reduction. Similarly, increasing

the target contrast resulted in a decrease in the decay time constant from ~168 ms at 10% contrast to an asymptotic value ~60 ms for contrasts >40%.

Overall, the model simulation results show a qualitatively, and in specific instances quantitatively, high level of correspondence with the experimental results of Wallace *et al* (2005), in particular in respect of the variation of both the peak offset bias in the target velocity and the decay of this bias to zero, with contrast level.

4 A neurally plausible model of the motion integration process based on an approximate form of the Kalman filter estimation algorithm

It is clear from the description of the Kalman filter algorithm in section 2 above, that the implementation of the model based on this algorithm would require the computation of an inverse matrix in (9), in order to calculate the Kalman gain matrix K_k. Such a calculation is not plausible as a neural computation, and therefore it is necessary to find an approximate, sub-optimal version of the algorithm which avoids this matrix inversion.

In the case of n local observations, each subject to a different, mutually independent measurement noise, ie with different variances, equation (9) becomes

$$K_k = [(P_k^{k-1})^{-1} + C_k^T \Sigma^{-1} C_k]^{-1} C_k^T \Sigma^{-1} \tag{15}$$

in which the measurement noise covariance matrix is denoted by $\Sigma \triangleq diag(\sigma_i^2)$, $i = 1,...,n$. To avoid the matrix inversion in (15), we first approximate the covariance matrix of the velocity estimate by a diagonal matrix, i.e.

$$P_k^{k-1} \triangleq \begin{bmatrix} (\sigma_{p,x}^2)_k^{k-1} & 0 \\ 0 & (\sigma_{p,y}^2)_k^{k-1} \end{bmatrix} \tag{16}$$

This assumes that the estimate of the velocity vector is uncorrelated in the x and y directions. Then the first term in the matrix inverse in (15) is given by

$$\left(P_k^{k-1}\right)^{-1} = \begin{bmatrix} 1/(\sigma_{p,x}^2)_k^{k-1} & 0 \\ 0 & 1/(\sigma_{p,y}^2)_k^{k-1} \end{bmatrix} \tag{17}$$

The second term is

$$C_k^T \Sigma^{-1} C_k = \begin{bmatrix} \sum_i \frac{1}{\sigma_i^2} I_{x,i}^2 & \sum_i \frac{1}{\sigma_i^2} I_{x,i} I_{y,i} \\ \sum_i \frac{1}{\sigma_i^2} I_{x,i} I_{y,i} & \sum_i \frac{1}{\sigma_i^2} I_{y,i}^2 \end{bmatrix} \tag{18}$$

We now assume that this can also be approximated by a diagonal matrix, ie

$$C_k^T \Sigma^{-1} C_k \simeq \begin{bmatrix} \sum_i \frac{1}{\sigma_i^2} I_{x,i}^2 & 0 \\ 0 & \sum_i \frac{1}{\sigma_i^2} I_{y,i}^2 \end{bmatrix} \tag{19}$$

Thus the matrix inversion in (15) becomes the inversion of a diagonal matrix, the inverse of which is given by

$$[(P_k^{k-1})^{-1} + C_k^T \Sigma^{-1} C_k]^{-1} = \begin{bmatrix} (\alpha_x)_k & 0 \\ 0 & (\alpha_y)_k \end{bmatrix} \triangleq M_k \tag{20}$$

where

$$(\alpha_x)_k \triangleq \frac{(\sigma_{p,x}^2)_k^{k-1}}{1 + (\sigma_{p,x}^2)_k^{k-1} \sum_i \frac{1}{\sigma_i^2} I_{x,i}^2} \tag{21}$$

$$(\alpha_y)_k \triangleq \frac{(\sigma_{p,y}^2)_k^{k-1}}{1 + (\sigma_{p,y}^2)_k^{k-1} \sum_i \frac{1}{\sigma_i^2} I_{y,i}^2} \tag{22}$$

With this approximation for the matrix inverse in (15), the Kalman gain matrix can be written in an approximate form which does not require any matrix inversion, i.e.

$$K_k = M_k \begin{bmatrix} \frac{I_{x,1}}{\sigma_1^2} & \cdots & \frac{I_{x,n}}{\sigma_n^2} \\ \frac{I_{y,1}}{\sigma_1^2} & \cdots & \frac{I_{y,n}}{\sigma_n^2} \end{bmatrix} \tag{23}$$

In order to update the diagonal elements of the approximate covariance matrix P_k^{k-1} defined in equation (16), we note that the *"at observation k"* update equation of the Kalman filter algorithm for the covariance matrix can be written as

$$\begin{aligned} P_k^k &= P_k^{k-1} - K_k C_k P_k^{k-1} \\ &= \begin{bmatrix} (\sigma_{p,x}^2)_k^{k-1} & 0 \\ 0 & (\sigma_{p,y}^2)_k^{k-1} \end{bmatrix} \\ &\quad - M_k \begin{bmatrix} \sum_i \frac{1}{\sigma_i^2} I_{x,i}^2 & \sum_i \frac{1}{\sigma_i^2} I_{x,i} I_{y,i} \\ \sum_i \frac{1}{\sigma_i^2} I_{x,i} I_{y,i} & \sum_i \frac{1}{\sigma_i^2} I_{y,i}^2 \end{bmatrix} \begin{bmatrix} (\sigma_{p,x}^2)_k^{k-1} & 0 \\ 0 & (\sigma_{p,y}^2)_k^{k-1} \end{bmatrix} \end{aligned} \tag{24}$$

Again setting the off-diagonal elements to zero results in the two independent update equations for the variances of the velocity estimate in each of the cardinal directions, i.e.

$$(\sigma_{p,x}^2)_k^k = \frac{(\sigma_{p,x}^2)_k^{k-1}}{1 + (\sigma_{p,x}^2)_k^{k-1} \sum_i \frac{1}{\sigma_i^2} I_{x,i}^2} \tag{25}$$

$$(\sigma_{p,y}^2)_k^k = \frac{(\sigma_{p,y}^2)_k^{k-1}}{1 + (\sigma_{p,y}^2)_k^{k-1} \sum_i \frac{1}{\sigma_i^2} I_{y,i}^2} \tag{26}$$

This approximate form of the algorithm produces a model of the motion integration process which avoids the need to compute a matrix inverse in the calculation of the

Kalman gain matrix, and therefore implementable as a neural computation. We note here that the temporal and spatial derivatives of the intensity, $I_t(x,y,t)$, $I_x(x,y,t)$ and $I_y(x,y,t)$, appear in the algorithm in the form of either squared terms: I_t^2, I_x^2, I_y^2, or the products: I_xI_t, I_yI_t, I_xI_y. As observed in Heeger and Simoncelli (1993), there are no known cells in V1 with receptive fields which behave as products of derivatives, ie I_xI_t, I_yI_t, I_xI_y. However such products can be expressed in the form

$$\begin{aligned} I_xI_y &= \frac{1}{4}\left\{(I_x+I_y)^2-(I_x-I_y)^2\right\} \\ I_xI_t &= \frac{1}{4}\left\{(I_x+I_t)^2-(I_x-I_t)^2\right\} \\ I_yI_t &= \frac{1}{4}\left\{(I_y+I_t)^2-(I_y-I_t)^2\right\} \end{aligned} \tag{27}$$

and therefore all of the required functions of intensity derivatives in the model can be expressed as the outputs of squared linear filters of the image intensity (Heeger and Simoncelli, 1993). This both simplifies the implementation of the model in a neural computational form, and allows the possibility of mapping it physiologically onto the V1-MT neural circuitry.

Whilst the above approximations make the model amenable to neural implementation, we now further simplify the algorithm to make it suitable for implementation in a *distributed, recurrent* neural form. This will further enhance the possibility of mapping the model onto the recurrent MT-V1 neural circuitry. Instead of continuously updating the variances of the velocity estimate as just described, we now use the diagonal elements of the prior covariance P_0 in equations (21) and (22) to calculate the elements of the M_k matrix at every *"update at observation k"* step of the algorithm, ie

$$(\alpha_x)_k = \frac{(\sigma_{p,x}^2)_1^0}{1+(\sigma_{p,x}^2)_1^0 \sum_t \frac{I_{x,t}^2}{\sigma_t^2}} \tag{28}$$

$$(\alpha_y)_k = \frac{(\sigma^2_{p,y})^0_1}{1+(\sigma^2_{p,y})^0_1 \sum_i \frac{I^2_{y,i}}{\sigma^2_i}} \tag{29}$$

in which, from the initial *"update between observations"* step of the algorithm, $(\sigma^2_{p,x})^0_1 = (\sigma^2_{p,x})^0_0$ and $(\sigma^2_{p,y})^0_1 = (\sigma^2_{p,y})^0_0$, the diagonal elements of P_0. This of course affects the calculation of the Kalman gain matrix using equation (23). But it avoids the need to update the elements of the covariance matrix as described in equations (24)-(26).

In summary, the approximate form of the Kalman filter estimation algorithm which forms the basis of our neural model of the motion integration process is described as follows:

1. initialization:

$$v^0_0 = \overline{v_0}$$

$$P^0_0 = P_0 = diag\left[(\sigma^2_{p,x})^0_0, (\sigma^2_{p,y})^0_0\right]$$

2. update of the velocity estimate between observations

(a) before the first observation

$$v^0_1 = v^0_0$$

$$(\sigma^2_{p,x})^0_1 = (\sigma^2_{p,x})^0_0$$

$$(\sigma^2_{p,y})^0_1 = (\sigma^2_{p,y})^0_0$$

(b) between subsequent observations

$$v^k_{k+1} = v^k_k$$

3. update of the velocity estimate at observation *k*

$$v_k^k = v_k^{k-1} + K_k(h_k - C_k v_k^{k-1})$$

where

$$K_k = M_k \begin{bmatrix} \frac{I_{x,1}}{\sigma_1^2} & \cdots & \frac{I_{x,n}}{\sigma_n^2} \\ \frac{I_{y,1}}{\sigma_1^2} & \cdots & \frac{I_{y,n}}{\sigma_n^2} \end{bmatrix}, \quad M_k \triangleq \begin{bmatrix} (\alpha_x)_k & 0 \\ 0 & (\alpha_y)_k \end{bmatrix}$$

$$(\alpha_x)_k = \frac{(\sigma_{p,x}^2)_1^0}{1 + (\sigma_{p,x}^2)_1^0 \sum_i \frac{I_{x,i}^2}{\sigma_i^2}}$$

$$(\alpha_y)_k = \frac{(\sigma_{p,y}^2)_1^0}{1 + (\sigma_{p,y}^2)_1^0 \sum_i \frac{I_{y,i}^2}{\sigma_i^2}}$$

We can think however of the elements $(\alpha_x)_k^{k-1}$ and $(\alpha_y)_k^{k-1}$ of the M_k matrix as *normalised forms* of the prior variances $(\sigma_{p,x}^2)_0^0$ and $(\sigma_{p,y}^2)_0^0$ of the velocity estimate.

5 Simulation experiments with the motion integration model based on approximate Kalman filter estimation algorithm

We have repeated the simulation experiments described in Figure 1 with the motion integration model based on the approximate form of the Kalman filter estimation algorithm, as described in section 4 above, where for simplicity we set $\sigma_i = \sigma$ for all i, and $(\sigma_{p,x}^2)_0^0 = (\sigma_{p,y}^2)_0^0 = \sigma_p^2$. The results are shown in Figure 2.

As in Figure 1, the value of σ / σ_p increases from the top row to the bottom, and results in a corresponding increase in the peak offset bias in the estimated direction for both types of stimuli. There is again a small, but non-zero peak offset bias in the case of the diamond stimulus, which is less than 1° in all but the case of the highest value of σ / σ_p and the lowest contrast (lower left-hand plot). For the thin rhombus stimulus however, there is a substantial peak offset bias (>30°) for all values of

σ / σ_p (note that there is no change in the vertical scale in the plots in the first two rows as there is in Figure 1). This is in contrast to the simulations described in Figure 1 for the optimal Kalman filter based model, where such values of direction error are obtained only for the highest value of σ / σ_p (lower right-hand plot in Figure 1).

Comparing the model simulation results for the thin rhombus stimulus (right-hand column of plots) with Figures 5 and 6 of Wallace *et al* (2005), it can be seen that the peak direction error in the model simulation now corresponds more closely to that described by Wallace *et al* than is the case for the model based on the optimal Kalman filter. In particular, in the case of all values of σ / σ_p used, the peak direction error of the model has a variation with contrast of between 30° and 40°. This is in close agreement with the variation in peak direction error observed by Wallace *et al*, which was between 30° and 45° (Figures 6A and 6B of Wallace et al (2005))

It can also be seen from Figure 2 that the decay time constant of the direction error also increases with both the value of the σ / σ_p ratio and with the contrast, as in the case of Figure 1, and again closely mimics qualitatively the variation in decay time constant with contrast observed experimentally by Wallace *et al* (Figure 6C of Wallace *et al* (2005)). In this case, however, compared to the optimal filter based model (Figure 1), the time constants are generally larger. Interestingly, the model mimics a significant characteristic of the Wallace *et al* data, in that the time constants stay approximately constant for all but the lowest levels of contrast where they increase substantially (~50% increase – see Figure 6C of Wallace *et al* (2005)). The model results also display this feature of the experimental data for all but the lowest value of the σ / σ_p ratio, as seen in the three lower right-hand plots of Figure 2. Overall, in the case of the approximate filter based version of the model there is a higher level of correspondence between the simulation results and the experimental results of Wallace *et al* (2005), than in the case of the optimal filter based version of the model.

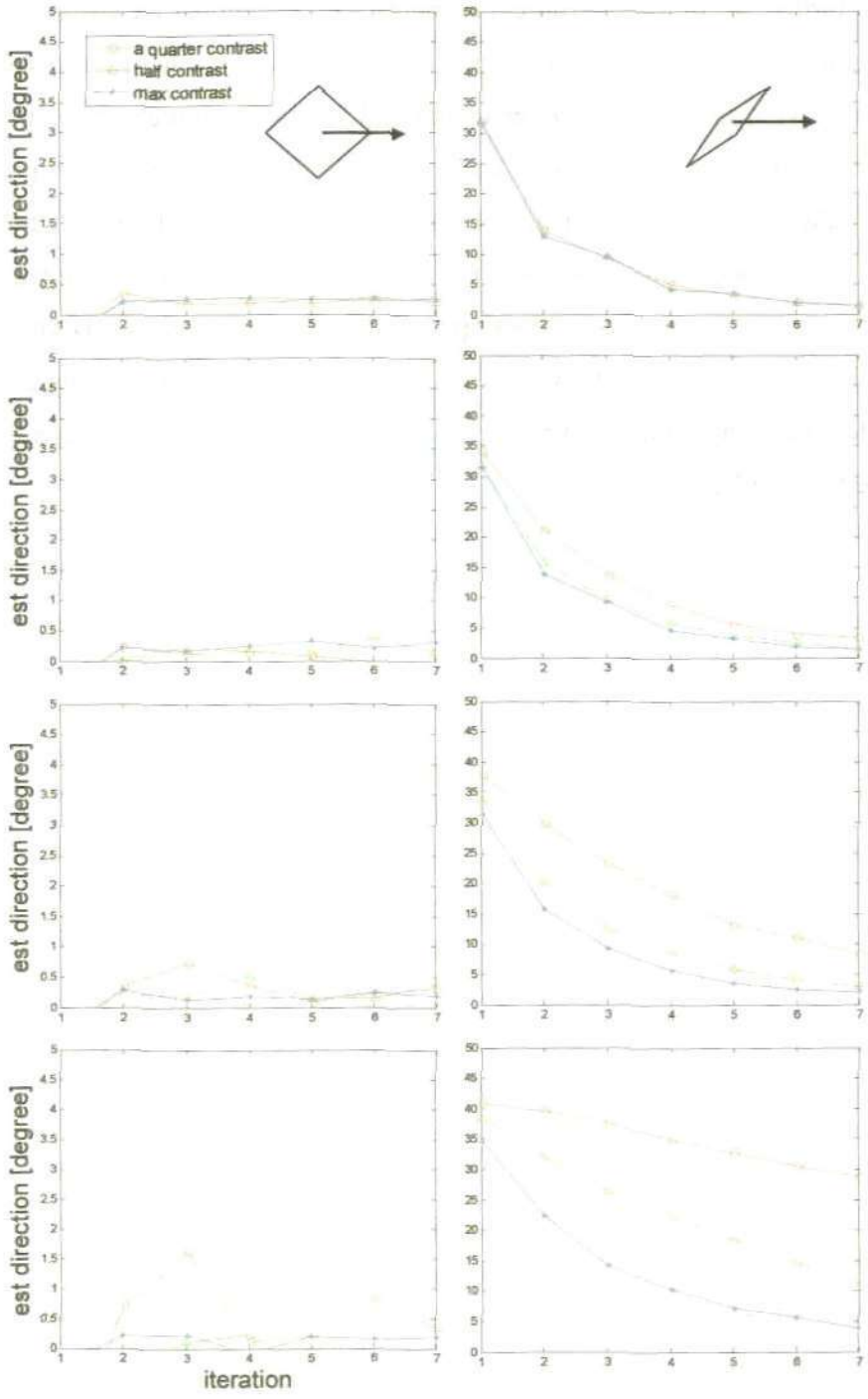


Figure 2. Results from the computer simulations of the motion integration model as in Figure 1, but for the model based on the approximate Kalman filter algorithm.

