The spatial resolutions of stereo and motion perception and their neural basis

Fredrik Allenmark

Doctor of Philosophy

Institute of Neuroscience

Newcastle University

December 2011

Abstract

Depth perception requires finding matching features between the two eye's images to estimate binocular disparity. This process has been successfully modelled using local cross-correlation. The model is based on the known physiology of primary visual cortex (V1) and has explained many aspects of stereo vision including why spatial stereoresolution is low compared to the resolution for luminance patterns, suggesting that the limit on spatial stereoresolution is set in V1. We predicted that this model would perform better at detecting square-wave disparity gratings, consisting of regions of locally constant disparity, than sine-waves which are slanted almost everywhere. We confirmed this through computational modelling and performed psychophysical experiments to test whether human performance followed the predictions of the model. We found that humans perform equally well with both waveforms. This contradicted the model's predictions raising the question of whether spatial stereoresolution may not be limited in V1 after all or whether changing the model to include more of the known physiology may make it consistent with human performance. We incorporated the known size-disparity correlation into the model, giving disparity detectors with larger preferred disparities larger correlation windows, and found that this modified model explained the new human results. This provides further evidence that spatial stereoresolution is limited in V1. Based on previous evidence that MT neurons respond well to transparent motion in different depth planes we predicted that the spatial resolution of joint motion/disparity perception would be limited by the significantly larger MT receptive field sizes and therefore be much lower than the resolution for pure disparity. We tested this using a new joint motion/disparity grating, designed to require the detection of conjunctions between motion and disparity. We found little difference between the resolutions for disparity and joint gratings, contradicting our predictions and suggesting that a different area than MT was used.

Acknowledgments

First I would like to thank my supervisor Jenny Read for giving me the opportunity to pursue a PhD and for all her invaluable support, encouragement and guidance.

I would also like to thank Ignacio Serrano-Pedraza, a former lab member, for all his kind help and advice and for our many stimulating scientific discussions.

Thank you, as well, to all the participants in my experiments who kindly volunteered their time.

Slutligen vill jag tacka mina föräldrar för deras stöd och uppmuntran under de senaste tre åren.

Table of Contents

Chapter 1. Introduction

This thesis is about a the spatial resolutions of the human abilities to see in depth and to perceive motion and about modelling what happens in early areas of the visual system in order to understand what it is that sets the limit of these resolutions. The visual system is generally thought of as having a hierarchical structure with lower and higher visual areas and where receptive field sizes grow larger when moving higher up in the hierarchy. The hierarchy starts with the retina which projects mainly to the lateral geniculate nucleus (LGN) but also to the pretectum which is important for papillary reflexes and to the superior colliculus which has a role in the guidance of eye movements. The LGN serves as a relay between the retina and primary visual cortex (V1). A property of LGN cells that is relevant to the topic of this thesis is that each individual LGN cells only receives input from one eye. V1 is the first visual area where binocular cells, that is cells that receive information from both eyes, are found. That is why area V1 is of particular interest in the study of spatial stereoresolution, the resolution with which we can perceive patterns of variation in depth. An important hypothesis splits the visual system into two visual streams, the dorsal and the ventral streams. These streams are sometimes called the "where" and "what" streams since the dorsal stream has been believed to be more involved in representation of object locations and the ventral stream more with form recognition and object representation. Area V2 is part of both the dorsal and the ventral stream. In depth perception this area is thought to have a role in perception of relative depth, but it also has other functions. After area V2 the ventral stream includes area V4 and the inferotemporal cortex which is thought to have a role in object recognition. The dorsal stream includes area MT which receives projections both directly from V1 and from the dorsal part of V2. Cells in area MT are strongly sensitive to object motion and also to depth. Cells in area V1 are also sensitive to motion but differ from area MT cells in that when motion in two different directions are combined in such a way that we perceive a total motion in a third direction V1 cells are only selective for the component directions while area MT cells are selective for the combined direction of motion that we perceive. Area MT projects to area MST which is involved in the perception of optic flow, i.e. the changes in the visual image that

occur as a consequence of our motion through the environment. Since higher areas than MT have such a more complex and specialized role in motion processing and area V1 cells in general responds inconsistently with human perception to combined motion consisting of component motions in two or more directions, area MT seems like an ideal candidate for a basic motion processing area. In Chapter 4 we will consider the hypothesis that area MT may be where the limit on motion resolution and joint depth/motion resolution is set.

Because of their difference in position the two eyes receive slightly different images of the surrounding world. The visual system is able to estimate the distances to different objects that we are looking at from these differences (binocular disparities). This process is comparatively well understood in terms of knowledge about the underlying physiology, computational models and data from psychophysical experiments. Therefore, in the field of stereovision, we are in a particularly good position to compare computational models to human performance. In particular models that use local cross-correlation to find matches between the two eye's images have been successfully used to explain the low human spatial stereoresolution for sinusoidal disparity gratings (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). This class of models is based on the known physiology of primary visual cortex (Nienborg, Bridge, Parker and Cumming 2004) and in particular the size of the correlation window corresponds to the receptive field size of V1 cells. The use of windowed cross-correlation implies that the model has an initial encoding of disparity as a set of frontoparallel patches. As a consequence these models would be predicted to perform better at detecting square-wave disparity gratings, which are built from segments of constant disparity than sine-waves.

Motivated by the prediction made in the above paragraph, the first subproject presented in this thesis looked at comparing the detectability of sinusoidal and square-wave disparity gratings, for the local cross-correlation model and in human subjects. It was found that the model performed better for the square-waves than for the sine-waves at high disparity amplitudes, in contrast to the human subjects who performed equally well with both wave-forms at all frequencies and amplitudes. This presented a challenge to the local cross-correlation model and to the idea that spatial stereoresolution is set in primary visual cortex. Before trying to introduce processing done at later stages in the visual system into the

model however it seemed natural to first try to include more of the known physiological detail of V1 cells. In particular the known size-disparity correlation (Tyler 1973; Smallman and MacLeod 1994; Prince, Cumming and Parker 2002) seemed a promising candidate for an addition to the model that might help reduce the difference between performance on square-waves and sine-waves, since introducing a size-disparity correlation into the model would make the high amplitude square-waves on average be detected by cells with larger receptive field sizes compared to sine-waves with the same amplitude and this would be expected lead to a reduction of performance on high frequency high amplitude square-waves compared to sine-waves with the same frequency and amplitude.

Motivated by this intuition the next subproject looked at a local cross-correlation model incorporating a size-disparity correlation and tested this model with the same stimulus and task used in the first subproject. The modified model performed consistently with the human results found in the first subproject. This provided further evidence that spatial stereoresolution is limited in area V1. The modified model also performed consistently with human results on the frequency dependence of the upper depth limit (Tyler 1973). This suggests that the disparity gradient limit may be a consequence of the size-disparity correlation as was originally suggested by Tyler (1973).

The final subproject was motivated by a prediction made based on previous evidence that cells in MT respond well to transparent motion in different depth planes (Bradley, Qian and Andersen 1995) and the much larger receptive field sizes in MT compared to V1 (Gattass and Gross 1981). The prediction was that the spatial resolution for perception of joint motion/disparity perception should limited by the large MT receptive field sizes and therefore be much worse that the resolution for pure disparity perception which is thought to be limited by V1 receptive field sizes. This hypothesis was tested in psychophysical experiments using pure disparity gratings, pure motion gratings and joint motion/disparity gratings, designed to require the detection of conjunctions between motion and disparity. The results supported at most a much smaller difference in receptive field size between cells used for joint motion/disparity perception and cells used for pure disparity perception than what would be predicted based on the difference in receptive field sizes between V1 and MT. The receptive fields for cells used in motion perception were found to be

significantly smaller than for either the pure disparity or the joint motion disparity, meaning that the pure motion task was also unlikely to be performed by area MT.

1.1 Thesis structure

Section 1.2 reviews previous literature on stereo vision and motion perception.

Chapter 2 describes the psychophysical experiments and modelling using the original cross-correlation model designed to test the hypothesis that square-wave disparity gratings should be easier to detect than sine-wave gratings.

Chapter 3 describes the modelling using the modified cross-correlation model with the size-disparity correlation.

Chapter 4 describes the experiments designed to compare the spatial resolution pure disparity gratings, pure motion gratings and joint motion/disparity gratings.

Chapter 5 contains conclusions, a discussion of limitations of the work presented here and ideas on how it could be taken further.

1.2 Literature review

1.2.1 *Psychophysics*

In this section some important research on human visual perception will be reviewed with the focus on depth perception and in particular those features of depth perception that will play an important role in this thesis.

1.2.1.1 Frequency analysis

Campbell and Robson (1968) provided evidence that, for contrast gratings, there exist independent channels sensitive to different ranges of spatial frequencies. They measured the contrast at which subjects found a contrast grating to be barely detectable for sine-wave and square-wave contrast gratings with a range of frequencies. They found that, for frequencies higher than 0.8 cycles/degree, the ratios between the detection thresholds for

the square-waves and those for the sine-waves were all close to the value of $4/\pi$ which would be predicted if the visibility of a square-wave grating depended primarily on the amplitude of the first harmonic. They also measured the lowest contrast at which square-wave gratings first became distinguishable from sine-wave gratings and found that this happened when the third harmonic of the square-waves reached their own threshold. Graham and Nachmias (1971) provided further evidence for the existence of multiple frequency channels. They measured the lower contrast threshold for detection of gratings with frequency f, gratings with frequency 3f and combinations of the two for values of f in the range from 0.9 to 6.3 cycles/degree. The combined gratings consisted of gratings with frequencies f and 3f added either with such a relative phase that the peaks of the f component were added to the peaks of the 3f component or with such a phase that the peaks were subtracted from each other. They found that for the combined gratings, the detection threshold was always close to the threshold of one of the component gratings, regardless of the relative phase used as well as the ratio between the contrasts of the f and 3f components.

The first results of a similar nature in the disparity domain were reported by Tyler (1975). Tyler demonstrated that the frequency of a disparity grating appears shifted after adaptation to another disparity grating with slightly different frequency, and that this effect only occurs if the adapting grating has a similar orientation to the test grating. He interpreted this as "evidence for channels tuned to stimulus size at the hypercyclopean level of processing, independently of any tuning prior to that level."

Schumer & Ganz (1979) used the same methodology as Graham and Nachmias to test whether there exists spatial frequency channels in the disparity domain. Like Graham & Nachmias they found that the results were independent of the relative phase used and that the thresholds for the combined gratings were close to the threshold of one of the component gratings most of the time. However, the thresholds for combined gratings where both components were close to their individual threshold were found to be somewhat lower than either of the individual thresholds. They argued that this may be explained in part by probability summation and in part by less than complete independence between the different frequency channels. They also did adaptation experiments were they had subjects

adapt to sine-wave disparity gratings of different frequencies and measured lower amplitude detection thresholds for sine-wave gratings of different frequencies before and after adaptation. They found that adaptation caused elevation of thresholds, where the elevation was highest close to the adapting frequency and fell off for larger and smaller frequencies. They argued that this may be seen as further evidence for multiple spatial frequency channels since if the same mechanisms were used to detect gratings of all different frequencies, then one would expect no frequency dependence of the threshold elevation caused by adaptation. The curves of elevation as a function of frequency had half-amplitude bandwidth in the range from 2 to 3 octaves.

Cobo-Lewis and Yeh (1994) provided further evidence for channels tuned to different spatial frequencies of disparity corrugations. They measured detection thresholds for sine-wave corrugations in random dot stimuli with added notched noise maskers, where the frequency of the signal was logarithmically centred between two intervals of narrowband filtered noise. They found that thresholds improved with increasing distance between the signal's frequency and the frequencies of the masking noise for several different signal frequencies. They found masking curves in the same bandwidth range reported by Schumer & Ganz (1979), but also narrower ones, with the narrowest being 1.1 octaves. They argued that this difference could be explained if the adaptation effect was nonlinear, in such a way that threshold increases slower than linearly with increasing intensity of the adapting stimuli.

More recently, Grove and Regan (2002) measured frequency discrimination thresholds for sine-wave disparity gratings with and without adaptation. They found that adaptation to a sine-wave grating at the frequency of the test grating did not result in any increase in discrimination threshold, but that adaptation to gratings with frequencies offset in both directions did produce an increase. They argued that these results could be understood if spatial frequency channels at neighbouring frequencies to the one most closely tuned to the frequency of the stimulus are used in discrimination tasks. Neighbouring channels would be expected to be more sensitive to changes in frequency if the tuning curves of the channels are relatively flat near the preferred frequency.

1.2.1.2 Disparity gradient

Burt and Julesz (1980) reported that the largest disparity at which fusion is possible is determined by a disparity gradient limit. They used a stimulus made up of rows of repeated patterns consisting of two horizontally and vertically separated dot pairs with different disparities. The disparities were the same in each row, but the separation between dot pairs was decreased, thereby increasing the disparity gradient, with increasing row number. The subjects were asked to report at which row fusion was no longer possible. This was repeated with stimuli with different disparities and different orientations for the separation between dot pairs as well as for different viewing distances and subjects always reported being unable to fuse for disparity gradients larger than a limit close to one.

One field of study where a disparity gradient limit would be expected to have an effect is the study of stereo transparency, where many studies have used random dot stereograms where different dots in the same region belong to surfaces at different depths. Such stimuli necessarily contain local disparity gradients, which increase with increasing relative disparity between the different surfaces as well as with the density of the surfaces. Several studies using such stimuli have found that the perception of transparency is impaired by increasing the relative disparity as well as the dot density (Akerstrom and Todd 1988; Gepshtein and Cooperman 1998; Tsirlin, Allison and Wilcox 2008). One study found that increasing the relative disparity had no effect on transparent surface segregation (Wallace and Mamassian 2004) but Tsirlin et al. (2008) argued that the most likely cause of this is that the disparities and the density used in that study were lower than the ones used in the other studies and below the limit were performance starts to drop. These results could potentially be explained by a disparity gradient limit, although none of the studies mentioned specifically set out to test that hypothesis.