We carried out a further set of simulation experiments with the model based on the approximate Kalman filter using the thin rhombus with varying length, in order to see the effect of the relative influence of 1-D versus the 2-D local motion measurements. In principle, a longer stimulus will contain a larger proportion of 1-D cues, compared to 2-D cues, and thus should result in an increase in the peak offset bias in the direction estimate and a longer time of decay of this bias to zero. This prediction was confirmed experimentally in a study for smooth pursuit initiation in monkeys (Born *et al*, 2006). They showed that the directional error is more pronounced and lasts longer for the longer tilted bars. Our simulation results are presented in Figure 3, using only one level of noise for simplicity.

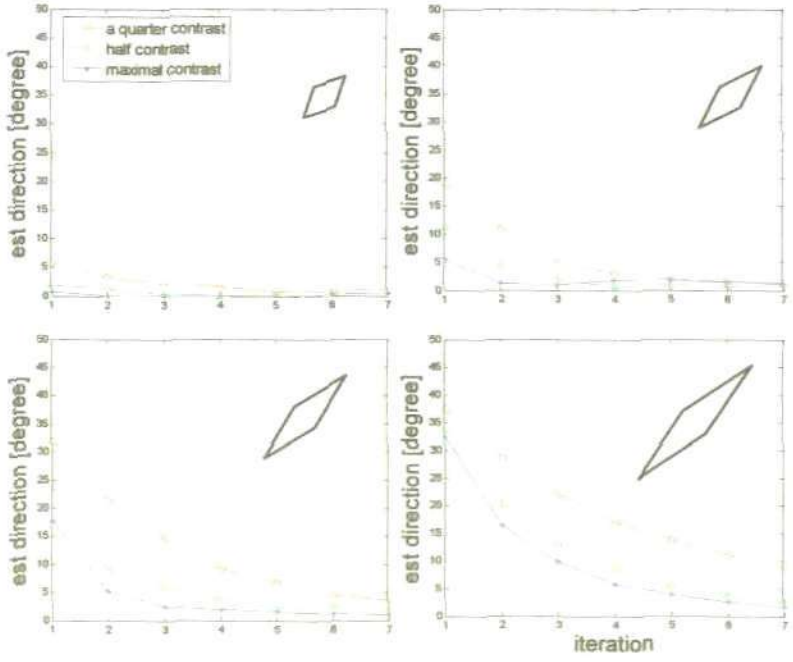


Figure 3. Results from the computer simulations of the motion integration model based on the approximate Kalman filter algorithm, for four different sizes of the thin rhombus and three values of contrast. The sizes are: 10, 20, 30 and 50 pixels length for the main diagonal and a constant size of 3 pixels for the short diagonal. The value of σ / σ_p ratio used was equal to 5.

When the length of the long diagonal of the 'thin' rhombus increases, both the peak directional error and the decay time increase, approaching the value of the vector

average direction for the longest stimulus used. These results compare well in a qualitative sense with those of Born *et al* (2006) (see their Figure 4).

6 A distributed recurrent model of the motion estimation process based on the approximate Kalman filter estimation algorithm

We now show that the model of the motion integration process based on the approximate version of the Kalman filter algorithm described above, leads to a distributed model of the motion integration process. In this form of the model, local observations of the spatial and temporal derivatives of the image intensity are made at a distributed set of spatial locations in the visual space, and are used together with the current global velocity estimate to form a corresponding set of local updates to the current estimate. The local update signals are then summed, and used to calculate a new global estimate of the velocity. The new global velocity estimate is then broadcast to each spatial location where it is again used to calculate a local update signal to the global estimate. This leads naturally to a recurrent, distributed form of the motion integration process, which as we show below can be mapped in a coarse sense onto the recurrent neural circuitry which exists between V1 and MT.

To describe this form of the model, we first define a local *"update"* signal for each spatially local observation window as

$$(e_k^{k-1})_i \triangleq \begin{bmatrix} \frac{I_{x,i}I_{t,i}}{\sigma_i^2} \\ \frac{I_{y,i}I_{t,i}}{\sigma_i^2} \end{bmatrix} - \begin{bmatrix} \frac{I_{x,i}^2}{\sigma_i^2} & \frac{I_{x,i}I_{y,i}}{\sigma_i^2} \\ \frac{I_{x,i}I_{y,i}}{\sigma_i^2} & \frac{I_{y,i}^2}{\sigma_i^2} \end{bmatrix} v_k^{k-1} \tag{30}$$

It then follows that the *"at observation k"* update equation for the velocity estimate in the approximate Kalman filter algorithm

$$v_k^k = v_k^{k-1} + K_k(h_k - C_k v_k^{k-1}) \tag{31}$$

where

$$K_k = M_k \begin{bmatrix} \frac{I_{x,1}}{\sigma_1^2} & \cdots & \frac{I_{x,n}}{\sigma_n^2} \\ \frac{I_{y,1}}{\sigma_1^2} & \cdots & \frac{I_{y,n}}{\sigma_n^2} \end{bmatrix}, \quad M_k \triangleq \begin{bmatrix} (\alpha_x)_k & 0 \\ 0 & (\alpha_y)_k \end{bmatrix}$$

can be written in terms of the local update signals as

$$v_k^k = v_k^{k-1} + M_k \sum_i (e_k^{k-1})_i \tag{32}$$

Thus the global update of the velocity estimate is expressed as a summation of local estimate update signals.

This results in a distributed version of the motion integration model, wherein: (i) for each spatially distributed observation window a local update signal $(e_k^{k-1})_i$ is computed based on the squared local temporal and spatial derivatives $I^2_{x,i}, I^2_{y,i}, I^2_{t,i}$, the variance of the local measurement noise σ_i^2, and the current velocity estimate v_k^{k-1} (equation (30)) ; (ii) the local update signals are summed and used to create a new velocity estimate v_k^k (equation (31)).

This distributed, recurrent form of the motion integration model is illustrated in Figure 4, in which we also suggest a mapping of the model onto the V1-MT recurrent circuitry. This mapping proposes that feedforward information is provided to MT from V1 neurons with spatially distributed receptive fields, which consists of : (i) the local estimate update signals $\left(e_k^{k-1}\right)_i$; (ii) the local measurements of the spatial derivatives of image intensity $I^2_{x,i}, I^2_{y,i}$, modulated by the local measurement noise variance σ_i^2, ie $\frac{I^2_{x,i}}{\sigma_i^2}, \frac{I^2_{y,i}}{\sigma_i^2}$. The latter information is necessary for the computation

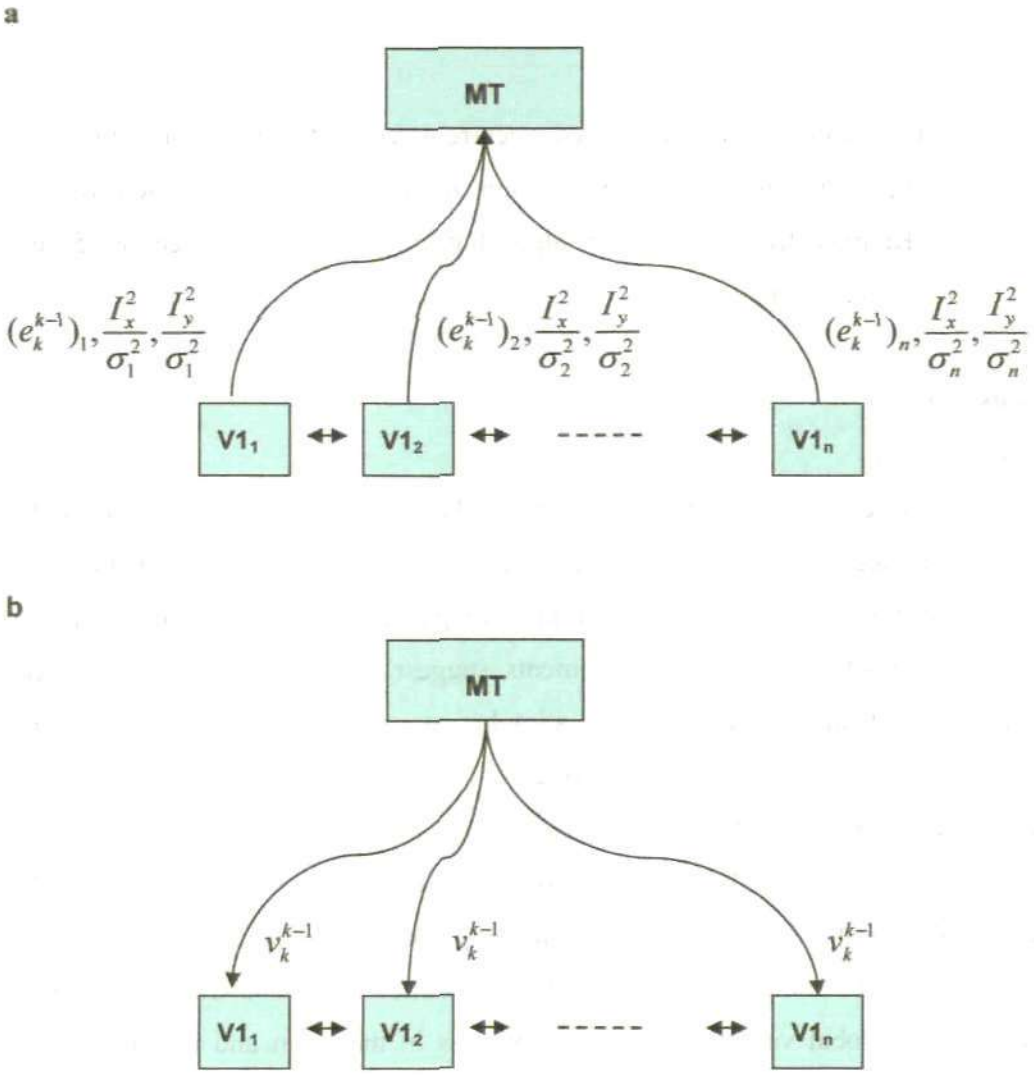


Figure 4. The distributed, recurrent form of the motion integration model based on the approximate Kalman filter estimation algorithm. a. Feedforward of the local update signals and modulated local spatial derivatives of the image intensity from neurons in V1 with receptive fields corresponding to the local spatial windows. b. Feedback of the global estimate of velocity of the image.

of the elements of the matrix M_k, ie for the "normalisation" of the MT-derived prior variances $(\sigma_{p,x}^2)_0^0$ and $(\sigma_{p,y}^2)_0^0$ according to equations (21) and (22). The recipient neurons in MT then compute a new velocity estimate by summing the local update signals and applying the normalisation M_k. These neurons then project the new velocity estimate back to the local V1 neurons to allow them to calculate new local

update signals based on a new set of observations of the local spatial and temporal derivatives $I^2_{x,j}, I^2_{y,j}, I^2_{t,j}$.

Clearly, the behaviour of this distributed, recurrent form of the motion integration model, will be identical to that the motion integration model based on the approximate Kalman filter estimation algorithm, as described in section 5 and illustrated in Figures 2 and 3.

7 Discussion

7.1 Model results

Our aim was to develop a theoretical, neurally plausible computer simulation model of the motion integration process which is intrinsic to the control of eye movements in the smooth pursuit of moving target stimuli. In particular psychophysical data for initiation of smooth pursuit eye movements suggest that the integration of local measurements of the velocity of the stimulus starts before the closing of the oculomotor control loop and is correcting for inaccurate, ambiguous local motion signals prior to the onset of any compensatory eye movement (Wallace *et al,* 2005; Born *et al*, 2006). This data also demonstrates a dynamical behaviour for the correction of the off-axis directional error in the global motion signal used to initiate eye movement, presumably reflecting the dynamics of an underlying motion integration and global velocity calculation process in the brain and the subsequent correcting eye movements. To develop our model, we used a recursive Bayesian estimation of the global stimulus velocity, based on the local 1-D and 2-D motion measurements. We implemented the estimation process as a Kalman filter algorithm for optimally determining the posterior distribution of the global velocity estimate. This algorithm makes some basic assumptions about the nature of the motion observation process, notably that it is linear and subject to additive Gaussian distributed measurement noise. We also assume that, as in the psychophysical experiments, the stimulus is moving with constant velocity and changes in image intensity at any location in visual space are due only to motion of the stimuli.

Given these assumptions the simulation results from the Kalman filter based motion integration model are shown to closely match the experimental data. In particular we show that such a model has a dynamic behaviour, the result of the recurrent Bayesian estimation process, which possesses qualitatively the same characteristics as the experimentally observed dynamics of the motion integration process during the intitial stages of smooth eye pursuit (Wallace *et al, 2005*; Born *et al, 2006*). Specifically, we show that variations in the stimulus contrast in our model simulations lead to changes in the peak directional error and the time constant for the decay of this error to zero which are qualitatively consistent with the changes in the dynamics of the directional error induced by similar contrast variations in the experimental situation (Wallace *et al,* 2005). Similar consistent results are obtained from the model in response to changes in stimulus length and the corresponding ratio of 1-D and 2-D local motion measurements (Born *et al*, 2006). These results strongly suggest that the brain may be using some form of Bayesian estimation process to correct for the presence of ambiguous 1-D local motion cues in the calculation of a veridical global stimulus velocity for smooth eye pursuit.

Furthermore, we aimed at demonstrating the neural plausibility of our model, both in terms of its computation using neurally plausible functions, ie avoiding the computation of matrix inversions as in the optimal Kalman filter based model, and in respect of its physiological plausibility in relation to the presumed underlying neural substrate of the V1-MT recurrent circuitry. This first required the development of an approximate version of the Kalman filter which avoided the use of matrix inversion. The simulation results from the motion integration model based on the approximate Kalman filter algorithm presented in our study, show an even closer correspondence with the experimental data, suggesting that the brain may well be adopting a neural network based, sub-optimal Bayes approach to the estimation of veridical stimulus velocity in the integration of local motion cues. We were also able to show that our neurally plausible model was capable, albeit in a relatively coarse way, of being mapped onto the recurrent neural circuitry connecting cortical areas V1 and MT. These areas have been specifically identified by a number of researchers as the physiological substrate for the processing of information for smooth eye pursuit and

for the integration of local motion cues (Komatsu and Wurtz, 1988, 1989; Groh *et al*, 1997; Simoncelli, 2003; Simoncelli and Heeger, 1998; Born *et al*, 2000, Pack and Born, 2001; Pack *et al*, 2003; Pack *et al*, 2004). We showed that the approximate Kalman filter algorithm can be formulated as a distributed sub-optimal recurrent estimation process which at least coarsely corresponds to the distributed architecture of the V1-MT circuitry. In particular, the convergent nature of the connections from the small subset of neurons in layer 4B of V1, less than 0.5% or ~40-60 cells in an orientation hypercolumn (Shipp and Zeki, 1989) which project to MT is suggested by the approximately ten times larger receptive field size of MT neurons compared to V1 neurons (Livingstone et al, 2001). The feedback projections from MT to layer 4B of V1 are widely divergent and have terminal connections which suggest that an MT cell which receives input from V1 layer 4B neurons has the opportunity to influence at least all of those V1 neurons which project to it (Rockland and Knutson, 2000).

It is important to emphasise here that, at this stage of its development, we make no claims for the biological plausibility of our model; that is, how it relates to the specific properties of V1 and MT nerve cells and their connections, for example the transmission delays in the V1-MT recurrent circuitry. In fact we have explicitly avoided any such attempt, which would require a far deeper analysis of how recurrent algorithms of the kind embodied in our model might be implemented by specific biological processes. In this paper we only seek to show that the distributed, recurrent form of the model's algorithm makes possible a coarse mapping of the recursive algorithm onto the convergent/divergent nature of the V1-MT recurrent circuitry, and thereby has possible implications for the role of feedback connections from MT to V1, an important aspect of the physiology of the biological system.

7.2 The origin of motion integration dynamics in the initiation of smooth eye pursuit

For some time it has been unclear as to the origin of the temporal dynamics which have been observed in the initiation of smooth eye pursuit in both humans and monkeys (Masson and Stone, 2002, Wallace *et al*, 2005) and the associated dynamic response properties of monkey MT neurons (Pack and Born, 2001; Born *et al*, 2002). One recent proposal has been that it is the consequence of a delayed response of end-

stopped cells resulting in a change in the "weighting" applied to the outputs of V1 direction-selective neurons as end-stopping eventually suppresses contour-related motion signals and emphasizes those from terminators (Pack *et al*, 2003; Born *et al*, 2006). However this suggestion has not resulted in the development of a specific model which can be tested against the behaviourally and physiologically observed dynamics.

Another suggestion which has been made to explain the dynamics of motion integration, in perceptual pattern coherence, is based on the computation of a delayed intersection-of-constraints (IOC) based 2-D signal from local 1D motion signals (Adelson and Movshon, 1982; Movshon *et al*, 1985). In the model of Simoncelli and Heeger (1998) such an IOC computation is achieved by an appropriate feed-forward weighting of local motion signals. However, this model lacks any temporal dynamics and would need to be substantially extended in order to account for the observed dynamics of motion integration in the initiation of smooth eye pursuit.