One study of stereo transparency which specifically addressed the question of what effect the disparity gradient limit has on the perception of transparency was performed by McKee and Verghese (2002). They asked test subjects to judge which of two dot patterns, presented in two separate time intervals, that contained a target consisting of four obliquely oriented dots hidden among pairs of noise dots with the same orientation as the target. The noise dots were located either in the same plane as the target dots, with the different pairs

distributed over two different depth planes on each side of the target or with the different dots in each pair placed in different depth planes allowing manipulation of disparity gradient without changing the relative disparity between the planes by changing the separation of dots within a pair. They found that separating the noise dots into different planes improved performance on the task and that the improvement decreased with increasing disparity gradient but remained significant up to disparity gradients of about four. They also tested how the perceived relative depth between the two planes depended on the disparity gradient, by asking subjects to compare to a reference, and found that, with the same relative disparity but increasing disparity gradient, the perceived relative depth decreased with increasing disparity gradient. They interpreted this as evidence that improvement of task performance with the introduction of different depth planes was dependent on the ability to clearly perceive transparency.

A challenge to the idea that there is a disparity gradient limit on the ability to perceive disparity corrugations came from Lankheet and Lennie (1996). They did experiments with dynamic RDS containing moving and static sine-gratings at different amplitudes and frequencies and with different levels of noise added to the disparities. They measured noise thresholds for detection of the gratings. For the static gratings they concluded that "Detection of binocular correlation depends on both spatial frequency and amplitude of disparity modulations, and cannot be reduced to a description in terms of gradient limits." However, the data of Lankheet and Lennie has received a different interpretation. Ziegler, Hess and Kingdom (2000) tested the ability of test subjects to discriminate between disparity gratings, based on random dot type stimuli with Gabor micropatterns instead of dots, at two different oblique orientations. They used sine-wave, square-wave and trapezoidal gratings. They argued that using trapezoidal gratings allowed them to vary frequency and disparity gradient independently by manipulating the ramp width. They measured upper disparity amplitude thresholds at a range of frequencies as well as ramp widths for the trapezoidal gratings. For the sine-waves they found similar results to Lankheet and Lennie (1996). They found that their data could be explained reasonably well by a model based on applying a disparity gradient limit after low-pass spatial filtering in the disparity domain.

1.2.1.3 Disparity gratings

There have been several studies using sinusoidal disparity gratings. Tyler (1974) was the first to use random dot stereograms of sinusoidal disparity gratings. Tyler used random dot stereograms containing horizontal sine-wave gratings of increasing frequency in the vertical direction and decreasing amplitude in the horizontal direction. Subjects were asked to indicate the borders where the grating appeared as a flat surface rather than a sine-wave. He found that the highest frequency at which the gratings could be perceived was around 4 cycles per degree and that the largest peak-to-peak disparities at which the grating could be perceived decreased with increasing frequency and could be reasonably well described by the equation $d_{max} = k/f$ where d_{max} is the peak-to-peak disparity, f is a frequency and k is a constant that differed between different subjects.

Bradshaw & Rogers (1999) showed that the lowest disparity amplitude at which sinusoidal disparity gratings can be perceived is lower for horizontal gratings than for vertical gratings at low spatial frequencies. They measured lower amplitude thresholds for sinusoidal disparity corrugations and found that the frequency where the lowest threshold was obtained was higher for vertical than for horizontal gratings and that thresholds were significantly higher for the vertical gratings for frequencies below 0.9 cpd. Serrano-Pedraza has shown that, at least with some subjects, this anisotropy appears when using sine-wave corrugations but not when using square-wave corrugations (Serrano-Pedraza and Read 2010).

Glennerster (1996) used square-wave disparity gratings in an experiment intended to test whether a coarse-to-fine algorithm or a co-operative algorithm provides a better model of human depth perception. The experiment involved comparing the exposure times necessary to perceive three different patterns in depth on a zero disparity or uncorrelated background, and it was argued that the two different algorithms make different predictions about what kind of pattern it should be easier to perceive. In particular the coarse-to-fine algorithm predicts that a high-frequency square-wave should require a longer exposure time than the two other pattern that had different average depth from the background. The predictions of the cooperative algorithm are instead based on smoothness which was the same for the square-wave and one of the other patterns, but different for the third. The results of this

experiment supported the coarse-to-fine algorithm. This was further supported another experiment were the frequency of the square-wave was varied and it was found that shorter exposures were required at lower frequencies. A study by Parker and Yang also concluded that a coarse-to-fine algorithm may be used in human stereo vision. They did experiments with random dot patterns with two planes in depth and examined the conditions under which one surface at the average disparity was perceived rather than two transparent surfaces. The inference that a coarse-to-fine strategy may be used was based on the finding that a larger disparity difference between the planes was required to see transparency rather than disparity averaging when the average disparity of the planes was offset from the plane of fixation. They took this to imply that coarser filters are used to detected larger disparities and that therefore "the neural apparatus is available to implement a coarse to fine strategy in stereo matching".

1.2.2 *Neuronal mechanisms*

In this section some of the most important research on the mechanisms responsible for depth perception and motion perception will be reviewed.

1.2.2.1 Energy Model

Ohzawa et al. (1990) measured the activity of cells in the visual cortex of cats while presenting dark and light bars at different positions on the two retinas and showed that a subset of complex cells are especially suited as binocular disparity detectors, having a fine disparity selectivity that is constant over the receptive field, and responding only to the correct contrast polarity. They developed a model of these cells where a complex cell received input from two pairs of simple cells, with a 90 phase shift between the pairs and where the output of each simple cell was half-wave rectified and squared before being summed together. This model was shown to produce similar results to real complex cells, when tested with bright and black bars in different positions. This so called energy model has been very influential, several computational models has been based on it and its properties have been subject to mathematical analysis (Qian 1994; Fleet, Wagner and Heeger 1996; Zhu and Qian 1996; Qian and Zhu 1997; Anzai, Ohzawa and Freeman 1999;

Tsai and Victor 2003). There has also been suggestions of minor changes to bring it closer to explaining all physiological data (Read, Parker and Cumming 2002; Haefner and Cumming 2008).

1.2.2.2 Cross-correlation based models

Banks, Gepshtein & Landy (2004) compared the predictions of a model of disparity detection, based on crosscorrelation between the images on the two retinas, to the results of psychophysical experiments with sinusoidal disparity modulations. They showed subjects random dot patterns containing sinusoidal corrugations at $\pm 20°$ orientation from horizontal and asked subjects to judge which orientation was being presented in each trial. They used an adaptive staircase procedure to obtain upper frequency thresholds. They found that, with disparity amplitude held constant, frequency thresholds rose with increasing dot density up to a maximum which depended on the disparity amplitude, and that the thresholds were close to the Nyquist limit up to a certain dot density where performance levelled off. They further found that performance levelled off at higher frequencies when the experiment was repeated with gratings with lower disparity amplitude. They interpreted this as evidence that resolution had been limited by the disparity gradient in their first experiment. They also explored the effect of optical blur by introducing extra blur using diffusing screens.

They found that performance levelled off at lower frequencies at higher levels of blur. They concluded that the luminance spatial frequency content of the stimulus is one factor that limits spatial stereoresolution, similar to what had been found in a previous study by Hess et al. (1999). They also presented the stimulus at different retinal eccentricities and found that performance levelled off at lower frequencies when the stimulus was presented at higher eccentricities. They argued that this was due to optical low pass spatial filtering and larger receptive fields in the periphery.

To test whether spatial stereoresolution was limited by the binocular matching process Banks, Gepshtein and Landy used an algorithm based on cross-correlation between the images on the two retinas to model binocular matching. They used the same images that were used in their psychophysics experiments, convolved them with the point-spread function of the well focused eye (Campbell and Gubisch 1966) to simulate optical blur and

computed windowed cross-correlation between the left and the right images. A square window was moved along a vertical line in the left eyes image and for each vertical position a window of the same size in the right eyes image was moved along all horizontal positions at the same vertical position and the cross-correlation of the contents of the window was computed for each combination of positions, resulting in a plot of correlation as a function of disparity and vertical position. The effects of window size, dot density and blur were examined, and it was found that resolution could be improved by decreasing the window size down to a certain limit that decreased with increasing dot density up to a certain limit determined by the level of blur. By testing the algorithm on corrugations with different disparity amplitudes it was found that the estimation of disparity became worse as the disparity gradient increased and that "The algorithm finds the highest correlations in the parts of the stimulus that are frontoparallel" (Banks, Gepshtein and Landy 2004). A similar result had been found in the physiological experiments of Nienborg et al. (2004) who measured the responses of V1 cells to sinusoidal disparity corrugations and found that the responses were as would be expected if the optimum disparity was constant across the receptive field of the cells.

The same model described above, except now using Gaussian windows, was compared further to human performance using a set of stimuli designed to test the effect of disparity gradient on performance (Banks, Gepshtein and Rose 2005; Filippini and Banks 2009). The same set of stimuli was used to test human subjects in psychophysics experiments and to test the model. The stimuli used were random dot stereograms of sawtooth disparity gratings of different frequency, phase, amplitude and parity (i.e. with the slats of the sawtooth waveform slanted either top-back or top-forward) containing different proportions of signal dots and noise dots. The task for both the human subjects and the model was to judge the parity of the gratings and for each combination of amplitude and frequency coherence thresholds were measured using the method of constant stimuli. The output of the model was correlated to a set of templates, and the model reported the parity to be that of the best matching template. The thresholds for both human subjects and the model was found to rise with increasing disparity gradient and the results for different frequencies overlapped reasonably well when plotting thresholds as a function of disparity gradient but not when plotting thresholds as a function of amplitude, indicating that disparity gradient and not

disparity amplitude was the main limiting factor on performance. The psychophysics experiments were repeated at two different viewing distances with similar results, showing that performance depended on the disparity gradient on the retina rather than the slant in the stimulus. The model was tested with three different window sizes with similar results except that thresholds were somewhat higher with the smallest window size. The main differences between the model and human performance were that the thresholds were somewhat lower in general for the model and that the rise of thresholds with disparity gradient was somewhat steeper for the human subjects. Banks et al. argued that the latter difference may be related to the fact that human disparity estimation is worse for large absolute disparities (Blakemore 1970), while the model had been designed to handle large disparities as well as it handles small ones.

Filippini and Banks (2009) also used the same model, with the Gaussian window and the template matching, to test quantitatively how well the model predicted human performance near the stereoresolution limit. The stimuli used were the same that were used by Banks et al. (2004). They found that the stereoresolution of the model was close to the Nyquist limit up to a limiting dot density where performance levelled off and that the highest attainable stereoresolution increased with decreasing window size. For large window sizes the highest attainable stereoresolution was found to be close to inversely proportional to the window size but it levelled off at smaller windowsizes. The highest stereoresolution that could be reached by decreasing the window size was higher for smaller levels of blur. The modelling data also showed that the dependence of stereoresolution on the level of blur was closest to that found in human data (Banks, Gepshtein and Landy 2004) when a window size of 6 arcmin was used.

1.2.2.3 Motion

Resolution for motion perception has also been studied and has been found to be slightly better than what is generally found for disparity (Anderson and Burr 1987; Georgeson and Scott-Samuel 2000) with an estimate of 2 arcmin for the smallest receptive field size of any motion detector unit (Anderson and Burr 1987; Anderson and Burr 1989). However, to the

authors' best knowledge, resolution for disparity and motion perception have never been studied and compared in the same subjects.

The Pulfrich effect (Morgan and Thompson 1975) is an illusion that has played an important role in the study of joint motion/disparity perception. The original Pulfrich effect can be demonstrated by viewing a pendulum swinging in a frontoparallel plane while introducing an interocular delay, for example by placing a dark filter in front of one eye. The pendulum then appears to follow an elliptical path in depth. The original Pulfrich effect can be explained as arising because the interocular delay introduces a disparity. This happens because the pendulum moves during the time occupied by the interocular delay and will therefore be at different positions in the images reaching the brain from the left and the right eyes.

Qian and Andersen modelled the integration of motion and disparity with a model based on the stereo energy model (Ohzawa, DeAngelis and Freeman 1990) and the motion energy model (Adelson and Bergen 1985) and showed that their model could account for the classical Pulfrich effect and a number of generalized Pulfrich phenomena (Qian and Andersen 1997). However, in the model of Qian and Andersen, individual model neurons encoded motion and disparity jointly, in the sense of having space-time inseparable receptive fields.

Disparity and motion are encoded separately in a large portion of V1 cells with only a small portion having joint encoding (Read and Cumming 2005b). Therefore, the Qian and Andersen model, and similar models that explain the Pulfrich effects based on joint encoding, assume that a large portion of all disparity-selective V1 neurons are ignored when viewing Pulfrich stimuli. Inspired by this, Read and Cumming showed that a model with separate encoding can also explain the Pulfrich phenomena with an appropriately chosen read-out rule. Qian and Freeman (2009) reproduced the results of Read and Cumming with a more physiologically detailed model and showed that, under the additional physiologically based assumptions made in this model, a population of cells that are tuned to a range of motions and a range of disparities combinatorially (referred to as joint encoding by Qian and Freeman) are required to explain the full range of Pulfrich

effects. The ability to detect correlation between disparity and motion has been studied before (Bradley, Chang and Andersen 1998) but not the resolution of this ability. In chapter 4 we study for the first time the resolution for detection of joint motion/disparity gratings.

Chapter 2. Detectability of sine- versus square-wave disparity gratings: a challenge for current models of depth perception

2.1 Introduction

Stereopsis, the ability to estimate 3D depth based on binocular vision, is one of the best understood aspects of human perception. 150 years of psychophysical experiments have documented in detail how binocular disparities between the eyes result in a depth percept (Howard and Rogers 1995), while in the last two decades, physiological experiments have mapped how disparities drive the firing rates of individual neurons in visual cortex (Roe, Parker, Born and DeAngelis 2007). Stereo vision has thus emerged as a paradigm for relating perceptual experience to neuronal activity.

A recent, highly successful example has been the development of a computational model explaining the spatial resolution of stereopsis in terms of the properties of neurons in primary visual cortex (Banks, Gepshtein and Landy 2004; Nienborg, Bridge, Parker et al. 2004; Filippini and Banks 2009). Stereo spatial resolution is traditionally assessed using sinusoidal "disparity gratings", corrugated surfaces which go back and forth in depth (Figure 1A). The upper frequency limit at which such disparity gratings can be perceived has been found to be around 3-4 cycles per degree (Tyler 1974; Bradshaw and Rogers 1999; Banks, Gepshtein and Landy 2004; Filippini and Banks 2009), much lower than the corresponding limit for luminance gratings, which can be as high as 50-60 cpd under optimal luminance conditions (Campbell and Green 1965). In a linked pair of papers (Banks, Gepshtein and Landy 2004; Nienborg, Bridge, Parker et al. 2004), Banks, Cumming and colleagues explained this limit in terms of the receptive field size of disparity-selective neurons in primary visual cortex (V1).