It has also been proposed that the independent computation of Fourier and non-Fourier motion signals followed by the computation of their vector average will lead to a dynamic response in motion integration owing to the delay in computing non-Fourier motion signals relative to Fourier motion signals (Löffler and Orbach 1999; Wilson et al. 1992) The time-delay has been found to be approximately 60 ms (Yo and Wilson, 1992), which is consistent with the observed dynamics of motion integration. Recent work by Barthélemy *et al* (2008) also shows that the Fourier and non-Fourier motion signals display different contrast dynamics, which may underlie the variation of the dynamics of motion integration with contrast in such a model. Loffler and Orbach (1999) developed a model based on this idea to explain the psychophysical results on the perception of plaid motion, which they tested on stimuli corresponding to stationary terminators, plaids, and moving terminators. Although they did not simulate the temporal dynamics of their model explicitly, they make a prediction for the dynamics of direction perception of briefly presented terminator stimuli, which suggests an initial directional bias approximately equal to the output of the Fourier pathway alone. As far as we are aware the model has not been used to

explain the dynamics of motion integration in smooth pursuit initiation, and it is not clear whether or not the results of Wallace et al (2005) and Born et al (2006), in particular the variation in magnitude and decay time of the directional bias with stimulus shape and length could be readily explained by this model, since at least in such variations would appear to have little or no effect on the time delay of the non-Fourier pathway. Furthermore, as pointed out in Smith *et al* (2005), there is no clear evidence of a separate cortical pathway for the computation of non-Fourier "pattern" motion cues, such as in areas V2 or V3, which do not apparently make an important separate contribution to the behaviour of pattern direction selective neurons in MT.

A number of models of motion integration have been based on the idea of the spatial propagation of 2-D motion signals from line terminators and their suppression of the ambiguous 1-D motion signals. The model of Lidén and Pack (1999) employs recurrent networks to spatially propagate motion signals across model MT cells, the propagation dynamics providing a good qualitative fit to the temporal dynamics of motion integration as observed in MT (Pack and Born, 2001). Similarly, the motion integration and segmentation model of Grossberg *et al* (2001) uses feedback connections between model MT and MST cells to select winning directions and suppress losing directions on a top-down matching process, resulting therefore in a temporal dynamics. It also uses lateral connectivity in MT cells to amplify 2-D feature tracking signals and suppress 1-D ambiguous direction signals, in a manner similar to that of Lidén and Pack (1999). Neither of these two models has been used to simulate the smooth pursuit initiation data of Wallace *et al* (2005), but in principle could provide an explanation of the dynamic nature of this data. Indeed, Lidén and Pack (1999) have used their model to simulate the perception of a horizontal line translating at 45° relative to its orientation. The model initially signals horizontal motion, and then gradually recovers the true line motion direction by propagating unambiguous motion signals generated by the terminators along the contour of the line. The model dynamics is therefore based on the dynamics of the spatial propagation of the motion signals, which the authors relate anatomically to the lateral connectivity between motion selective cells in MT. Some evidence exists for laterally extending fibres in MT (Van Essen, Maunsell & Bixby 1981), but further detailed physiological

examination of these connections and their propagation delays would be necessary to reveal whether or not the dynamics conveyed by this connectivity is consistent with the psychophysically (Wallace *et al*, 2005) or physiologically observed (Born *et al*, 2006) dynamics of the perceptual bias in the initiation of smooth pursuit.

The model of motion disambiguation proposed by Bayerl and Neumann (2004) also suggests that the dynamics of motion integration results from the spatial propagation of signals in MT, and whilst it has not been used to explain specifically the data on smooth pursuit initiation, such propagation dynamics would similarly predict the observed dynamics of smooth pursuit initiation. It would further predict that the spatial distance of MT cells from unambiguous 2-D motion features will directly influence the time for the decay of the perceptual directional bias to zero. An additional feature of this model compared to that of Lidén and Pack (1999) is that it uses feedback from MT to V1 to attentionally gate by excitatory modulation the unambiguous V1 motion signals, and thus takes some account of the feedback connections of the MT-V1 circuitry.

In the Bayesian estimation framework, Koechelin *et al* (1999) describe a model of motion integration based on the V1-MT circuitry that also employs mechanisms of recurrent lateral interactions. Their model proposes a multiplicative combination of feedforward input and the result of lateral integration, which leads to the proposal that the model represents a neural implemention of Bayesian motion estimation.

Also using a Bayesian approach, a recursive extension of the Bayesian model of motion estimation of Weiss *et al* (2002) has recently been proposed (Montagnini et al, 2007). They used human experimental data on smooth pursuit eye movements in response to dot and line stimuli to derive, respectively, the variances of two likelihood functions, one for the 1-D cues and one for the 2-D cues. These likelihood functions are used together with the same prior as in Weiss et al (2002) in a recursive version of the Weiss *et al* model, to produce a discrete evolution of velocity estimates. They show that the temporal evolution of the velocity estimates expressed in terms of tracking error, coarsely matches their experimental data for mean eye velocity and

different target stimulus speeds. However they have not attempted to explain the observations of Wallace et al (2005) on smooth pursuit initiation.

In contrast to the above models, the model we propose uses a neurally plausible recursive estimation process based on an approximate form of the Kalman filter, and has been specifically addressed to the issue of motion integration dynamics in the initiation of smooth eye pursuit. As we have shown, our model demonstrates the main characteristics of experimentally observed dynamics, and the dependence of the model dynamics on contrast qualitatively replicates the experimentally observed contrast variation in the motion integration dynamics (Wallace *et al,* 2005; Born *et al,* 2006). We were also able to show similar qualitative agreement between the model simulations and the experimental data of Born *et al* (2006), for the dependence of the motion integration dynamics on stimulus length.

An important feature which distinguishes our model from many others described above is that it does not depend on the concept of lateral propagation of unambiguous motion signals along line contours in order to achieve a veridical estimation of motion direction in the presence of ambiguous direction cues. Rather, we suggest that true motion perception is achieved through a recursive estimation process which produces successively improved velocity estimates by eliminating the initial estimation errors introduced by the ambiguous motion signals of local detectors. We suggest that this can be achieved by the divergent feedback connections which are known to exist from MT motion selective cells back to V1 cells, and which connect to a retinotopic area of V1 which corresponds to the receptive field size of the MT cell (Shipp and Zeki, 1989). At the present time, more is known about these connections than is about the lateral connections in MT, in particular the fast transmission times of V1-MT connections (Raiguel et al, 1989; Schmolesky et al, 1998; Hupe et al 2001). It is very clear however that much further work will be needed, in relating the models to their proposed physiological and anatomical basis in the visual system, before it becomes clear as to whether the origin of motion integration dynamics lies predominantly in: (i) the delayed activation of end-stopped cells (Pack *et al*, 2003; Born *et al*, 2006); (ii) separate Fourier and non-Fourier motion pathways (Löffler and Orbach 1999); (iii)

lateral connectivity in MT (Lidén and Pack, 1999; Bayerl and Neumann, 2004; Koechelin *et al*, 1999)); (iv) feedback connectivity between MT and MST (Grossberg *et al*, 2001); or (v) feedback connectivity between V1 and MT.

However our model does provide support for the notion that motion integration in the brain might be based on a Bayesian estimation process, as has been suggested by many psychophysical studies, and suggests that the observed motion integration dynamics, and their dependence on stimulus contrast and length, may result from the recursive nature of this motion estimation process. The model also reflects the recurrent MT-V1 circuitry, the pooling of information from local V1 motion detectors, the convergence of the pooled local motion signals in projections to MT, and the divergent feedback of MT velocity signals to V1 cells. However far more work needs to be done before we could show that the recurrent MT-V1 circuitry is biologically capable of supporting the kind of recursive Bayesian estimation algorithm inherent in our model.

We are also aware that our model, along with all the models we have reviewed here, is limited by its ability to explain the dynamics of motion integration in the initiation of smooth eye pursuit only in relation to intensity-based motion stimuli, and not moving targets defined in other ways, eg cyclopean targets defined by means of random-dot stereograms. We note however that although it is likely that different neural pathways are involved in luminance motion and stereomotion perception, the roles played by the pooling of local motion detectors and moving stereoscopic line terminators appear similar in both cases (Donnelly *et al*, 1997; Patterson *et al*, 1998), and therefore similar motion estimation processes involving Bayesian inference may be involved.

7.3 The effect of the free parameter σ / σ_p

We have varied the free parameter in our model, σ / σ_p, the ratio of standard deviations of the probability density functions of the measurement noise and the prior velocity estimate, as shown in Table 1, in order to demonstrate the dependence of the model behaviour on this parameter. As has been pointed out by Stocker and

Simoncelli (2006), Bayesian models of visual motion perception are difficult to validate quantitatively owing to the fact that it is hard to attribute precise values to these variables. Assuming that a reasonable estimate of the prior velocity estimate can be made on theoretical grounds, or based on the statistics of natural visual stimuli, uncertainty in the value of the measurement noise variance remains. It is interesting however to note that this variance always appears in our model as a divisive modulation of the squared image intensity derivatives, that is, as $\frac{I^2_{x,i}}{\sigma_i^2}, \frac{I^2_{y,i}}{\sigma_i^2}$. Thus we can think of the measurement noise variance in neural terms as a signal which increases or decreases the magnitude of the output of the squared linear filter representation of the neurons which code for $I^2_{x,i}, I^2_{y,i}$. The obvious suggestion for a neural implementation of our model is that the influence of the measurement noise variances in the Bayesian formulation could be identified with the divisive normalisation role of the intracortical connections between orientation hypercolumns (Carandini and Heeger, 1997). Thus neurons reporting the directional derivatives (or their orientated versions) of the image intensity would be suppressed by the activity of neurons in the surrounding spatial region which were responding to stimuli with a different orientation/direction, indicating an increased uncertainty in the local motion measurement consistent with a decrease in signal to noise ratio for this measurement. If we adopt this approach, we have no need of a precise measurement for the measurement noise variance, which in any case is highly unlikely to be represented in any V1 neuronal responses, but instead we mimic the effect of the divisive role of this variance in the model by the effect of divisive normalisation via intracortical lateral connections.

7.4 Predictions of the model

The recursive nature of the calculation of the velocity estimate in our motion integration model suggests that blanking of the stimulus for short periods of time will not result in the reintroduction of an offset directional error in the eye movement when the stimulus reappears at the end of the blanking period. This follows from the fact that the update of the velocity estimate, as expressed in equation (31), depends on the error signal $(h_k - C_k v_k^{k-1})$. Owing to the definitions of h_k and C_k (following

equation (4)) in the absence of any stimulus during the blanking period, both h_k and C_k will be zero. Thus no updates of the velocity estimate will take place during the blanking period, and at the end of the period the original velocity estimate will be used in further updates. Thus the estimation procedure will continue as if it had not been interrupted. This "prediction" of the model has already been confirmed by the experiments of Masson and Stone (2002), in which the motion of a tilted elongated diamond stimulus (thin rhombus) was transiently blanked for a period of 90 ms during steady state pursuit. They observed a small decrease in the eye velocity in the veridical direction during the blanking period and a transient increase in this velocity immediately after the blanking period ended, but no post-blanking reappearance of an offset direction error. The small decrease in the eye motion in the veridical direction could be explained by a slow decay of the velocity estimate in our model in the absence of any updating, which would maintain the directional component of the estimate but decrease its magnitude. The update of the estimate post-blanking would then rapidly correct for any magnitude error without introducing any directional error.

A further, physiological prediction of the model, in terms of our proposed coarse mapping of the recursive estimation process onto the V1-MT circuitry, is that as the estimated velocity converges to the true velocity, the local update signals from the V1 neurons which project these signals to MT will reduce in magnitude to nearly zero. This, we admit, would probably be difficult to test. Firstly only a small number of V1 neurons project to MT, from layer 4B (<0.5%) and layer 6 (<0.05%) in monkey V1 (Shipp and Zeki, 1989), which means that it would be difficult to find and record from such neurons. Also, many of the cells will be projecting directional information about the stimulus from V1 to MT (Rust *et al*, 2006; Livingstone *et al*, 2001) as suggested in the model, so a very small number of V1 cells may be projecting local estimation update signals, according to the model.

Acknowledgement

The work reported here was partly supported by funding under the Sixth Research Framework Programme of the European Union under the grant no. 15879 (FACETS).

References

Adelson EH, Movshon, JA (1982) Phenomenal coherence of moving visual patterns. Nature 300: 523-525

Barthelemy FV, Perrinet LU, Castet E, Masson GS (2008) Dynamics of distributed 1D and 2D motion representations for short latency ocular following. Vision Res 48: 501-522

Bayerl P, Neumann H (2004) Disambiguating visual motion through contextual feedback modulation. Neural Comp 16: 2041–2066

Born RT, Pack CC, Zhao R (2002) Integration of motion cues for the initiation of smooth pursuit eye movements. Prog Brain Res 140:225–237.

Born RT, Pack CC, Ponce CR, Si Yi (2006) Temporal Evolution of 2-Dimensional Direction Signals Used to Guide Eye Movements. J Neurophysiol 95:284-300

Born RT, Groh JM, Zhao R, Lukasewycz SJ (2000) Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements. Neuron 26: 725–734

Bowns L. (1996) Evidence for a feature tracking explanation of why type II plaids move in the vector sum directions at short durations. Vision Res 36: 3685–3694

Burke D, Wenderoth P (1993) The effect of interactions between one dimensional component gratings on two dimensional motion perception. Vision Res 33: 343–350

Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci 17:8621-44.

Castet E, Lorenceau J, Shiffrar M, Bonnet C (1993) Perceived speed of moving lines depends on orientation, length, speed and luminance Vision Res. 33, 1921-1936

Chey, J., Grossberg, S., & Mingolla, E. (1997). Neural dynamics of motion grouping: from aperture ambiguity to object speed and direction. J Opt Soc Amer, 14: 2570–2594

Donnelly M, Bowd C, Patterson R (1997) Direction discrimination of cyclopean (stereoscopic) and luminance motion. Vision Res 37(15):2041-6

Fennema CL, Thompson W (1979) Velocity determination in scenes containing several moving objects. Computer Graphics and Image Processing, 9: 301-315

Groh JH, Born RT, Newsome WT (1997) How is a sensory map read out? Effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. J Neurosci 17:4312–4330

Heeger DJ, Simoncelli EP (1993) Model of visual motion sensing. In: Harris L, Jenkin M (eds) Spatial Vision in Humans and Robots, Cambridge University Press, pp 367-392

Hupé JM, James AC, Girard P, Lomber SG, Payne BR, Bullier J (2001) Feedback connections act on the early part of the responses in monkey visual cortex. J Neurophysiol 85: 134–145

Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82: 35-45

Koechlin E, Anton JL, Burnod Y (1999) Bayesian inference in populations of cortical neurons: a model of motion integration and segmentation in area MT. Biol. Cybern 80: 25–44

Komatsu H, Wurtz RH (1988) Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. J Neurophysiol 60: 580–603

Komatsu H, Wurtz RH (1989) Modulation of pursuit eye movements by stimulation of cortical areas MT and MST. J Neurophysiol 62: 31–47

Lidén L, Pack C (1999) The role of terminators and occlusion cues in motion integration and segmentation: a neural network model. Vision Res 39: 3301–3320

Linder A, Ilg UW (2000) Initiation of smooth-pursuit eye movements to first-order and second-order motion stimuli. Exp Brain Res 133: 450–456

Lisberger SG, Westbrook LE (1985) Properties of visual inputs that initiate horizontal smooth pursuit eye movements in monkeys. J. Neurosci. 5: 1662-1673

Livingstone MS, Pack CC, Born RT (2001) Two-dimensional substructure of MT receptive fields. Neuron 30: 781–793

Löffler G, Orbach HS (1999) Computing feature motion without feature detectors: a model for terminator motion without end-stopped cells. Vision Res 39: 859–871, 1999.

Lorenceau J, Shiffrar M, Wells N, Castet E (1993) Different motion sensitive units are involved in recovering the direction of moving lines. Vision Res 33: 1207–1217

Marr D, Ullman S (1981) Directional selectivity and its use in early visual processing. Proc R Soc Lond B Biol Sci 211: 151–180

Masson GS, Castet E (2002) Parallel motion processing for the initiation of short-latency ocular following in humans. J Neurosci 22: 5149–5163

Masson GS, Rybarczyk Y, Castet E, Mestre DR (2000) Temporal dynamics of motion integration for the initiation of tracking eye movements at ultra-short latencies. Vis Neurosci 17: 753–767

Masson GS, Stone LS (2002) From following edges to pursuing objects. J Neurophysiol 88: 2869–2873

Maunsell JH, Van Essen DC (1983) The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. J Neurosci 3: 2563–2586

Montagnini A, Mamassian P, Perrinet L, Castet E, Masson GS (2007) Bayesian modeling of dynamic motion integration. Journal of Physiology - Paris 101: 64-77

Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1985). The analysis of moving visual patterns. In Pattern Recognition Mechanisms, ed. C. Chagas, R. Gattass, C.Gross (Pontificiae Academiae Scientiarum Scripta Varia **54**: 117-151). Rome: Vatican Press. (Reprinted in Experimental Brain Research, Supplementum **11**: 117-151, 1986)

Pack CC, Born RT (2001) Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. Nature 409: 1040-1042

Pack CC, Livingstone MS, Duffy KR, Born RT (2003) End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. Neuron 39: 671–680

Pack CC, Gartland AJ, Born RT (2004) Integration of contour and terminator signals in visual area MT. J Neurosci 24: 3268–3280

Patterson R, Bowd C, Donnelly M (1998) The cyclopean (stereoscopic) barber pole illusion. Vision Res 38:2119-25

Polat U, Mizobe K, Pettet MW, Kasamatsu T, Norcia AM (1998) Collinear stimuli regulate visual responses depending on cell's contrast threshold. Nature 391: 580–584

Raigel SE, Lagae L, Gulya SB, Orban GA (1989) Response latencies of visual cells in macaque areas V1, V2 and V5. Brain Res 493: 155–159

Rao RPN (2004) Bayesian computation in recurrent neural circuits. Neural Comp 16: 1–38

Rockland A, Knutson T (2000) Feedback connections from area MT of the squirrel monkey to areas V1 and V2. J Comp Neurol. 425:345-368

Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. Nat Neurosci 9:1421-1431

Schmolesky MT, Youngchang Wang, Hanes DP, Thompson KG, Leutgeb S, Schall JD (1998) Signal timing across the macaque visual system. J Neurophysiol 79: 3272–3278

Sceniak MP, Ringach DL, Hawken MJ, Shapley R (1999) Contrast's effect on spatial summation by macaque V1 neurons. Nat Neurosci 2: 733–739

Sceniak MP, Hawken MJ, and Shapley R (2001) Visual spatial characterization of macaque V1 neurons. J Neurophysiol 85: 1873–1887

Shibata T, Tabata H, Schaal S, Kawato M (2005) A model of smooth pursuit based on learning of the target dynamics using only retinal signals. Neural Networks 18: 213-225.

Shipp S, Zeki S (1989) The organization of connections between areas V5 and V1 in macaque monkey visual cortex. Eur J Neurosci 1: 309–332

Simoncelli E, Adelson E, Heeger D (1991) Probability distributions of optical flow. In: Proc. IEEE Conf. Comput. Vision Pattern Recog, IEEE, Washington DC, pp 310-315

Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. Vision Res 38: 743-761

Simoncelli EP (2003) Local analysis of visual motion. In: Chalupa LM, Werner JS, The Visual Neurosciences, MIT Press, pp 1616-1623

Smith MA, Majaj NJ, Movshon JA (2005). Dynamics of motion signaling by neurons in macaque area MT. Nature Neuroscience **8**: 220-228.

Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci 9: 578-585

Stone L, Watson A, Mulligan J (1990) Effect of contrast on the perceived direction of a moving plaid. Vision Res 30: 1049-1067

Stone L, Thompson P (1990) Human speed perception is contrast dependent. Vision Res 32: 1535-1549

Thompson P (1982) Perceived rate of movement depends on contrast. Vision Res 22: 377-380

Wallace JM, Stone LS, Masson GS (2005) Object motion computation for the initiation of smooth pursuit eye movements in humans. J Neurophysiol 93:2279-2293

Wilson HR, Ferrera VP, Yo C (1992) A psychophysically motivated model for two-dimensional motion perception. Visual Neurosci 9: 79–97

Wallach H (1935) Über visuell wahrgenommene Bewegungsrichtung. Psychologische Forschung 20: 325-380

Weiss E, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. Nat Neurosci 5:598-604

Wuerger S, Shapley R, and Rubin N (1996). On the visually perceived direction of motion by Hans Wallach: 60 years later. Perception, 25: 1317-1367

Yo C Wilson H (1992). Moving two-dimensional patterns can capture the perceived directions of lower or higher spatial frequency gratings. Vision Res, 32: 1263–1270

A model of plaid motion perception based on recursive Bayesian integration of the 1-D and 2-D motions of plaid features.

Kameliya D. Dimova & Michael J. Denham

Centre for Theoretical and Computational Neuroscience, University of Plymouth, Plymouth, PL4 8AA, UK

Abstract. We describe a theoretical and computational model of the perception of plaid pattern motion which fully accounts for the majority of cases in which misperception of the direction of motion of Type II plaids has been observed [Yo, C. & Wilson, H. (1992) Perceived direction of moving two-dimensional patterns depends on duration, contrast, and eccentricity. *Vision Res.* **32**, 135–147]. The model consists of two stages: in the first stage local motion detectors signal both the one-dimensional and two-dimensional motion of the high luminance features (blobs) in the plaid pattern; in the second stage these local motion signals are combined using a recursive Bayesian least squares estimation process. We demonstrate both theoretically and using simulations of the computational model that the estimated direction of the plaid motion for Type II plaids is initially dominated by the 1-D motion of the longer edges of the elongated blobs, which is in a direction close to the vector sum direction of the component gratings. The recursive estimation process which combines the local motion signals in the second stage of the model results in a dynamic shift in the estimated plaid direction towards the direction of the 2-D motion of the blobs, which corresponds to the veridical plaid direction.

Keywords. Plaid, motion, Bayes, computational model, blobs

1. Introduction

The problem of how the visual system combines the motion of two moving gratings to form the percept of a coherent moving plaid pattern is still unsolved after nearly thirty years of research. It has long been known that the plaid motion can be computed by a velocity space construction, known as the intersection-of-constraints (IOC) (Fennema & Thompson, 1979). Based on this, Adelson & Movshon (1982) proposed a two-stage model for the analysis of plaid motion in which the one-dimensional (1-D) motions of the plaid's two component gratings are first determined, and then combined in a weighted summation corresponding to the IOC construction. This model has dominated research in the area for almost thirty years, despite the psychophysical (Welch, 1989; Derrington and Suero, 1991; Derrington and Badcock, 1992; Stone, Watson and Mulligan, 1990) and physiological (Movshon, Adelson, Gizzi and Newsome, 1985; Movshon and Newsome, 1996; Tinsley, Webb, Barraclough, Vincent, Parker and Derrington, 2003) evidence being equivocal. In particular, the available evidence is based entirely on experiments using symmetric Type I plaids (Ferrera and Wilson, 1990), for which the plaid velocity vector lies between the velocity vectors of the two component gratings, which have equal magnitude. The strongest evidence against the Adelson and Movshon (1982) model was obtained when Type II plaids, the velocity vector of which lies outside of the velocity vectors of the two component gratings, were used in psychophysical experiments (Yo and Wilson, 1992). These experiments demonstrated that the direction of the plaid motion during the initial period (up to ~60 ms) of stimulus presentation is misperceived, with a strong bias in the perceived direction towards the vector sum (VS) of the velocities of the component gratings. Whilst it is possible that the Adelson and Movshon (1982) model is correct for Type I plaids, and that another mechanism is responsible for Type II plaid motion perception, this would seem highly unlikely.

Subsequent to the Yo and Wilson (1992) experiments, and prior experiments which showed identified misperceptions in the direction and speed of Type II plaids (Ferrera and Wilson, 1990, 1991), several models have been proposed which attempt to explain these misperceptions. Wilson, Ferrera and Yo (1992) suggested a model, subsequently extended by Wilson and Kim (1994), which consisted of two parallel processing pathways, one signalling the direction of the component gratings (presumed to be mediated by neurons in area V1 of visual cortex) and the other (presumed to be end-stopped neurons in area V2) signalling, after a hypothesised delay of ~77 ms, the direction of "the motion of illusory lines formed by the nodes of the Type II pattern" (Yo and Wilson, 1992). The signals of the first pathway are combined (by neurons in extrastriate area MT to which both V1 and V2 neurons project) to form a cosine-weighted sum of the component grating velocities. The signals of the second pathway are derived after full-wave rectification of the stimulus and orientation filtering at a lower spatial frequency than that of the component gratings (postulated to take place in V2). A cosine-weighted sum of the two pathways is then followed by competitive feedback inhibition in order to predict the perceived plaid direction. The delay in the second pathway accounts for the initial misperception of the plaid direction towards the vector sum direction of the component gratings' velocities. Whilst this model offers a compelling explanation of the observed misperception, it is deficient in several respects, as discussed in Alais, Wenderoth and Burke (1997), who carried out experiments on the effect of the size and number of plaid features, or blobs, the "nodes of the Type II pattern" referred to above, on the misperception. They concluded that a more likely explanation is based on "a feature sensitive mechanism which responds to the motion of plaid features and which is tuned to their various qualities" (Alais, Wenderoth and Burke, 1997). The plaid blobs which they examined and refer to are the high luminance regions which are formed at the intersection of the component gratings and which, in particular for Type II plaids, are the most visually salient features in the plaid pattern for a human observer.

In this paper we show that the misperception of the plaid direction, its dependence on the angular separation and contrast of the component gratings, and its decrease with lengthening stimulus duration, can all be fully explained by a two-stage model which is based on the detection of both the one-dimensional (1-D) and two-dimensional (2-D) motion of the blobs, and their combination by a recursive Bayesian velocity estimation process.

In the first stage of our proposed model, local motion detectors respond to both the 1-D and 2-D motion of the blobs within the plaid. We hypothesise that these detectors are based on the complex and hypercomplex (end-stopped) neurons in V1 (Hubel & Weisel, 1965; Pack, Livingstone, Duffy, and Born, 2003). This stage of the model differs from that of Wilson et al (1992) in that: (i) the 1-D motion signals are derived not from the motion of the component gratings but from the edge motion of the blobs; (ii) there is no requirement for the separate combination or any explicit weighting, cosine or otherwise, of the 1-D signals; (iii) the 2-D motion signals are derived directly from the terminations (end-points) of the blobs, without the requirement for full-wave rectification (or squaring) of the plaid stimulus. In the second stage of the model, the 1-D and 2-D motion signals are combined using a recursive Bayesian least squares estimation process, which we postulate to occur in the recurrent V1-MT circuitry. This also differs from the Wilson et al (1992) model in that cosine weighting of the 1-D and 2-D signals is not required, nor is there the need for a final stage of competitive inhibition.

In the remainder of the paper, we first examine in detail the specific geometric properties of the blobs which, we claim, play the main role in the perception of plaid motion. In particular we show that the shape of the blobs, specifically the extent of their elongation, is defined by the angular difference in the directions of motion of the component gratings, and that the orthogonal direction of motion of the longer edges of the elongated blobs is given by the mean of the directions of motion of the component

gratings. We also show that as the blobs become more elongated, the orthogonal direction of motion of the longer edges of the blobs tends towards the vector sum of the directions of motion of the component gratings. Although the blobs have been implicated in the perceptual process by several authors (Wilson et al, 1992; Burke and Wenderoth, 1993; Wenderoth, Alais, Burke and van der Zwan, 1994; Alais, Wenderoth and Burke, 1994, 1997), as far as we are aware this is the first time that the geometric properties of the blobs and their relationship to the directions of motion of the component gratings have been precisely defined. Related plaid features and their properties have been described by Bowns (1996, 2006), with similar properties, and we compare these in our discussion (section 4) with the blob features which we have defined.

Next we show theoretically how these particular properties of the blobs can be used to predict the misperception of the direction of Type II plaids which has been observed psychophysically (Yo and Wilson, 1992; Bowns, 1996; Burke and Wenderoth, 1993). To demonstrate this more fully, we use a computational version of our model to simulate the observed misperception, and show that the magnitude of the direction bias, its dependence on angular separation and contrast, and the convergence of the perceived plaid direction towards the veridical direction with increasing stimulus presentation duration, are all accurately predicted by the model.

Finally we discuss how our model differs from the two-stage model of Adelson and Movshon (1982), yet is consistent with the available physiological and psychophysical evidence, and how it relates to a recent Bayesian extension of the Adelson and Movshon model (Weiss and Adelson, 1998; Weiss, Simoncelli and Adelson, 2002), and the models proposed by Bowns (1996, 2006).

2. Model Description

2.1 Geometric analysis of the plaid blobs.

The high luminance regions of the plaid, i.e. the blobs, which are formed at the intersections of the component gratings, can be precisely defined by representing the plaid as the *product* of two gratings rather than as a *sum* of two gratings, its normal form of representation. Typically a plaid is described by the sum of two sine or cosine gratings, i.e. the spatiotemporal luminance intensity function of the stimulus is defined by

$$I(x,y,t) = \sin(\omega_1) + \sin(\omega_2) \tag{1}$$

where $\omega_i = 2\pi s_i(x\cos\theta_i + y\sin\theta_i + r_i t)$; s_i = spatial frequency (cycles/°); θ_i = direction of motion (°); and r_i = speed (°/sec), for the i th. grating, $i \in \{1,2\}$. Using a simple trigonometric identity, this expression can be rewritten as,

$$I(x,y,t) = 2\sin((\omega_1+\omega_2)/2)\cos((\omega_1-\omega_2)/2) \tag{2}$$

i.e. as the product of two anti-phase gratings, henceforth referred to as the *product* gratings to distinguish them from the *component* gratings used in the summation form (1) of the plaid. The two product gratings comprise: (i) a sine grating which moves in the direction $\phi = (\theta_1+\theta_2)/2$, and which has a spatial frequency $s_\phi = (s_1\cos\theta_1 + s_2\cos\theta_2)/2\cos\phi$, and a speed $r_\phi = (r_1+r_2)\cos\phi/(\cos\theta_1+\cos\theta_2)$; and (ii) a cosine grating which moves in the direction $\varphi = \phi - 90^\circ$, has a spatial frequency $s_\varphi = (s_1\cos\theta_1 - s_2\cos\theta_2)/2\cos\varphi$, and a speed $r_\varphi = (r_1-r_2)\cos\varphi/(\cos\theta_1-\cos\theta_2)$. For simplicity we have assumed henceforth that $s_1 = s_2 = s$.

Derrington and Ukkonen (1999) used this representation to describe a specific instance of a plaid with component gratings oriented symmetrically about the vertical and a specific relationship between the spatial frequency of the gratings and their orientations. In this case they obtain a vertically oriented, horizontally moving product grating and horizontally oriented, stationary product grating.

{Figure 1 about here}

Figure 1 shows three examples of equivalent plaid representations in terms of their component and product gratings. These clearly demonstrate that the high luminance regions, or blobs, in the plaid which occur at the intersections of the component gratings are precisely defined by the anti-phase modulation of one product grating by the other. In particular, where the spatial frequencies of the product gratings differ substantially (Figs. 1a and 1c), the blobs are readily seen to correspond to the high luminance bands of the higher frequency product grating modulated by the lower frequency product grating. The shape of the blobs, in terms of the ratio of their long and short edges, is thus determined by the ratio of the spatial frequencies of the product gratings, which can be expressed in terms of the angular difference $\theta_1 - \theta_2$ in the directions of the component gratings as

$$s_\phi / s_\varphi = 1/\tan((\theta_1 - \theta_2)/2) \qquad (3)$$

The direction of motion of the blobs (and therefore the IOC direction of the plaid) is given by the vector sum direction of the two *product* gratings.

We are mostly concerned here with Type II plaids (e.g. Figure 1a), so in order to simplify the presentation we will derive the main characteristics of the blobs only for this case, although similar equations describing the characteristics of Type I plaids (e.g. Figures 1b and 1c), can be easily obtained. For the Type II property, that the IOC velocity vector lies outside of the two component grating velocity vectors, to hold the ratio of the speeds of the component gratings, r_1 / r_2, must be greater than one, and the difference in their directions of motion, $\theta_1 - \theta_2$, must be less than 90°. It follows from (3) that, as the difference in the direction of motion of the two component gratings, $\theta_1 - \theta_2 < 90^\circ$, decreases, the ratio of the spatial frequencies of the product gratings, s_ϕ / s_φ, will increase, and the blobs will become more elongated in shape. Moreover, the shape of the blobs only depends on the difference between the directions of the two component gratings, $\theta_1 - \theta_2$, and not on the ratio of their speeds. It also follows from (3) that $s_\phi > s_\varphi$,

and thus the motion of the longer edges of the blobs orthogonal to their orientation will be in the direction $\phi = (\theta_1 + \theta_2)/2$, the mean of the directions of motion of the component gratings.

Most importantly, we can express the difference between the orthogonal direction of motion of the longer blob edges, ϕ, and the vector sum direction of the two component gratings, denoted by θ_{VS}, as

$$\phi - \theta_{VS} = \arctan\left(\frac{(r_1/r_2)-1}{(r_1/r_2)+1} \cdot \frac{s_\varphi}{s_\phi}\right) \tag{4}$$

This shows that, for a fixed ratio of component grating speeds, r_1/r_2, as the difference between directions of the two component gratings, $\theta_1 - \theta_2$, decreases, and the shape of the blobs become more elongated, the angular difference between the orthogonal direction of motion of the longer blob edges and the vector sum direction of the component gratings will decrease. It is also worth noting that for a fixed difference in the directions of the component gratings, $\theta_1 - \theta_2$, as the speed ratio r_1/r_2 increases, the angular difference expressed by (4) will increase, causing the orthogonal direction of motion of the longer edges of the blobs to move away from the vector sum direction of the component gratings.

2.2 Theoretical predictions of the model.

The geometric analysis of the blobs, as expressed by equations (2) - (4), allow us to make theoretical predictions about the behaviour of our model in response to Type II plaids. In the first stage of the model, we propose that local motion detectors signal both the 1-D (edge) and 2-D (end-point) motion of the blobs present within the plaid. Thus in the case of Type II plaids, for which the blobs are elongated, the majority of the local motion detectors will respond to the 1-D motion of the longer edges of the blobs. Since a local motion detector signals the velocity of 1-D edge motion in the direction orthogonal to the

orientation of the edge, owing to the aperture effect (Wallach, 1935; Wuerger, Shapley and Rubin, 1996, Marr and Ullman, 1981), the majority of the local motion detectors will signal motion in the orthogonal direction of motion of the long edges of the blobs. The geometric analysis of the previous section shows that for a fixed ratio of component grating speeds, r_1 / r_2, as the difference between the directions of the two component gratings, $\theta_1 - \theta_2$, decreases and the shape of the blobs become more elongated, the orthogonal direction of motion of the longer edges of the blobs, ϕ, will tend towards the vector sum direction of the component gratings. Thus the majority of the local motion detectors will signal motion in a direction which is increasingly biased, as $\theta_1 - \theta_2$ decreases, towards the vector sum direction of the component gratings.