Their analysis was based on the stereo energy model, in which disparity is encoded by a local cross-correlation between the two eyes' images (Ohzawa, DeAngelis and Freeman 1990; Ohzawa, DeAngelis and Freeman 1997; Banks, Gepshtein and Landy 2004;

Nienborg, Bridge, Parker et al. 2004; Filippini and Banks 2009). In this model, interocular correlation is measured locally within a finite window corresponding to the neuronal receptive field, and the stereoresolution limit is determined by the smallest window-size available. When the frequency is high enough that the disparity changes significantly within this window, the effective interocular correlation is reduced and the signal is lost in the noise (compare Figure 2B to Figure 2A). It is this which eventually limits the ability to resolve the grating. Banks, Cumming and colleagues showed that the stereoresolution of human and monkey observers was remarkably consistent with the size of receptive fields in V1. Thus, the local cross-correlation model is a noteworthy example of how perceptual abilities can be successfully related to the properties of nerve cells recorded in cerebral cortex.

An important feature of this model is that the initial encoding of disparity is piecewise-frontoparallel. That is, the model neurons respond best when the disparity within their receptive field is constant. This explains why the resolution for disparity gratings is so much lower than for luminance gratings. V1 receptive fields typically have several different ON or OFF subregions which respond to different luminance polarities. The limiting period for luminance gratings reflects the size of these subregions, not the receptive field as a whole. In contrast, in the stereo domain, V1 receptive fields appear to prefer uniform disparity (Nienborg, Bridge, Parker et al. 2004).

Sine-wave disparity gratings (Figure 1A) are always a sub-optimal stimulus for this population, since their disparity is never even locally constant. Square-wave disparity gratings, on the other hand, consist of regions of locally constant disparity (Figure 1B). When the grating's period exceeds the window-size used for local cross-correlation, the disparity within the window is constant. Neurons with the optimal tuning (black ellipses in Figure 2) should thus experience an interocular correlation of near unity. Critically, this statement is true independent of the grating's amplitude (compare Figure 2C with Figure 2A, blue curves). In contrast, for sine-wave gratings, the range of different disparities falling in a window depends on the amplitude of the grating (compare Figure 2C with Figure 2A, red curves and green shaded regions).

From this qualitative argument, we expected that the piecewise-frontoparallel model should find it easier to detect square-wave disparity gratings than sine-wave gratings, especially at high frequencies and/or amplitudes. If this prediction were borne out in human observers, this would be a powerful confirmation of the model. In this chapter, then, we first carry out computer simulations to establish whether the piecewise-frontoparallel model really does respond better to square-wave than sine-wave disparity gratings, and whether this is sensitive to the precise way in which the model is implemented. We next carry out psychophysical experiments to compare human performance to the predictions of the piecewise-frontoparallel model.

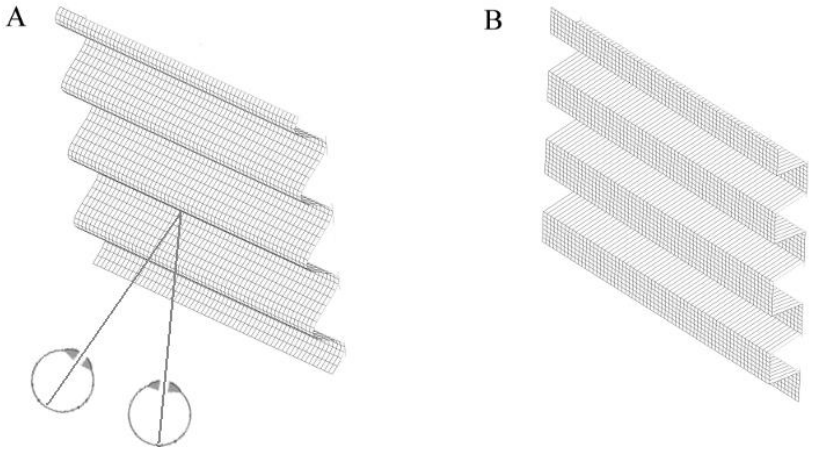


Figure 1: Physical surfaces implied by (A) sine-wave and (B) square-wave disparity gratings

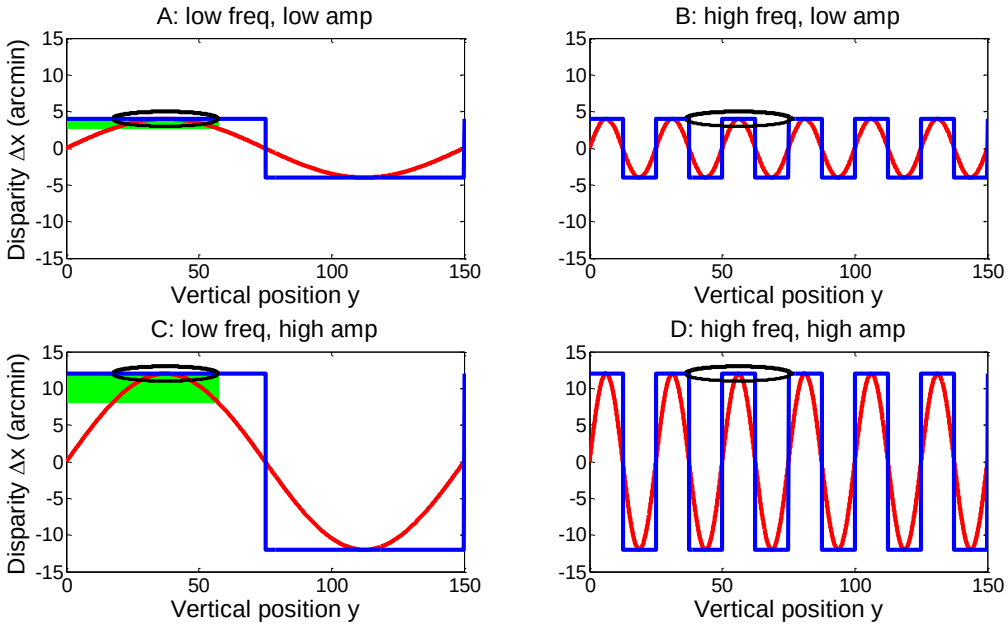


Figure 2: Sketch of sine- and square-gratings and a receptive field. The red and blue curves show the profile of (respectively) sine- and square-wave disparity gratings, with disparity varying as a function of vertical position in the visual field. The 4 panels show gratings with low (AC) and high (BD) spatial frequencies, and with small (AB) and large (CD) amplitudes. The black ellipse shows the receptive field of a model neuron tuned to the largest disparity in the grating. At low frequencies (AC), the period of the grating is large compared to the correlation window (black lines), and the grating can be resolved. At high frequencies (BD), the period is small compared to the window, and the grating cannot be perceived. At low frequencies (AC), the square-wave presents only a single disparity within the local correlation window. This is not so for the sine-wave. For low amplitudes, the range of disparities within the receptive field is small (green shaded region in A), but as the amplitude increases, this range increases (green shaded region in C), even for low-frequency sine waves.