In the second stage of the model, we propose that the outputs of the local motion detectors are combined using a recursive Bayesian estimation process. The estimate computed in the first iteration of the estimation process will thus form the model's prediction of the perceived plaid velocity in a short initial period of stimulus presentation. As we have already discussed, this estimate will be dominated by the majority of local motion detectors which signal the orthogonal motion of the blobs in the ϕ direction. We have also shown, in equations (3) and (4) respectively, that as the difference between component grating directions, $\theta_1 - \theta_2$, decreases: (i) the long edge of the blob will become longer and therefore drive an increasing majority of local motion detectors; and (ii) the orthogonal motion of the blobs in the ϕ direction approaches the vector sum direction of the component gratings. Hence it follows that, as the angle between the component gratings decreases, the first velocity estimate formed by the model, and therefore the initial plaid velocity predicted by the model, will be increasingly biased towards the ϕ direction, which itself will approach the vector sum direction of the

component gratings. This is precisely what Yo and Wilson (1992) observed in their psychophysical experiments.

For example, consider one of the Type II plaids used by Yo and Wilson in their experiments. The parameters of the component gratings of this plaid are: $\theta_1 = 70.5^\circ$, $\theta_2 = 48.2^\circ$, $r_1 = 1.33$, and $r_2 = 2.67$. Then, for these values: $\theta_1 - \theta_2 = 22.3^\circ$, $\theta_{IOC} = 0.2^\circ$, $r_{IOC} = 3.9$, $\theta_{VS} = 55.6^\circ$, $r_{VS} = 4.0$, $\phi = 59.4^\circ$, $\phi - \theta_{VS} = 3.8^\circ$ and $s_\phi / s_\varphi = 5.1$. The blobs are thus elongated (edge ratio of 5:1) and move orthogonally to their longer edges in a direction which is less than 4° from the vector sum direction of the component gratings. In Yo and Wilson's experiment, the perceived direction of the plaid motion in the initial period of presentation was observed to be approximately 60°. This is close to the vector sum direction of 55.6°, and almost exactly equal to the orthogonal direction $\phi = 59.4^\circ$ of motion of the longer edges of the blobs.

The velocity estimate formed by the model during subsequent iterations of the recursive estimation process will also be influenced by the majority of local motion detectors which signal the orthogonal direction ϕ of the longer edges of the blobs, although this influence will gradually decrease with each iteration (see the simulation model description below) leading to convergence to a steady-state velocity estimate. Thus for long stimulus presentations the perceived direction of the plaid motion predicted by the model will continue to be biased, but to a lesser extent, in the direction $\phi = (\theta_1 + \theta_2)/2$, the mean of the component gratings' directions. This is precisely what Ferrera and Wilson (1990) observed, i.e. that the perceived direction of the plaid motion has a small residual bias, after approximately 150ms of presentation time, of between 8°

and 10° towards the mean of the component gratings' directions, in this case for plaids with component grating separations of between 22.3° and 51.6°. A similar residual bias was observed by Burke and Wenderoth (1993). They found in addition that as the difference in component grating directions decreased from 40° to 10°, the residual bias increased from 2° to 17°. This dependence of the residual bias on the difference in component grating directions was observed for a constant value of $\phi = (\theta_1 + \theta_2)/2$. Hence they argued that the bias could not be due to the orthogonal direction of motion of the elongated blobs which remained constant in this experiment. In our model however the strength of both the initial and the residual bias is determined by the length of the long edges of the blobs, since this determines the number of local motion detectors which signal the orthogonal 1-D motion of the blob edges in the ϕ direction. Since the elongation of the blobs increases with decreasing difference in the direction of motion of the component gratings, as shown by equation (3), it follows that the residual bias will always be towards the ϕ direction, but will increase as the difference in component grating directions decreases.

The theoretical predictions of the model, presented above, are largely qualitative in nature, but will be confirmed in a more quantitative form in section 3 of the paper, where we describe the results from using of a computational version of our model to simulate the perceptual experiments of Yo and Wilson (1992), Bowns (1996) and Burke and Wenderoth (1993). The form of the computational model is described in the next section.

2.3 Computational model description

To quantify the predictions of our model and, in particular, to demonstrate the convergence of the estimated direction of the Type II plaid motion towards the true IOC

direction, we will use a computational version of the model to simulate the psychophysical experiments of Yo and Wilson(1992) and Burke and Wenderoth (1993).

A detailed description of the model has been given previously (Dimova and Denham, 2009), where the model was used to explain the initial direction bias in the velocity of smooth eye pursuit eye movements (Masson and Stone, 2002; Wallace, Stone and Masson, 2005). Briefly, the input to the model is the luminance function $I(x,y,t)$ of equation (1), describing the plaid pattern and its motion, which is presented in a 200 x 200 pixel visual space. Local motion detectors measure the spatial and temporal derivatives I_x, I_y, I_t of $I(x,y,t)$ in a number of 10 x 10 pixels, non-overlapping windows uniformly distributed across the visual space, using a simple spatial and temporal shift mechanism. These measurements are then related to the pattern velocity vector $\begin{bmatrix} v_x \\ v_y \end{bmatrix}$ by the gradient-based equation (Fennema and Thompson, 1979):

$$I_t = \begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} + \eta \qquad (5)$$

where η is additive zero mean, normally distributed measurement noise. A recursive algorithm, well-known as the Kalman filter (Kalman, 1961), is used to determine a least squares estimate of the velocity vector based on the set of measurements from the local motion detectors, as the best-fit solution to the corresponding set of gradient-based equations (5). The velocity estimate in the estimation algorithm is initialised to zero, which corresponds assigning a zero mean, *a priori* velocity distribution in the Bayesian formulation of the estimation algorithm.

As we have described, the bias in the first velocity estimate formed by the model results from the large number of local motion detectors for which the measured derivatives I_x, I_y, I_t correspond to the 1-D motion of the longer edges of the blobs. For these detectors many solutions to the corresponding gradient-based equations (5) are possible, corresponding to the aperture effect (Wallach, 1935; Wuerger, Shapley and

Rubin, 1996, Marr and Ullman, 1981). The zero-valued initial velocity estimate provides a constraint on the estimate formed by the first step of the algorithm, which results in an best-fit solution being selected for which the magnitude of the velocity estimate is smallest. This corresponds to the solutions to (5) for each local motion detector for which the selected velocity is in the direction orthogonal to the longer edges of the blobs. Thus the estimate formed in the first step of the algorithm will be strongly biased in this direction, with the strength of the bias dictated by the number of motion detectors signalling the direction. As we have shown, the bias will be stronger as the difference between the directions of the component gratings decreases, since this results in a greater elongation of the blobs.

In contrast, measurements of I_x, I_y, I_t from the local motion detectors which signal the 2-D motion of the end-points of the blobs result in a unique (within the noise) solution to the corresponding set of gradient equations (5) . This solution corresponds to the vector sum direction of the *product* gratings, and thus, equivalently, to the veridical, IOC direction of the plaid. These local motion detectors will therefore influence the estimate of plaid direction towards the IOC direction, both in the initial step of the algorithm and in all further steps. However, lowering the contrast of the plaid stimulus, or equivalently reducing the signal to noise ratio in equation (5) , will result in a weaker influence of this solution, and thus allow a greater bias in the estimated direction of the plaid towards the vector sum direction of the *component* gratings.

As the number of iterations of the recursive estimation algorithm increases, the effect of the 1-D local motion detectors will decrease in relation to that of the 2-D motion detectors, since the velocity estimate formed in each iteration of the algorithm becomes the prior estimate for the next iteration. This gradually relaxes the effect of the zero prior constraint on the solution to (5) corresponding to the set of outputs of the 1-D motion

detectors, allowing the solution to (5) corresponding to the set of outputs of the 2-D motion detectors to increasingly influence the velocity estimate at each iteration.

In the following section we will show by simulations of the computational version of the model that, in accordance with the above theoretical predictions, the model also yields quantitative predictions of the perceived direction of plaid motion which closely resemble the experimentally obtained data of Yo and Wilson (1992), Bowns (1996) and Burke and Wenderoth (1993).

3. Simulation Results

3.1 Simulation of the Type II plaid experiments of Yo and Wilson (1992) and Bowns (1996)

In Figures 2a - 2f, we show the results obtained using the computational model to simulate the psychophysical experiments of Yo and Wilson (1992) and Bowns (1996). Yo and Wilson (1992) used as stimulus three different Type II plaids (see their Figure 2), although the results were not given for all three plaids for each of the experiments. The main experiments, which we simulate here, observed the perceived direction of plaid motion as functions of presentation duration and pattern contrast. Their results on presentation duration are given for the plaid with the following parameters: $\theta_1 = 70.5^\circ$, $\theta_2 = 48.2^\circ$, $r_1 = 1.33$, $r_2 = 2.67$, $s_1 = s_2 = 1.5$, $\theta_1 - \theta_2 = 22.3^\circ$, $\theta_{IOC} = 0.2^\circ$, $r_{IOC} = 4.02$, $\theta_{VS} = 55.6^\circ$, $r_{VS} = 3.93$. For the product plaid representation, these parameters give: $\phi = 59.4^\circ$, $\phi - \theta_{VA} = 3.8^\circ$, $r_\phi = 2.0$, $r_\varphi = 3.5$, $s_\phi = 1.5$, $s_\varphi = 0.3$ and $s_\phi / s_\varphi = 5.0$. Note that both the speed and spatial frequency of the higher spatial frequency product grating are similar to those of the component gratings.

For this plaid, two observers reported a perceived direction of motion of approximately 60° after 60 ms of presentation, reducing to approximately 15° and 30° respectively after 90 ms, and to approximately 0° (the IOC direction) after 150 ms. Note

that the reduction in the bias was apparently gradual rather than occurring discontinuously as might be expected if additional 2-D motion information became available after some fixed time delay, as was suggested in the Wilson et al (1992) model. When the plaid contrast was varied, with values of 5%, 50% and 100%, the observed initial bias at 60 ms was 60°, 40° and 30° respectively, and the length of time for the bias to reduce lengthened considerably with decreasing contrast. For a contrast of 5-10%, a substantial bias of approximately 25° was observed after 1 sec of presentation.

{Figure 2 about here}

We can compare these experimental results with the graph shown in Figure 2a, which shows the model results for this plaid. As the graph shows, the bias in the estimated direction at the first iteration for the three values of contrast, 25%, 50% and 100%, are remarkably similar to the initial perceived bias observed experimentally. We note also that the convergence time decreases substantially with increasing contrast, and that there is a considerable steady-state bias for all contrasts of up to 25° for this plaid, again as observed experimentally. Figures 2b and 2c show the same simulations for the other two plaids used by Yo and Wilson (1992), but for which they did not report the results as fully as for the first plaid. These graphs show similar characteristics of the variation in magnitude and convergence rate of the direction bias with contrast as in Figure 2a, but with the steady-state bias reducing with increasing difference in the directions (47.4° and 51.6° respectively) of the component gratings in Figure 2b and 2c, to between 4° and 10°. In Ferrara and Wilson (1990), the perceived steady-state bias for similar Type II plaids was approximately 6°.

In Bowns (1996), a number of experiments were carried out which attempted to establish whether or not the misperception of the plaid direction observed by Yo and Wilson (1992) generalises to all Type II plaids and is due to a temporal delay in Fourier and non-Fourier motions processing as proposed in the parallel pathway model of Wilson et al (1992). Here we have simulated their Experiment 3 which used Type II plaids very similar to those used by Yo and Wilson (1992). The component gratings for these plaids had the same spatial frequencies (1.3 cycles/°) and orientations (202° and 225°) but differed in the ratio of their speeds, which ranged from 1:0.45 to 1:0.75, with the speed of one of the component gratings held constant at 3.13 °/sec. The experiments used a simple forced choice response which required subjects to report either a plaid direction to the right or to the left of "the vertical", i.e. 90°. The component grating directions and speeds were such that the vector sum direction remained virtually constant , varying from 29° to 32° to the left of the vertical, for the varying speed ratios, whereas the IOC direction varied from 28° to 2° to the right of the vertical.

The experiments revealed that for the two speed ratios at the extreme ends of the above range, subjects reported a perceived direction of plaid motion which shifted from 100% in the vector sum direction (i.e. left of vertical), for a speed ratio of 1:0.75, to 100% in the IOC direction (i.e. right of vertical), for a speed ratio of 1:0.45. This was interpreted in Bowns (1996) as: "a rather surprising complete reversal of the perceived motion in the direction of the IOC".

We simulated the cases of the two plaids at the extremes of the ranges of speed ratios referred to above. The experimental data was also simulated by Weiss and Adelson (1998) - see our discussion of their model in section 4. The simulation results from our

model are described in Figure 2d (for a speed ratio of 1:0.75) and 2e (for a speed ratio of 1:0.45) in the form of vector space diagrams. As these Figures show, changing the ratio of the component grating speeds from 1:0.75 to 1:0.45 is sufficient to move the both the estimate formed in the first step of the estimation algorithm, v_{e1}, and the steady- state estimate, v_{es} of the perceived plaid direction from being on the left of the vertical (vector sum side) to being on the right of the vertical (IOC side).

The difference in the directions of the first step velocity estimate v_{e1} for the two speed ratios is 21° (108° vs. 87°). However the difference in the first step direction bias estimate (relative to the IOC direction) is only 5° (20° from IOC vs. 25°). For the steady-state velocity estimate v_{es}, the estimated direction differs by 24° (102° vs. 78°) for the two speed ratios, but the difference in the estimated bias is only 2° (14° from IOC vs. 16°).

Thus the change in the estimated bias is small with this change in speed ratio, both in the first step of the algorithm and after convergence, and we suspect that the change in the perceived bias is also small. The simple forced choice response of left or right of the vertical appears however to have resulted in an interpretation in Bowns (1996) that there is a large change in bias which leads to a reversal in the perception of the plaid motion direction from the IOC to the vector sum direction.

We suggest an alternative interpretation, supported by our simulation results (see Figures 2d and 2e), that the value of the perceived bias for the two speed ratios is almost the same, but that the change in speed ratio results in a shift in the IOC direction towards the vector sum direction, causing the perceived motion direction to switch from right side of the vertical to the left side.

In the Discussion section we will describe also our simulation results for the set of plaids used in Experiment 2 of Bowns (1996), with angular differences between component gratings in a range between 10° and 90°, and a speed ratio of 1:0.5.

3.2 Simulation of the experiments of Burke and Wenderoth (1993)

In Figure 3, we show the results obtained using the computational model to simulate the psychophysical experiments of Burke and Wenderoth (1993), in which they used Type II plaids to study the dependence of the steady-state misperception of plaid direction on the angular difference between the component grating directions. The plaids were constructed using component gratings with angular differences of 10°, 20°, 30° and 40°. The true plaid direction was 270° and the stimulus was presented for 10 sec. Two experiments were carried out: in the first, the component directions were chosen so that the mean was constant at 295°; in the second, one component direction was kept constant at 315°. Figures 3a and 3b show the results from each experiment, both the perceived plaid direction obtained in the Burke and Wenderoth (1993) study (▲symbols) and the direction estimated by the model (● symbols). In Figure 3a the mean component direction is 295°, and in Figure 3b this direction varies and is shown by the dashed line.

{Figure 3 about here}

The graphs in Figure 3 show that the estimated plaid direction from the model simulation varies with the difference in component grating direction and displays in both cases the same trend in the variation as observed in the Burke and Wenderoth (1993) study, although with a slightly greater bias towards the mean component direction of up to 9°. Importantly the model shows in Figure 3b the same non-linear variation of the estimated direction with component separation as was observed experimentally for the perceived direction.

3.3 Robustness of the model

In the above cited experiments and those that are described later in the Discussion, the stimuli were presented in a circular windows with the following diameters: Yo and Wilson (1992) - diameter = 8°; Bowns (1996) - diameter = 3°; Stone, Watson, and Mulligan (1990) - diameter = 5.4° ; Champion, Hammett,and Thompson (2007) - diameter = 6° ; Alais, Wenderoth, and Burke (1997) - diameter = 3°, 6° and 12°. We do not have any information on the size of the stimulus used in the experiments of Burke and Wenderoth (1993). In the simulations described in Section 3.2 and in the Discussion, we display the image in a circular aperture of diameter 200 pixels; thus the size of our 10 x 10 pixel window corresponds to between 0.15° and 0.6°. This is in close accordance with an average receptive field diameter measurement, for V1 cells in humans, of approximately 0.25° at the fovea, rising linearly to approximately 0.6° at 6° eccentricity (Smith, Singh, Williams and Greenlee, 2001).

Since our results closely match the experimental results in each of these experiments, we can infer that our model results are robust if the 10 x 10 pixel window represents a receptive field diameter of between 0.15° and 0.6°, which is the approximate physiological range for V1 cells.

Our model breaks down when the simple algorithm we use to calculate the image intensity derivatives fails to produce acceptably accurate results. This happens when the spatial frequency of the stimulus is sufficiently high that the spatial period falls within a single window, ie is less than 10 pixels, corresponding to a frequency of 0.1 cycles/pixel, or between 6.7 cycles/° (corresponding to window size of 0.15° and a stimulus aperture diameter of 3°) and 1.6 cycles/° (corresponding to a window size of 0.6° and a stimulus

aperture diameter of 12°). Thus, for the simulations of the Alais et al (1997) experiments described in the Discussion, in which the aperture diameter is 3°, we did not simulate the result for a stimulus of 6 cycles/°.

It is important to note that the parameters of the model were held constant for all the simulation results described in this Section, i.e. for the Yo and Wilson (1992), the Bowns (1996), and the Burke and Wenderoth (1993) experiments.