2.2 Methods

2.2.1 *Psychophysics*

2.2.1.1 Experimental Setup

The experiments were performed using a mirror stereoscope. The stimuli were displayed on the left and right halves of a single LCD-monitor with a physical display size of 41x25.5 cm and a resolution of 1440x900 pixels. The size of the images was 350x350 pixels. With the viewing distance of 308 cm, the images subtended 1.8°x1.8° and each pixel subtended

0.3 arcmin. The mirrors were aligned to make the vergence distance 308 cm. The monitor was linearized (gamma-corrected) using a Minolta LS-100 photometer. White pixels were 240 cd/m^2 and black pixels were 0.26 cd/m^2.

2.2.1.2 Stimuli

Stimuli were presented using Matlab (The Mathworks, Natick, MA, USA; www.mathworks.com) with the Psychophysics Toolbox (Brainard 1997; Pelli 1997). The grating stimuli used were random-dot stereograms depicting horizontally-oriented depth corrugations with either sine-wave or square-wave profiles (Figure 1). We varied the amplitude, frequency and phase of the gratings. Amplitude is defined as half the peak-to-trough range of the waveform, (max-min)/2, except in the section Frequency analysis, where the amplitude of the fundamental is specified. The dots were 2x2 pixels, 0.6x0.6 arcmin, and were white on a black background. Anti-aliasing, implemented in-house in our own Matlab code, was used to place dots at sub-pixel locations. The long viewing distance (308 cm) and small pixel size (0.3 arcmin, less than the retinal cone spacing) were used to ensure that the range of disparities and frequencies perceived by human observers was not limited by the resolution of the display. For the highest grating frequencies used in this study (5.7 cpd), sine- and square-wave profiles could be readily perceived and distinguished from one another when the stimuli were viewed up close in anaglyph, although they became invisible as the observer walked further away. This demonstrates that the limits on grating detectability were contained in the observer's visual system, not the physical display.

2.2.1.3 Task

A two-interval forced-choice task was used, where one temporal interval contained a disparity grating and the other contained disparity noise (described below). The task was to report, by a button press, which interval contained the disparity grating. For three subjects the length of the temporal intervals was 500 ms, with a 100 ms blank between the intervals. A fourth subject was allowed to view each of the intervals for as long as he wanted before making a choice. Experimental trials were organized in blocks, most of which consisted of 240-280 trials, where the frequency of the gratings was kept constant in each block. The

two waveforms and the different phases were always interleaved in blocks of experimental trials and in most cases different amplitudes were interleaved as well.

On each trial, the disparity noise image was generated by assigning each dot a disparity drawn at random from the same distribution as the disparity grating presented in the other interval. Thus, in trials where the grating was a square-wave with amplitude A, the disparity noise dots had disparity +A or -A with equal probability. On sine-wave trials, they had a disparity in the range [-A,+A] . In the grating stimuli, all dots at a given vertical position had the same disparity, but in the noise stimuli, disparity was picked without reference to vertical position, so dots in the same row would have different disparities.

2.2.1.4 Observers

The 4 observers were the two authors, one additional experienced psychophysical observer and one inexperienced observer.

2.2.1.5 Data analysis

A truncated probability density function of a gamma distribution was fit to the data for each frequency. This was simply a descriptive function without any theoretical significance. The Matlab function FIT, using non-linear least squares, was used to do the fitting.

2.2.2 *Model*

2.2.2.1 Stimuli and task

The same stimuli that were used in the psychophysics were also used in the modeling. The model had the same task as the human subjects: in each trial it was presented with two image-pairs, one containing a grating and one containing a noise pattern and it had to judge which one contained the grating.

2.2.2.2 Preprocessing

The model used here was based on the piecewise-frontoparallel local cross-correlation based model of Banks et al. (2004). The left- and right-eye images were first preprocessed to simulate the effects of the eye's optics, and then passed to a cross-correlator.

The preprocessing consisted of convolving the images with the point-spread function of the well-focused eye:

$$h(x,y) = a * h_1(x,y) + (1-a) * h_2(x,y)$$

where

$$h_i(x,y) = \left(s_i\sqrt{2\pi}\right)^{-2} e^{-0.5(x^2+y^2)/s_i^2}$$

and a=0.583, s_1=0.443 arcmin and s_2=2.04 arcmin (Geisler and Davila 1985; Filippini and Banks 2009). The images were then scaled to make the distance between rows and columns 0.6 arcmin. This was done to make sure the resolution of the images was no higher than the spacing between cones at the fovea (Geisler and Davila 1985; Filippini and Banks 2009; Rossi and Roorda 2009).

2.2.2.3 Cross-correlator

The preprocessed images were then passed to the cross-correlator. A window was moved along a vertical line in one eye's image. For each vertical position of that window, a second window in the other image at the same vertical position was moved across an interval of horizontal positions centered on the horizontal position of the first window. For each combination of window-positions the correlation between the content of the windows was recorded. The correlation was defined as:

$$C(y,\Delta x) = \frac{cov(L_w, R_w)}{\sqrt{cov(L_w, L_w)cov(R_w, R_w)}}$$

where L_w and R_w are the contents of the windows in the left and the right image multiplied by the window function and cov is the covariance. The window functions used to obtain the main results presented here were Gaussians centered on the current window position (that is, (Δx/2,y) in one eye and (-Δx/2,y) in the other) and cut off two standard deviations from the centre in each direction. The output from the cross-correlator was a two-dimensional image of correlation as a function of the horizontal disparity, Δx, between the windows as well as the vertical position of the windows, y (see Figure 3). The disparities used were in the range from -25 to 25 arcmin with a step of 0.6 arcmin (1 pixel in the scaled images). The step in the y-position was also 1 pixel in the scaled images.

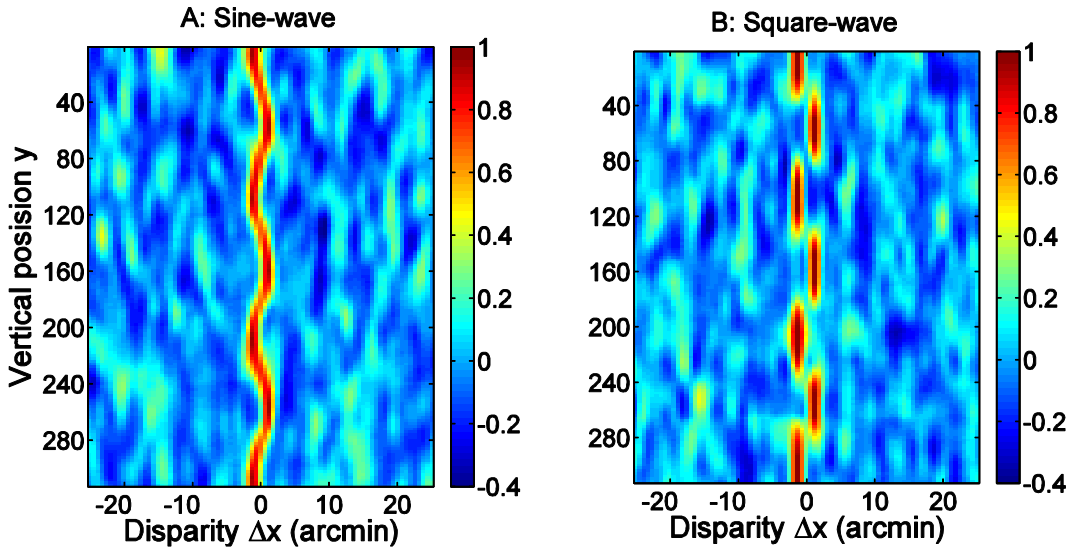


Figure 3: Examples of output from the cross-correlator for one sine-wave and one square-wave grating, both with a frequency of 1.9 cpd. A Gaussian window with 2*σ = 6 arcmin was used.