4. Discussion

The original two-stage model (Adelson and Movshon, 1982; Movshon et al, 1985)) has dominated research in plaid motion perception for almost thirty years, leading to an almost universal view that the first stage of plaid motion analysis involves the detection of the 1-D motion of the component gratings, carried out by component-direction selective neurons in V1 (see the review by Pack and Born (2008)). It is important to note however that the available evidence is almost entirely based on using symmetric Type I plaids, in which the component gratings move with equal speeds. For the psychophysical experiments of Movshon et al (1985) the difference in directions of the component gratings was 120°, for their physiological experiments in cat and monkey V1 and in monkey MT (Movshon et al, 1985) the angular difference was 90°, and for Movshon and Newsome's (1996) physiological experiments in monkey V1 the difference was 90° or 45°. For such plaids, a neuron in primary visual cortex (V1) which responds optimally to the motion of a single grating, produces little response to a plaid moving in its optimal direction, as would be predicted from the orientations of the component gratings if the neurons were only responding to the 1-D motion of the gratings (Movshon et al, 1985).

Our proposed model suggests that neurons in V1 respond both the 1-D and 2-D motion of the blob features of the plaid, and in the case of Type II plaids are driven by the 1-D edges and 2-D end-points of the elongated blobs. Moreover, we suggest that the 2-D motion is detected by end-stopped cells in V1, as observed by Pack et al (2003). As we have discussed above, this model leads to theoretical and simulation results which closely mimic the physiological observations of perceived direction for such plaids. So how does the model explain the component-selective responses for V1 neurons in the case of Type I plaids, as observed by Movshon et al (1985) and Movshon and Newsome (1996), in particular as the neurons observed by Movshon and Newsome (1996) were apparently mostly of the end-stopped variety ?

For Type I plaids in which the difference in the component grating directions is 90°, the blobs take the form of small square regions of high luminance which are aligned in the same orientations as the component gratings. Therefore, a neuron which is optimally responsive in the direction of the plaid motion, and with a long, narrow receptive field oriented orthogonally to the plaid direction, will respond sub-optimally to the two lines of blobs, each moving at 45° to the optimal direction for the neuron, in exactly the same way as if it were responding to the component gratings themselves, as shown by Tinsley et al, 2003. Little or no 2-D motion signal in the direction of the plaid would be detected due to the absence of well defined end-points in the stimulus, in contrast to the case of Type II plaids with elongated blobs. It is significant however that Movshon and Newsome (1996) observed a degree of pattern-selective response in two of the nine neurons they measured. Thus, we suggest, for such neurons and for symmetric Type I plaids, it is not possible to distinguish whether the neurons are responding to the

component gratings or to the lines of small square blobs (which we will refer to as blob-lines) present in the plaid moving in the same directions as the component gratings. The lines formed by the blobs are certainly more perceptually salient to the human observer than the individual component gratings.

In our model, the outputs of the local motion detectors signalling the two orthogonal 1-D motions of the blob-lines described above will be combined in the second stage by the estimation algorithm to yield the vector sum of these two motions, the direction of which corresponds exactly to the IOC direction of motion in the case of a Type I plaid. Note that no initial or steady-state perceived direction bias was observed for Type I plaids by Ferrera and Wilson (1990) or Yo and Wilson (1992). It is also possible that the 2-D motion of the individual blobs may be signalled by V1 neurons with short, wide receptive fields, as observed by Tinsley et al (2003). The combination of the outputs of the 2-D motion detectors and the 1-D motion detectors in the second stage of the model would reinforce the computation of the velocity estimate in the true plaid direction.

In summary, we argue that for symmetric Type I plaids, with a difference between the component grating directions of around 90°, the 1-D motion detectors in V1 will respond in exactly the same way to the blob-lines as to the component gratings. Since the blob-lines and the component gratings are indistinguishable, in terms of their orientation, direction, spatial frequency and speed, it is thus impossible for any experiment with such Type I plaids to distinguish between a model in which the first stage responds to the 1-D motion of the component gratings and one in which the first stage responds to the 1-D motion of the blob-lines. Since, in addition, the direction of a symmetric Type I plaid is

given by the simple averaging (vector average) of the 1-D motion directions, it is impossible to distinguish between a model in which the second stage computes the IOC direction from one in which the second stage computes the vector average direction. We therefore conclude that the psychophysical experiments (Welch, 1989; Derrington and Suero, 1991; Derrington and Badcock, 1992) using symmetric Type 1 plaids, which have apparently confirmed the two-stage model of Adelson and Movshon (1982), are wholly inadequate in this respect. In contrast, the psychophysical experiments with Type II plaids (Yo and Wilson, 1992) strongly challenge the Adelson and Movshon model.

It is worth noting here that our theoretical analysis of the plaid blobs indicates, for asymmetric Type I plaids with an angular separation of component directions of >90° (e.g. the plaid in Figure 1c), that a similar elongation of the blobs occurs, and that the longer edges increase in length as the angular separation increases. Also the orthogonal direction of the longer edges of the blobs approaches the vector sum direction of the component gratings. Hence our model predicts for such plaids a significant bias in the perceived direction of plaid motion towards the vector sum direction of the component gratings and away from the true IOC direction, of comparable magnitude to that observed for Type II plaids. As far as we are aware, no psychophysical or physiological experiments have been carried out for such Type I plaids.

A recent model of motion integration (Weiss and Adelson, 1998; Weiss et al, 2002) aimed at extending the Adelson and Movshon (1982) model to accommodate the Yo and Wilson (1992) results. According to Weiss and Adelson (1998) and Weiss et al (2002), their model captures the uncertainty in the 1-D motion of the component gratings in the case of low contrast by using a Bayesian estimation process. The Bayesian

formulation of the model results in the identification of a distribution of 1-D and 2-D velocity measurements which correspond to local likelihood functions. The model therefore represents the 1-D motion of each of the component gratings, corresponding to the first stage of the Adelson and Movshon model, as a pair of "fuzzy" (Weiss and Adelson, 1998) constraint lines in velocity space, the degree of fuzziness being dependent on contrast. The estimate of the plaid direction is then given by the mean/maximum of the posterior probability distribution, which is computed from the product of the local likelihoods and the prior distribution for the velocity estimate. The latter is assumed to be Gaussian with zero mean according to a "slow and smooth" (Weiss and Adelson, 1998; Weiss et al, 2002) hypothesis based on suggestions that human observers prefer the slowest motion consistent with the visual input (Ullman, 1979).

In fact, the model described by Weiss and Adelson (1998) and Weiss et al (2002) is identical to the first step of the recursive Kalman filter estimation algorithm in our model, and therefore produces an identical, biased first step estimate of plaid direction. There appears therefore to be a contradiction between the explanation in Weiss and Adelson (1998) of the behaviour of the model in predicting plaid motion, which is solely in terms of the 1-D motion of the component gratings, and our explanation, which is in terms of the 1-D and 2-D motion of the edges and end-points blobs. The explanation in Weiss et al (2002) is essentially the same as that in Weiss and Adelson (1998) but less detailed and with no supporting diagrams.

To resolve this contradiction, we first consider the plaid used to produce the simulation results shown in Figure 2a, and previously discussed in Section 2.2. This plaid is also used in Weiss and Adelson (1998) and Weiss et al (2002) as their main example

for demonstrating the misperception of the direction of Type II plaids. The parameters of the component gratings of this plaid are, as given before: $\theta_1 = 70.5^\circ$, $\theta_2 = 48.2^\circ$, $r_1 = 1.33$, and $r_2 = 2.67$, yielding the following values: $\theta_1 - \theta_2 = 22.3^\circ$, $\theta_{IOC} = 0.2^\circ$, $r_{IOC} = 3.9$, $\theta_{VS} = 55.6^\circ$, $r_{VS} = 4.0$, $\phi = 59.4^\circ$, $r_\phi = 2.03$, $r_\varphi = 3.46$, $\phi - \theta_{VS} = 3.8^\circ$ and $s_\phi / s_\varphi = 5.1$. For a contrast of 50%, our model calculates the initial estimate of the plaid velocity vector v_{e1} as: speed $r_{e1} = 1.84$ and direction $\theta_{e1} = 40^\circ$.

{Figure 4 about here}

Figure 4a illustrates the plaid, clearly showing the elongated blobs. Figure 4b is a velocity space diagram on which the velocity vectors of the component gratings, v_1 and v_2, together with their constraint lines, the initial velocity estimate, v_{e1}, and the IOC and vector sum velocity vectors, v_{IOC} and v_{VS} respectively, are shown (0° is vertically upward in this diagram). The velocity space diagram in Figure 15d of Weiss and Adelson (1998) is redrawn as an inset in Figure 4b.

In their Figure 15d, the latter authors indicate the magnitude of the vector average (VA) velocity of the component gratings, rather than the vector sum velocity. Although the *direction* of these two velocity vectors are the same, the *magnitude (speed)* of the vector average velocity is half that of the vector sum velocity.

Weiss et al (2002) explain the bias in the perceived direction towards the vector sum/average direction by the statement that "the vector average velocity [speed] is much slower than the IOC solution and hence it is favored [by the zero prior] at low contrasts". They assume that the Bayes estimate of plaid velocity is based on "local likelihoods

[which] are 'fuzzy' constraint lines" (Weiss and Adelson, 1998) defined by the component grating velocities. Thus any bias in the estimate towards a speed slower than the IOC speed, as a consequence of the zero prior, will automatically result in a shift of the direction of the estimated velocity away from the IOC direction and towards the vector sum/average direction, i.e. the velocity estimate will be constrained to fall along, or close to, the dashed line depicted in Figure 4b.

The explanation of the perceived direction bias in Weiss et al (2002) and Weiss and Adelson (1998) is thus based on the Adelson and Movshon (1982) model of plaid perception, in which only the 1-D motion of the component gratings are detected in the first stage of analysis of the plaid motion, and their model is presented as a Bayesian extension of this model. This is clearly reflected in their explanation since they indicate that their model forms local likelihoods as "fuzzy" constraint lines defined by the 1-D motion of the component gratings. However their model, as ours, undoubtedly detects both the 1-D and 2-D motion that is present in the stimulus in the form of the motion of the edges and end-points of the blobs, as is clearly demonstrated by their depiction (in Figure 3 of Weiss et al (2002)) of the likelihood functions generated by their model for a moving diamond stimulus. It is surprising therefore that no reference is made to the likelihood functions formed from the 2-D motion in the plaid stimulus, and their role in forming the estimate.

We offer here an alternative explanation for the perceived plaid motion, which is based on the 1-D and 2-D motion of the edges and end-points of the blobs. This is illustrated in Figure 4c. Here we show the velocity vectors corresponding to the orthogonal motion of the longer and shorter edges of the blobs in the plaid,

$$v_\phi \triangleq \begin{bmatrix} r_\phi \cos\phi \\ r_\phi \sin\phi \end{bmatrix}, v_\varphi \triangleq \begin{bmatrix} r_\varphi \cos\varphi \\ r_\varphi \sin\varphi \end{bmatrix}$$

respectively, and their constraint lines, together with the IOC and vector sum velocities, v_{IOC} and v_{VS} respectively, of the component gratings. Also shown are the velocity estimates from our model for the first step ($v_{e1} = 1.83$; $\theta_{e1} = 40^\circ$), the fifth step ($v_{e5} = 2.38$; $\theta_{e1} = 25^\circ$), and in the steady-state ($v_{es} = 2.62$; $\theta_{es} = 18^\circ$).

It is clear that the initial estimate v_{e1} lies very close to the velocity vector v_ϕ, corresponding to the orthogonal motion of the long edge of the blob, and to the maximum of the likelihood function (the "fuzzy" constraint line) for v_ϕ. Subsequent velocity estimates in further iterations of the recursive algorithm get closer to this maximum, and also to the velocity vector v_{IOC}. Note that the effective prior for each step in the estimation algorithm is given by the velocity estimate in the previous step, which together with the influence of the likelihood function corresponding to the 2-D velocity of the end-points of the blobs, v_{IOC}, leads to the convergence of the estimate towards the IOC velocity.

To reinforce our account of the model behaviour, we provide a further piece of evidence that the first stage of plaid motion perception is based on the 1-D and 2-D motion of the blobs rather than the 1-D motion of the component gratings. Stone, Watson and Mulligan (1990) investigated the effect on the perceived plaid direction of making the contrasts of the component gratings unequal. They based their investigation on the Adelson and Movshon (1982) model, assuming their first stage in which the 1-D velocities of the component gratings were detected to be correct. They hypothesised that the low contrast grating would be detected at a lower speed than the true value and that if

this erroneous value were used in a second stage IOC calculation of plaid direction, a significant contrast-dependent error in the perceived plaid direction would result. They used a Type I plaid with angular separation of the component gratings of 120°, and changes in the ratio of the speeds of the component gratings to vary the true direction of the plaid whilst maintaining a constant plaid speed. In this way they found that the perceived plaid direction was biased towards the direction of the higher contrast grating and this bias increased for increasing contrast ratio, and also for decreasing total contrast (the sum of the grating contrasts). At 5% total contrast, the average observed bias varied between 0°, at a contrast ratio of 1, to ~16°, at a contrast ratio of 4:1. A maximum bias of 20° was observed for a total contrast of 10% and a contrast ratio of 8:1. The modified Adelson and Movshon (1982) model proposed by Stone et al (1990) using perceived rather than actual component speeds appeared to give qualitatively similar results to those observed (see their Figure 11). However, similar experiments by Champion, Hammett and Thompson (2007) appeared to invalidate the modified IOC model of Stone et al (1990), since it would also predict a bias towards the direction of the low contrast component at high component grating speeds due to an increase in the perceived speed of low-contrast gratings for grating speeds above ~12 deg/s (Champion et al, 2007). Champion et al. observed an increasing bias with component speed which was always towards the direction of the high contrast component except for the very lowest component grating speeds, but a *decrease* in the bias at the highest component speeds (above 12 deg/s), consistent with their observed switch in the contrast-related misperception of grating speed for higher speed gratings. It should be noted however that Champion et al used plaids of total contrast equal to 90%, compared to the total contrast

values of between 5% and 40% used by Stone et al. They also used component gratings with angular separation of 90°, compared with the 120° angular separation used by Stone et al. Champion et al also suggest that their results are inconsistent with the Bayesian IOC model of Weiss et al (2002), since that model relies upon the perceived speed of the gratings being smaller for lower contrast, and hence higher uncertainty, owing to the greater influence of the "slow" prior. Champion et al also claim that their results are inconsistent with several other models of plaid perception including the 1-D and 2-D parallel pathways model of Wilson et al (1992), and the blob tracking model of Alais, Wenderoth and Burke (1994).

{Figure 5 about here}

Applying our model to this data shows that it replicates the misperception of the direction of plaid motion towards the direction of the higher contrast grating, but the magnitude of the bias in the estimated direction is dependent on the spatial frequency of the component gratings. The case of a plaid with a separation of component gratings of 120° , 60° either side of the vertical (0°) and a contrast ratio of 4:1 is shown in Figure 5a. It is clear that the salient feature of this plaid is a set of blob-lines which are formed from a joining-up of the plaid blobs. The direction of motion of the blobs is the IOC direction of the component gratings, i.e. the plaid direction of 0°, but the orthogonal direction of motion of the blob-lines is 300°, the direction of the higher contrast component grating. The estimated plaid direction computed by our model is 308°, giving a bias of 52°away from the IOC direction of 0° towards the higher contrast grating direction, much greater than that measured by Stone et al (1990), where the direction error was up to 20° for this

contrast ratio (4:1). However, our result was obtained for a grating spatial frequency and a viewing aperture shown for the plaid illustrated in Figure 5a, corresponding to ~14 cycles of the component gratings being present within the viewing aperture. If the gratings' spatial frequency and the viewing aperture are changed to approximate that used by Stone et al (1990) and Champion et al (2007), approximately 6 cycles of the component gratings are present in the viewing aperture, as illustrated by the plaid in Figure 5b. The blob lines are still clearly visible but the size of the blobs is greater by about a factor of two. In this case, the estimated plaid direction computed by our model is 342°, giving a bias towards the direction of the higher contrast grating of 18°, comparable to that measured by Stone et al (1990) for this contrast ratio (4:1). The bias computed by our model for the contrast ratio of 2:1 was 7°, which is consistent with the Stone et al result of approximately 7° for the 5% contrast case, and with the results of Champion et al (2007) who used a contrast ratio of 2:1 and obtained a maximum bias of approximately 7°.

The above example illustrates the importance, both in psychophysical experiments and in modelling, of the choice of the spatial frequency of the component gratings in relation to the viewing angle/aperture of the stimulus. Our model results would suggest that if the psychophysical experiments of Stone et al (1990) or Champion et al (2007) had been carried out using a higher component grating spatial frequency, a far greater bias towards the higher contrast grating would have been obtained, owing to the greater salience of the 1-D motion of the blob-lines in the direction of the higher contrast grating, compared to that of the 2-D motion of the blobs themselves, when viewing the plaid.

Alais et al (1997) investigated the effect of blob size and number on perceived plaid direction, in this case for Type II plaids. They showed, by varying both spatial frequency and viewing aperture size, that there is a large effect of blob size on the perceived direction bias, of up to 14.1°, due to changes in the component spatial frequency, but a small effect of blob number, of about 5°, obtained by changing aperture size whilst spatial frequency is held constant. We simulated their experiments with our model, keeping the viewing aperture constant and varying the spatial frequency of the component gratings. We used three values of spatial frequency: 0.6, 0.3 and 0.2 cycles/pixel. For the sake of comparing our simulation results with the experimental results, we assumed that these spatial frequencies corresponded to the experimental values of 3.0, 1.5, and 1.0 cycles/°.

{Figure 6 about here}

We show in Figure 6 the results from Alais et al (1997) (their Figure 5) giving the perceived direction as a function of spatial frequency for the 3° aperture case (● symbols), together with the steady state direction estimates from the model (▲symbols), *for each of the component grating spatial frequencies. We do not simulate the 6.0 cycles/°* owing to the limitations of our model in dealing with such high frequencies due to our choice of window size. As can be seen from Figure 6, the estimates of the plaid direction are very similar to the perceived experimental values and, importantly, show the same trend, with a decrease in the misperceived direction bias as the spatial frequency of the component gratings increases.

It is not clear how a model of plaid perception based on the Adelson and Movshon (1982) model might account for the dependence of the misperception of plaid direction on component grating spatial frequency, as observed by Alias et al (1997) and modelled by us. Varying the spatial frequency of the component gratings should have no effect on the computation of the 1-D velocity of the gratings, or on the IOC calculation, even in the case of where uncertainty in the component directions is taken into account as in the Bayesian IOC model of Weiss et al (2002) and Weiss and Adelson (1998). On the other hand, our model, which depends on both the 1-D and 2-D motion of the blobs, is entirely consistent with the Alais et al (1997) results. As noted by them: "These results provide further support for the existence of a feature-sensitive mechanism which responds to the motion of plaid features and which is tuned to their various qualities". Our model provides just such a mechanism.

Other approaches based on a feature tracking mechanism have been proposed which are related to the mechanisms that we have described here. In particular, Bowns (1996) proposed a feature tracking explanation for the misperception of Type II plaids as observed by Yo and Wilson (1992) which is based on specific plaid features, "avgL", "minL" and "maxL" which she introduces, and which clearly relate to the blob features that we have defined in Section 2.

In Figure 6 of Bowns (1996), these features and their motion are illustrated for a plaid in which the directions of motion of the two component gratings differ by 10° (directions of 90° and 100°). According to our analysis, the blobs in this plaid, which appear to correspond approximately in shape to the maxL feature, have an edge ratio of 1:0.09, i.e. the blobs are highly elongated, and the longer edges move in an orthogonal

direction of 95°, almost exactly equal to the vector sum direction of 93°. For a component grating speed ratio of 1:0.5, our model gives for this plaid an initial direction estimate of 90°and a final direction estimate of 65°, i.e. 25° to the right hand (IOC = 19°) side of the vertical, implying that in a forced choice decision of left or right of the vertical, as in the Bowns (1996) experiments, a consistent IOC choice would be likely. At this point we refer the reader back to our description and simulations of Bowns' 1996 experiments in Section 3.1. A different explanation is however given in Bowns (1996) for consistent IOC result; namely that, as stated in the legend to Figure 6 "there are no edges that move in the vector sum direction for this plaid". Hence it is concluded that the choice will always be in the IOC direction.

In Figure 7 of Bowns (1996), avgL, maxL and minL are again illustrated for a plaid in which the directions of motion of the two component gratings differ by 80° (directions of motion of 90° and 170°). In the legend to the Figure it is stated again that "there are no edges that move in the vector sum direction for this plaid". However, for this plaid our analysis shows that the blobs are not elongated, having an edge ratio of 1:0.84, which would predict a velocity estimate close to the IOC direction. Also for this a plaid, the IOC direction (108°) is close to the vector sum direction (114°), and both are thus to the left of the vertical. Our model gives initial and final velocity estimates for this plaid which are both approximately equal to the IOC direction, thus predicting, in a forced choice of left or right of the vertical, a decision of left (vector sum), corresponding to the outcome in the actual experiment, as indicated in the legend to Figure 7 of Bowns (1996).

Finally, in Figure 8 of Bowns (1996), avgL, maxL and minL are illustrated for a plaid in which the directions of motion of the two component gratings differ by 40°

(directions of motion of 90° and 130°). In this case, the legend to Figure 8 indicates that whilst neither of the features maxL or minL have edges moving in the vector sum direction, avgL has an edge which moves in this direction. The inference is made that the presence of this motion resulted in subjects performing variably with this plaid, one perceiving it in the IOC direction (right of the vertical) and one in the vector sum direction (left of the vertical). From our analysis, for this plaid, and a speed ratio of 1:0.5, the blob edge ratio is 1:0.36, i.e. the blobs are somewhat elongated, and their long edges move in an orthogonal direction of 110°, close to the vector sum direction of 103°. Our model estimates a plaid velocity direction in the first step of the estimation algorithm of 95° (5° to the left of the vertical) and a final estimate of 73° (17° to the right of the vertical). The IOC direction is 67°.

{Figure 7 about here}

The results from our simulations of the full range of plaids used in Experiment 2 of Bowns (1996), of which those discussed above are a subset, are shown in Figure 7. The plots in Figure 7 show the initial and final estimated plaid directions as a function of the angular difference between the component grating directions for these plaids. The shaded area in the centre of the graph indicates the range of component grating angular differences which resulted in an inconsistent choice by subjects between "vector sum direction" and "IOC direction" for the corresponding plaids. From these results, and our discussion above, we suggest that the reason for the observed variability between subjects in their choice of IOC or vector sum direction (Bowns, 1996) lies in the variability of subjects in terms of the dependence of their direction perception on the duration of the stimulus. As Yo and Wilson (1992) showed, subjects can display considerable

differences in this dependence. In Figure 6 of Yo and Wilson, one subject (HRW) reported a direction bias of 30° after ~90 msec. stimulus duration, from an initial bias of ~60° at 60 msec. Another subject (HJ) reported a direction bias of 15° after ~90 msec., from approximately the same initial bias at 60 msec. Significantly, the stimulus duration used in the Bowns (1996) experiments was 80 msec, which would imply that a significant variation in perceived bias between subjects at this duration was possible. A similar variability to that reported by Yo and Wilson (1992) would therefore probably be sufficient to cause the difference in direction choice between the two subjects in the Bowns (1996) experiments.

Whilst our explanation contrasts with that of Bowns (1996), her explanation does clearly indicate that there is present in the plaid pattern both motion in the vector sum direction (in our analysis the orthogonal 1-D motion of the longer edges of the blobs) and in the IOC direction (in our analysis the 2-D motion of the blob end-points). She uses this fact to propose that the variation between subjects may result from a competition between these two sets of motion information. Our analysis suggests that a Bayesian process which using both information sets can also predict this result.

Another analysis of Type II plaid misperception based of the motion of features in the plaid was presented in Bowns (2006). Here a squaring operation is performed on the plaid and two "components" are identified: "sqHF" and "sqLF" which are derived from the squared plaid. The description in Bowns (2006) shows that the "components" are in fact two gratings formed from the squared plaid pattern, a high spatial frequency grating and a low spatial frequency grating, with spatial frequencies and orientations defined in the Appendix. They clearly relate to the product gratings, and have the same orientations

and direction of motion as these, as illustrated in Figure 1(d) of Bowns (2006). Examples of the values for the direction of motion of the sqHF and sqLF are also given for three Type II plaids, which were also used in Bowns (1996), showing that the direction of motion of the sqHF "component" is close to the vector sum direction. This led to the proposal that the direction of motion of the "components" provided a better overall predictor of the misperceived direction of these plaids than either the vector sum, as suggested by Yo and Wilson (1992), or the IOC direction, as suggested by Adelson and Movshon (1982). We clearly concur with this conclusion, as our predictions based on the motion of the blob edges show. However, Bowns also suggests that there is no motion energy in the plaids in the IOC direction, so that a full explanation of the misperception would "a model that incorporates both squaring and the IOC". Our model however incorporates both the 1-D motion of the blob edges, which contain motion energy close to the vector sum direction, and the 2-D motion of the blob end-points, which contain motion energy in the IOC direction. Used together in a recursive Bayes estimation framework, we have shown that this model closely predicts a wide range of results on perceived direction of plaid motion.

In addition to providing plausible explanations for a wide range of existing psychophysical and physiological results, new directions for experimental investigation are suggested by our model, including monitoring the response to Type II plaid motion of end-stopped cells (Hubel and Weisel, 1965; Pack et al, 2003) in layer 4b of area V1, the layer which contains the majority of V1 neurons projecting to MT. We predict that such experiments will indicate that these neurons signal the 2-D motion of the high luminance regions in the plaid, for Type II plaids. Additionally, studies of the dynamic response of

MT neurons to Type II plaids have, as far as we are aware, not been done, although a stimulus consisting of a field of short bright bars (Lorenceau, Shiffrar, Wells, and Castet, 1993) mimics the high luminance regions in Type II plaids. For this bar-field stimulus, Pack and Born (2001) showed that MT neurons initially respond primarily to the component of motion perpendicular to a contour's orientation, but over a period of approximately 60 ms the responses gradually shift to encode the true stimulus direction, regardless of orientation. Thus the responses of the MT cells closely parallel the psychophysical responses of human observers to the motion of Type II plaids (Yo and Wilson, 1992). Similar studies in which the responses of MT neurons are selectively inhibited, by lesioning or reversibly cooling (Hupé, James, Payne, Lomber, Girard and Bullier, 1998; Supèr and Lamme, 2007) might also be able to test our hypothesis that the local 1-D and 2-D local motion signals are combined to provide the perception of plaid motion via a recursive estimation process, which we hypothesise is implemented in the recurrent interaction between V1 and MT, an interaction which has been strongly implicated in the perceptual awareness of visual motion (Lamme and Roelfsema, 2000; Sterzer, Haynes and Rees, 2006)

References

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

Alais, D.M., Wenderoth, P.M., & Burke, D.C. (1994). The contribution of 1-D motion mechanisms to the perceived direction of drifting plaids and their aftereffects. Vision Res. 34, 1823-1834.

Alais, D.M., Wenderoth, P.M., & Burke, D.C. (1997). The size and number of plaid blobs mediate the misperception of Type-II plaid direction. Vision Res. 37, 143-150.

Bowns, L. (1996). Evidence for a feature tracking explanation of why Type II plaids move in the vector sum direction at short durations. Vision Res. 36, 3685-3694.

Bowns, L. (2006). `Squaring' is better at predicting plaid motion
than the vector average or intersection of constraints. Perception 35, 469-481

Burke, D., & Wenderoth, P. (1993). The effect of interactions between one dimensional component gratings on two dimensional motion perception. Vision Res. 33, 343-350.*Champion, R. A., Hammett, S. T., & Thompson, P. G. (2007). Perceived direction of* plaid motion is not predicted by component speeds. Vision Res. 47, 375-383.

Derrington, A. M., & Badcock, D. R. (1992). Two-stage analysis of the motion of 2-dimensional patterns: what is the first stage? Vision Res. 32, 691–698.

Derrington, A. M., & Suero, M. (1991). Motion of complex patterns is computed from the perceived motions of their components. Vision Res. 31, 139-149.

Dimova, K., & Denham, M. (2009). A neurally plausible model of motion integration in smooth eye pursuit based on recursive Bayesian estimation. Biol. Cybern. 100, 185-201.

Fennema, C.L., & Thompson, W.B. (1979).Velocity determination in scenes containing several moving objects. Comp. Graph. Image Proc. 9, 301-315.

Ferrera, V., & Wilson, H. (1990).Perceived direction of moving two-dimensional patterns. Vision Res. 30, 273–287.

Ferrera, V., & Wilson, H. (1991).Perceived speed of moving two-dimensional patterns. Vision Res. 31, 877–893.

Hubel, D. H., & Weisel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Neurophysiol. 28, 229-289.

Hupé, J.M., James A.C., Payne B.R., Lomber S.G., Girard P., & Bullier J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394, 784-787.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. 82, 35-45.

Lamme, V.A., & Roelfsema P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. 23, 571-579.

Lorenceau, J., Shiffrar, M., Wells, N. & Castet, E. (1993). Different motion sensitive units are involved in recovering the direction of moving lines. Vision Res. 33, 1207-1217

Marr, D., & Ullman, S. (1981). Directional selectivity and its use in early visual processing. Proc. R. Soc. Lond. B Biol. Sci. 211, 151–180.

Masson, G. S., & Stone, L. S. (2002). From following edges to pursuing objects. J. Neurophysiol. 88, 2869–2873.

Movshon J.A., Adelson E.H., Gizzi M.S., & Newsome W.T. (1985). The analysis of moving visual patterns. In: C. Chagas, R. Gattass, & C. Gross (Eds.), Pattern recognition mechanisms (pp. 117–151). New York: Springer.

Movshon, J. A., & Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. Vis. Neurosci. 16, 7733–7741.

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, 1040-1042.

Pack, C. C., & Born, R. T. (2008). Cortical mechanisms for the integration of visual motion. In: A. I. Basbaum, Akimichi Kaneko, G. M. Shepherd, & G. Westheimer, (Eds.). The Senses: A Comprehensive Reference, Vol 2, Vision II (pp. 189-218), Thomas D. Albright and Richard Masland: Academic Press.

Pack, C.C., Livingstone, M.S., Duffy, K.R., & Born, R.T. (2003). End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. Neuron 39, 671–680.

Smith, A.T., Singh, K.D., Williams, A.L. & Greenlee, M.W. (2001). Estimating receptive field Size from fMRI Data in human striate and extrastriate visual cortex. Cerebral Cortex, 11, 1182-1190.

Sterzer, P., Haynes, J.D., & Rees, G. (2006) Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT+/V5. Neuroimage 32, 1308-1316.

Stone, L., Watson, A. & Mulligan, J. (1990). Effect of contrast on the perceived direction of a moving plaid. Vision Res. 30, 1049-1067.

Supèr, H., & Lamme, V.A. (2007). Altered figure-ground perception in monkeys with an extra-striate lesion. Neuropsychologia. 45, 3329-3334.

Tinsley, C. J., Webb, B. S., Barraclough, N. E., Vincent C. J., Parker, A., & Derrington, A. M. (2003). The nature of V1 neural responses to 2D moving patterns depends on receptive field structure in the marmoset monkey. J. Neurophysiol. 90, 930–937.

Ullman, S. (1979). The Interpretation of Visual Motion. Cambridge: MIT Press.

Wallace, J. M., Stone, L. S., & Masson, G. S. (2005). Object motion computation for the initiation of smooth pursuit eye movements in humans. J. Neurophysiol. 93, 2279-2293.

Wallach, H. Ueber visuell whargenommene bewegungrichtung. (1935). Psychol. Forsch. 20, 325–380.

Wenderoth, P.M., Alais, D.M., Burke, D.C., & van der Zwan, R. (1994). The role of "blobs" in determining the perception of drifting plaids and motion aftereffects. Perception 23, 1163-1169.

Weiss, Y., & Adelson, E. H. (1998). Slow and Smooth: a Bayesian theory for the combination of local motion signals in human vision. Massachusetts Institute of Technology, Dept. of Brain and Cognitive Sciences, A.I. Memo. No. 1624.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. Nat. Neurosci. 5, 598 - 604.

Welch, L. (1989). The perception of moving plaids reveals two processing stages. Nature 337, 734–736.

Wilson, H.R., Ferrera, V.P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. Visual Neuroscience 9, 79-97

Wilson, H.R., & Kim, J. (1994). A model for motion coherence and transparency. Visual Neuroscience 11, 1205-1220.

Wuerger, S., Shapley, R., & Rubin, N. (1996). "On the visually perceived direction of motion" by Hans Wallach: 60 years later. Perception 25, 1317–1367.

Yo, C. & Wilson, H. Perceived direction of moving two-dimensional patterns depends on duration, contrast, and eccentricity. Vision Res. 32, 135–147 (1992).

Figure Legends

Figure 1. Three examples (a-c) of the representation of a plaid (centre) as the sum (left) or the product (right) of two gratings. The velocity space diagram above each plaid shows the velocity vectors for each component grating, v_1 and v_2 together with the IOC and vector sum velocity vectors v_{IOC} and v_{VS}. The arrows on the gratings and plaids also show their directions of motion, with the dashed arrow on the plaid showing the vector sum direction.

Figure 2. a-c: Simulations of the computational model for three cases of Type 2 plaids used in the experiments of Yo and Wilson (1992):

a. $\theta_1 = 70.5^\circ, \theta_2 = 48.2^\circ, v_1 = 1.33, v_2 = 2.67, \theta_{IOC} = -0.2^\circ, \theta_{VA} = 55.6^\circ$;

b. $\theta_1 = 84.3^\circ, \theta_2 = 36.9^\circ, v_1 = 0.25, v_2 = 2, \theta_{IOC} = 0^\circ, \theta_{VA} = 41.7^\circ$;

c. $\theta_1 = 85.2^\circ, \theta_2 = 33.6^\circ, v_1 = 0.4, v_2 = 4, \theta_{IOC} = 0^\circ, \theta_{VA} = 37.8^\circ$.

and for three different stimulus contrast levels. The results illustrate the dependence on stimulus contrast of the initial and final estimates of plaid direction, and of the convergence rate of the estimate towards the true IOC plaid direction. Presentation duration is represented by the number of iterations of the algorithm. d-e: Vector space diagrams showing the model simulation results for Experiment 3 of Bowns (1996). The initial v_{el} and final v_{es} plaid velocity estimates from our model are shown together with the component grating velocity vectors v_1 and v_2, and the vector sum and IOC velocity vectors.