2.2.2.4 Decision rule 1: Autocorrelation

Two different methods were used to make a decision on which interval contained the gratings based on the correlation images. The first was based on autocorrelation, and the second on template matching. The method based on autocorrelation started by finding the maximum correlation across all horizontal window positions, Δx, for each vertical window position, y, and recording the difference in horizontal position between the two windows as an estimate of the horizontal disparity at that vertical position:
$\Delta x_{est}(y) = argmax(C(y, \Delta x))$.

The autocorrelation of the resulting curve of estimated disparity as a function of vertical position, $\Delta x_{est}(y)$ was then calculated as:

$$ac_n = \frac{\sum_{i=1}^{N-n}(\Delta x_{est}(y_i) - \mu)(\Delta x_{est}(y_{i+n}) - \mu)}{((N-n) * \sigma^2)}$$

where μ is the mean and σ is the standard deviation of Δx_{est}. Two examples of what the auto-correlograms looked like are given in Figure 4. Finally both a sine-wave and a triangular wave, which are the auto-correlation functions of a sine-wave and a square-wave respectively, with the same frequency used in the stimulus were fit to the auto-correlogram and the r^2-value of the best fit was recorded. For each pair of a wave and a noise pattern,

making up a single trial, the image pair which got the highest r^2-value was guessed to contain the grating.

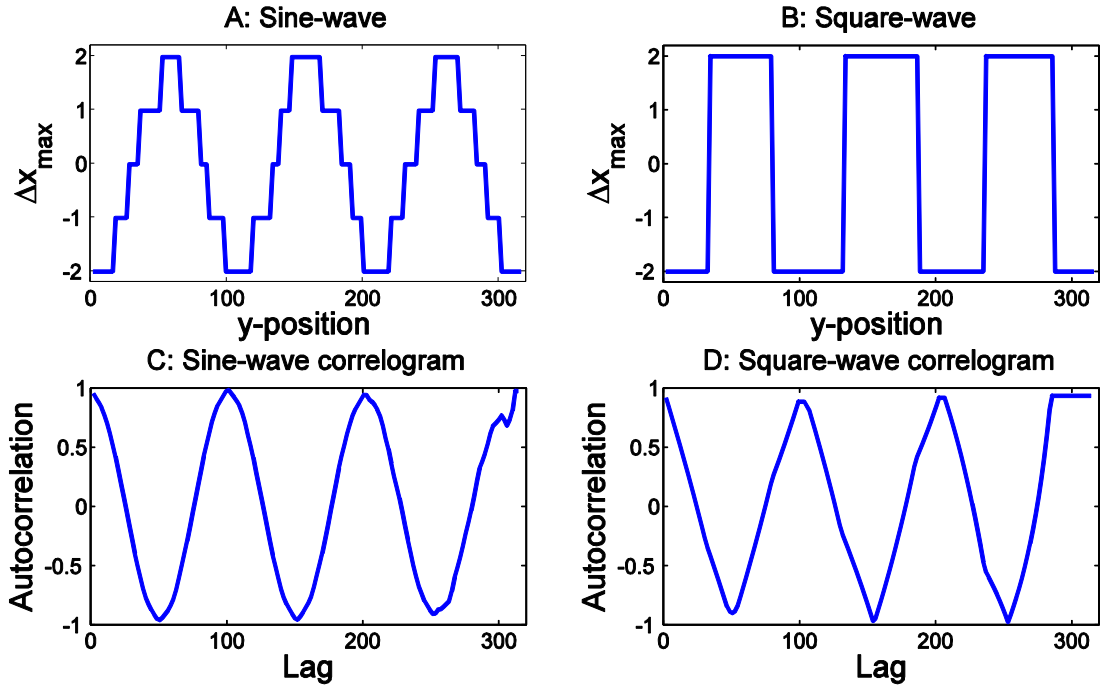


Figure 4: Examples of estimated disparity curves and their autocorrelograms for one square-wave and one sine-wave both with a frequency of 1.9 cpd. A Gaussian window with 2σ = 6 arcmin was used. The estimated disparity curve for the sine-grating is quantized because the model only included detectors tuned to integer pixel disparities.

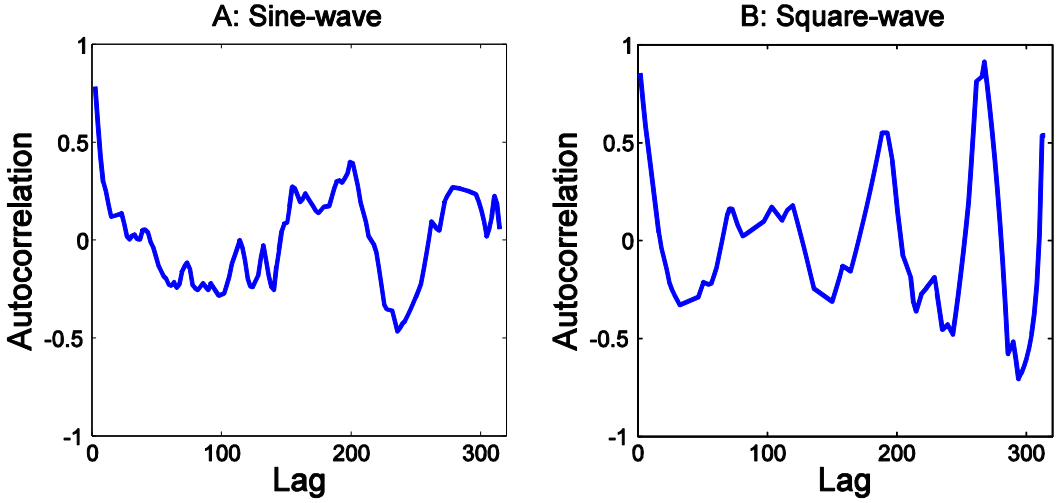


Figure 5: Example of autocorrelation for the corresponding noise patterns to one square- and one sine-wave. A Gaussian window with 2*σ = 6 arcmin was used.