Figure 3. Results from simulations of the computational model for the plaids used in the experiments of Burke and Wenderoth (1993), showing in a and b both the perceived plaid direction obtained in the experimental study (▲symbols) and the plaid direction

estimated by the model (• symbols). In Figure 3a the mean component direction is 295°, and in Figure 3b this direction varies and is shown by the dashed line. The graphs a and b show that the plaid direction estimated by the model varies with the difference in component grating direction and displays in both cases the same trend in the variation as observed in the Burke and Wenderoth (1993) study, although with a slightly greater bias towards the mean component direction of up to 9°. From the original diagram for the experimental results in this study, perceived errors were in the region of ±3°.

Figure 4. Velocity space diagrams of the plaid used in the experiments of Yo and Wilson (1992) and for which the model simulation results are shown in Figure 2a. a. diagram illustrating the plaid, clearly showing the elongated blobs; b. velocity space diagram on which the velocity vectors, v_1 and v_2, of the component gratings, together with their constraint lines, the IOC and vector sum velocity vectors v_{IOC} and v_{VS}, and the velocity estimate from our model for the first step, v_{e1}, are shown (0° is vertically upward in this diagram). The inset diagram. is redrawn from Figure 15d of Weiss and Adelson (1998); c. velocity space diagram showing the velocity vectors corresponding to the motion of the longer and shorter edges of the blobs in the plaid, v_ϕ and v_φ, together with their constraint lines, the IOC and vector sum velocity vectors, and the velocity estimates from our model for the first step, v_{e1}, the fifth step v_{e5}, and in the steady-state v_{es}. An explanation of the diagrams is given in the text.

Figure 5. The two plaids used the model simulations of the experiments of Stone et al (1990), showing the effect of the spatial frequency of the component gratings on blob size and number. Both plaids correspond to an angular separation of component gratings of 120°, 60° either side of the vertical (0°) and for each the contrasts of the component gratings are in the ratio of 4:1. For the plaid in a. the spatial frequency of the component gratings is twice that for the plaid on the right.

Figure 6. Results from simulations of the computational model for the plaids used in the experiments of Alais et al (1997) showing the perceived plaid direction as a function of spatial frequency obtained in the experimental study, for the 3° aperture case (▲symbols) and the plaid direction estimated by the model (● symbols). The estimates of the plaid direction are very similar to the perceived values (within 4°) and show the same trend, with a decrease in the misperceived direction bias as the spatial frequency of the component gratings increases. From the original diagram for the experimental results in this study, perceived errors were in the region of ±2°.

Figure 7. Results from simulations of the computational model for the plaids used in Experiment 2 of Bowns (1996), showing the initial and final estimated plaid directions as a function of the angular difference between the component grating directions for these plaids. The shaded area in the centre of the graph indicates those component grating angular differences which resulted in an inconsistent choice by subjects between "vector sum direction" and "IOC direction" for the corresponding plaids (see the text in the Discussion section for a further discussion of these results).

Figure 1

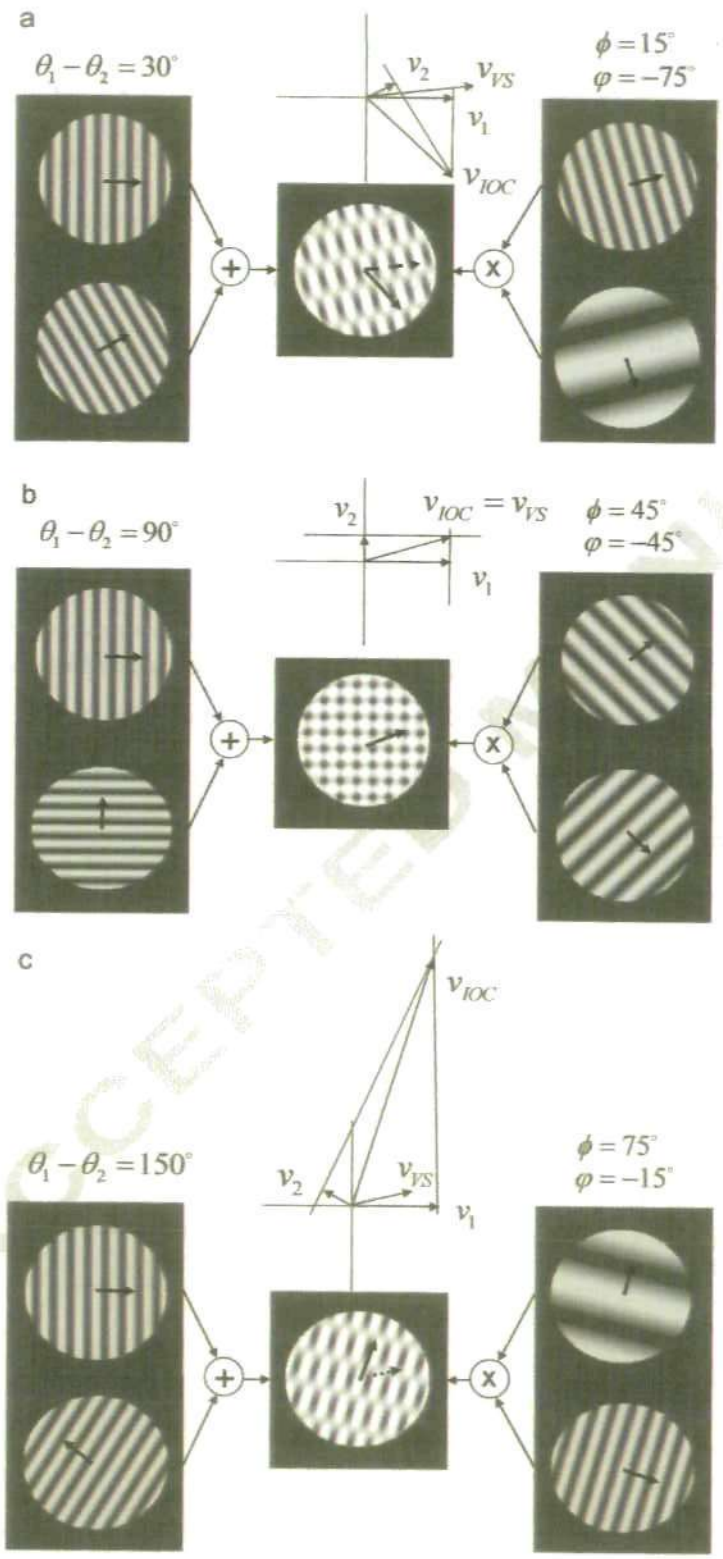

a
$\theta_1 - \theta_2 = 30^\circ$
v_2
v_{VS}
v_1
v_{IOC}
$\phi = 15^\circ$
$\varphi = -75^\circ$
+
x
b
$\theta_1 - \theta_2 = 90^\circ$
v_2
$v_{IOC} = v_{VS}$
v_1
$\phi = 45^\circ$
$\varphi = -45^\circ$
+
x
c
v_{IOC}
$\theta_1 - \theta_2 = 150^\circ$
v_2
v_{VS}
v_1
$\phi = 75^\circ$
$\varphi = -15^\circ$
+
x

Figure 2

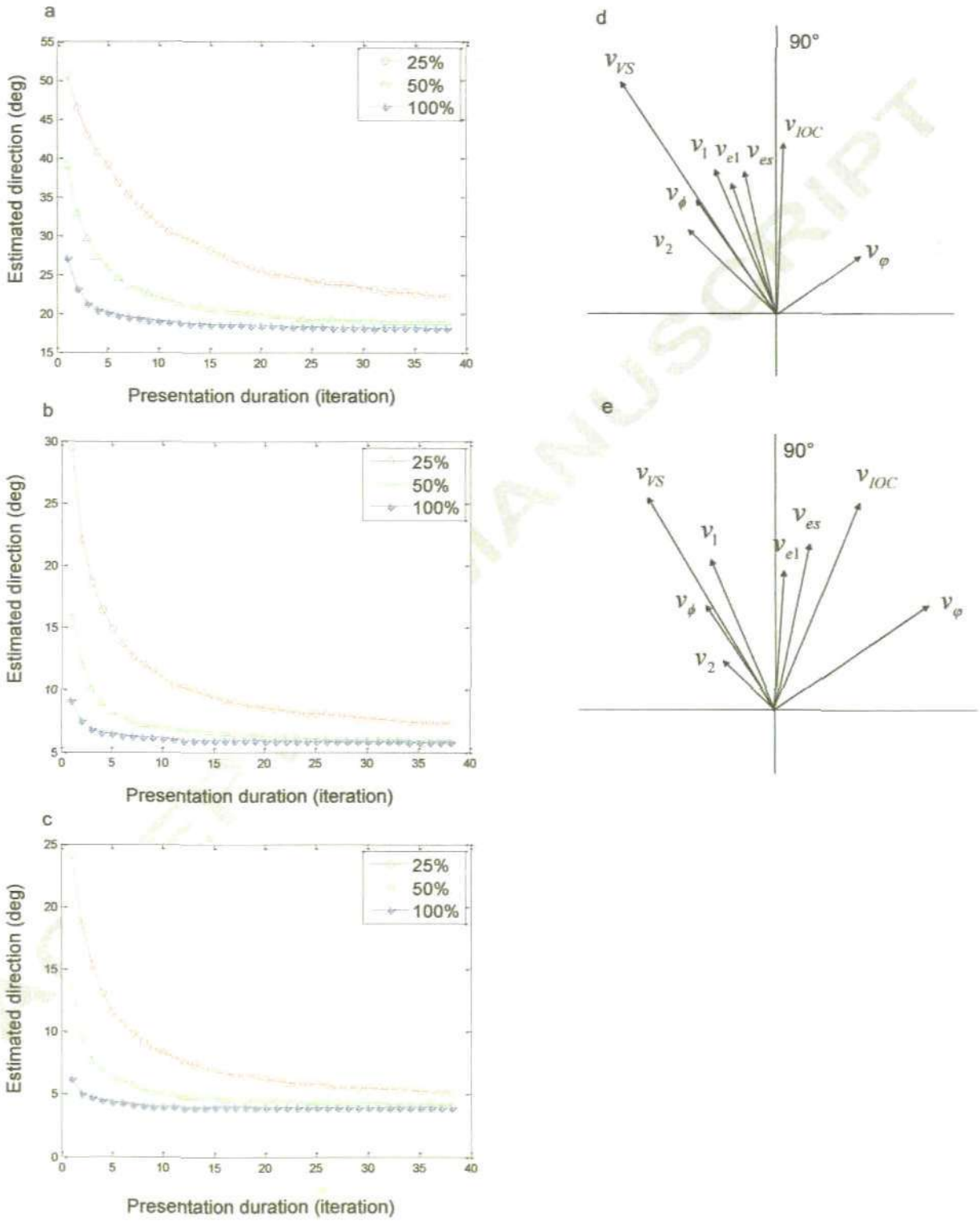


Figure 3

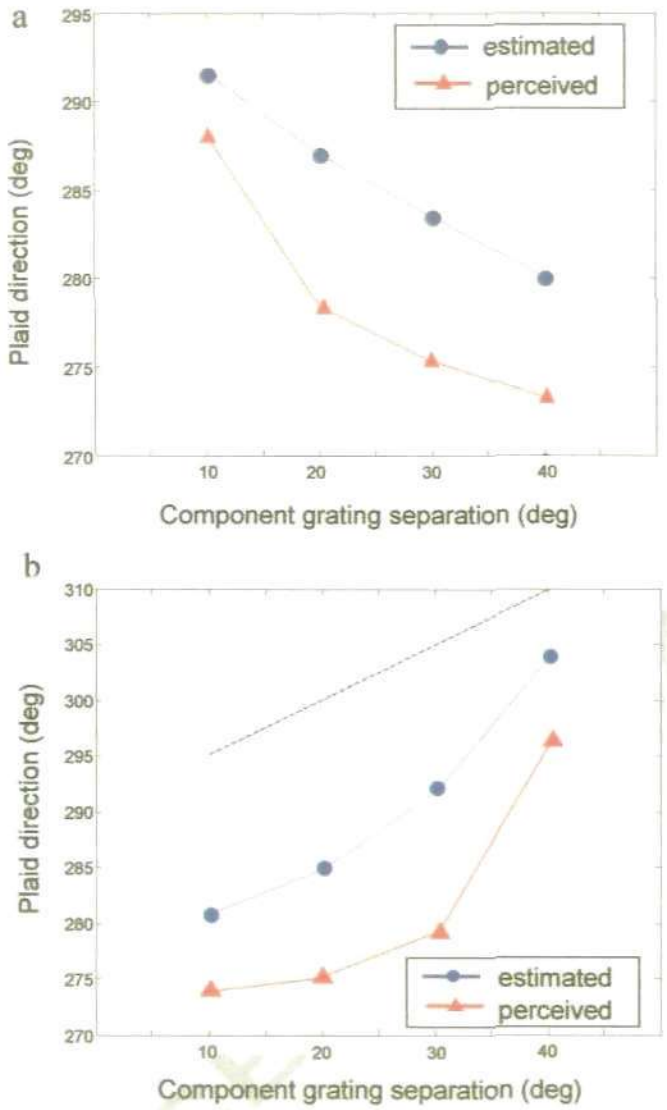
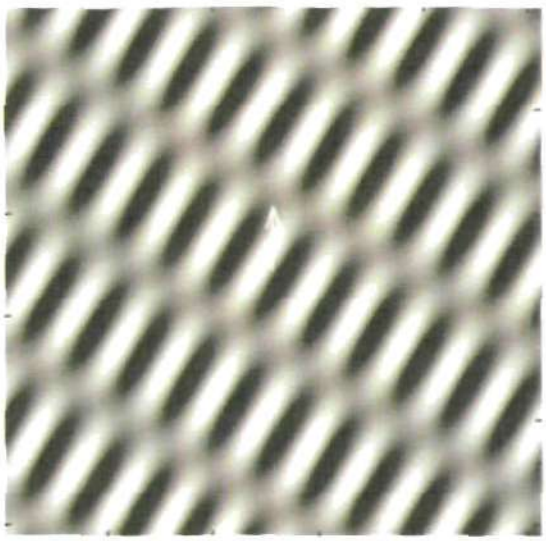
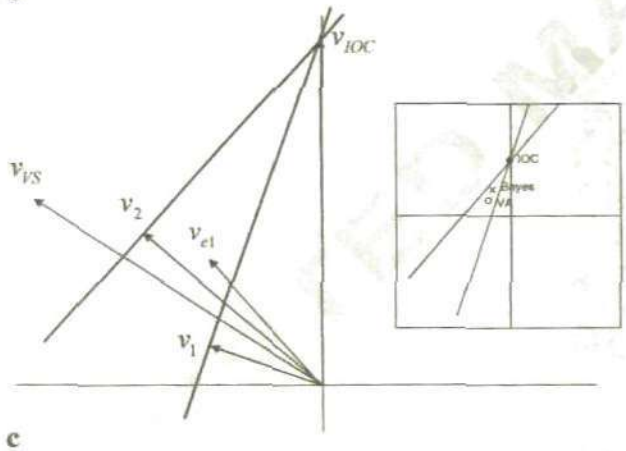


Figure 4

a



b


v_{IOC}
v_{VS}
v_2
v_{e1}
v_1
IOC
Bayes
NA

c

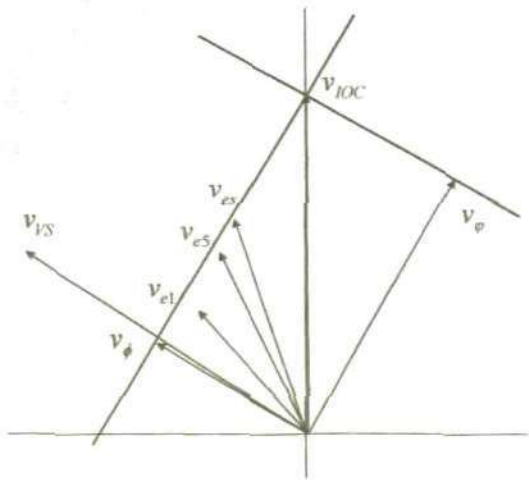
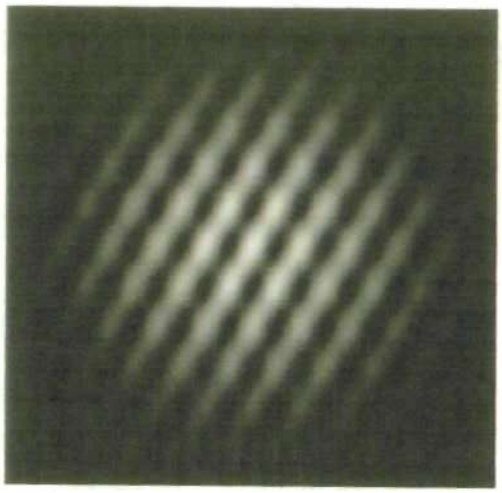

v_{IOC}
v_{es}
v_{VS}
v_{e5}
v_{e1}
v_ϕ
v_φ

Figure 5

a



b

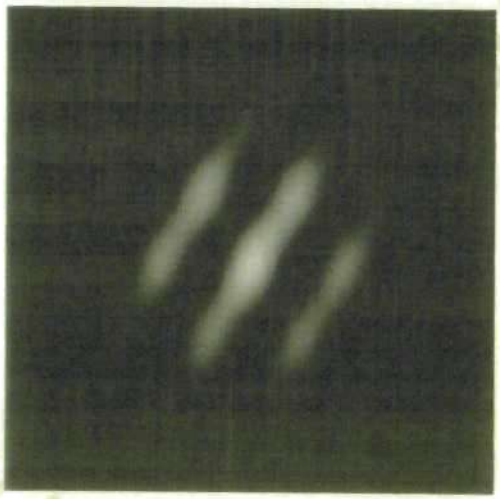


Figure 6