2.2.2.5 Decision rule 2: Template matching

This method used a set of templates of the correlator output for the disparity gratings (grating templates) as well as a set of templates of the correlator output for the two types of noise patterns (noise templates). The set of grating templates covered all frequencies, amplitudes and phases used in the simulations as well as both wave forms. The set of noise templates covered all the amplitudes (the noise patterns were by their nature independent of frequency and phase). For each interval, the grating template and the noise template with the highest correlation to the correlator output were chosen. The correlations were calculated as follows:

$$C_n = \frac{\sum((CO(\Delta x, y) - \mu_{CO})(T_n(\Delta x, y) - \mu_{T_n}))}{\sqrt{\sum(CO(\Delta x, y) - \mu_{CO})^2}\sqrt{\sum\left(T_n(\Delta x, y) - \mu_{T_n}\right)^2}}$$

where CO is the correlator output, T_n is one of the templates, μ_{CO} and μ_{T_n} are the means over all disparities Δx and all y-positions of the correlator output and template T_n respectively. All sums were performed over disparity and y-position. The interval for which the difference between the correlation to the grating template and the correlation to the noise template was the highest was guessed to contain the grating.

For each grating profile (sine- vs square-wave), frequency, amplitude and phase, the corresponding template was generated by presenting 100 different random dot stereograms to the cross-correlator, after the same preprocessing steps used in the main model. The resulting set of 100 correlation images were then used to calculate the average for each pixel (see Figure 6). The phase of the disparity gratings was varied in steps of 10° when generating the templates and when testing the model the phase was randomly chosen at each trial to be one of the 36 different phases represented in the set of templates. The template amplitudes were 0.3, 1.3, 2.5, 5.1, 7.6, 10.1, 15.2 and 20.2 arcmin. The template frequencies were 1.9, 2.5, 3.2, 3.8, 4.4, 5.1, 5.7, 6.3, 7.0 and 7.6 cpd. Thus, there were 5760 grating templates and 16 noise templates.

We have also examined a somewhat different template matching rule, where the correlator output was matched only to templates of the same frequency, where no noise templates

were used and where the matching was based on sums of squared differences instead of correlation. This decision rule performed slightly worse in general but the results were qualitatively very similar to the results with the decision rule described in this section.

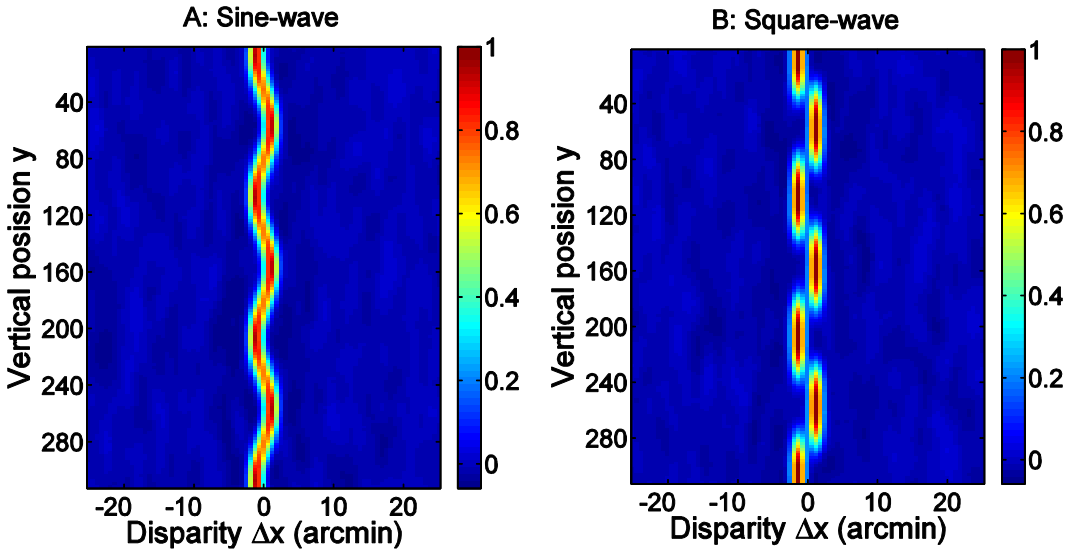


Figure 6: Examples of templates for sine-waves (left) and square-waves (right) with a frequency of 1.9 cpd.

2.3 Results

We begin by examining the behavior of the two correlation-based models, and then compare this to the performance of our human observers.

2.3.1 *Model*

2.3.1.1 Decision rule 1: Autocorrelation

Figure 7 shows the results for the model with decision rule 1 (autocorrelation). The boxed panel summarizes the results by plotting maximum performance over all amplitudes against frequency. In this and all further graphs the error bars show 95% confidence intervals, the red curves show data for sine-waves and the blue curves show data for square-waves. When the maximum performance over all amplitudes is plotted against frequency there is very little difference between the curves for the two different waveforms (boxed panel (K) in Figure 7). However, when we examine how performance depends on disparity amplitude (Figure 7A-J), a key difference emerges between the two waveforms. At the smallest amplitude tested, performance is near chance, but rapidly rises to its peak value. For sine-

wave disparity gratings, performance then declines again as the disparity amplitude increases further. For the square-wave gratings, in contrast, performance remains at its peak value as the amplitude increases.

This is readily explicable. The model is built from local correlation detectors tuned to constant disparity. They respond well if disparity remains roughly constant over their window, and they do not respond well when there are steep disparity gradients. For sine-wave gratings, increasing the disparity amplitude also increases the disparity gradient at every point (except the turning-points), reducing performance. This reduction of performance with increasing disparity gradient was also found by Banks et al. (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009) and was predicted by Kanade and Okutomi for a slightly different cross-correlation model (Kanade and Okutomi 1994). For square-wave gratings, the disparity gradient is zero everywhere except at the discontinuities, and this remains true as the amplitude increases. Thus, performance remains high, as long as the amplitude does not go outside the range of disparities to which the model is sensitive. The reason why performance is low for the lowest amplitude is because this amplitude, 0.3 arcmin, is lower than the step in the range of correlation detectors which is 0.6 arcmin, The closest detectors are therefore at 0 and 0.6 arcmin, which are equally far from 0.3 arcmin and they will therefore be close to equally strongly activated by this disparity. The autocorrelation-based rule only uses the detector with the strongest response at each y-position and the detector for 0.6 arcmin can only be the most strongly activated one when the entire window or very close to the entire window is seeing 0.3 arcmin. This can only happen for the squarewaves, and it is only for the lowest frequency that it happens for a large enough range of y-values to allow detection.
Given that the model is built to respond to locally-constant disparity, it is perhaps surprising that at low amplitudes (above 0.6 arcmin) it performs as well with sine-waves as for square-waves. Figure 3 shows that the peak cross-correlator output reached for sine-wave gratings does fluctuate across the cycle, being – unsurprisingly – lower where the disparity gradient is higher. However, recall that our model estimates disparity from the correlation-detector reporting the largest response. Thus, so long as the peak is above the background noise, the correct disparity will still be identified. In addition, the decision rule (here, based on the auto-correlation of the estimated disparity profile) can still correctly

identify which interval contains the grating, even if the estimated disparity is not accurate everywhere.

We have examined the behavior of this model with different window sizes. Quantitatively, as the window size increases, performance naturally starts dropping at lower frequencies. Banks et al. found that decreasing the window size improves performance up to a limit which depends on the level of blur (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). For optical levels of blur they found the limiting window size to be about 6 arcmin, the value used in Figure 7. Window size does not affect the qualitative behavior of the model. In particular, we continue to find that (1) maximum performance as a function of frequency remains the same for both sine- and square-wave gratings (see the boxed panel in Figure 7); and (2) performance declines as a function of amplitude for sine-wave gratings, but remains at its peak value for square-wave gratings (see Figure 7).

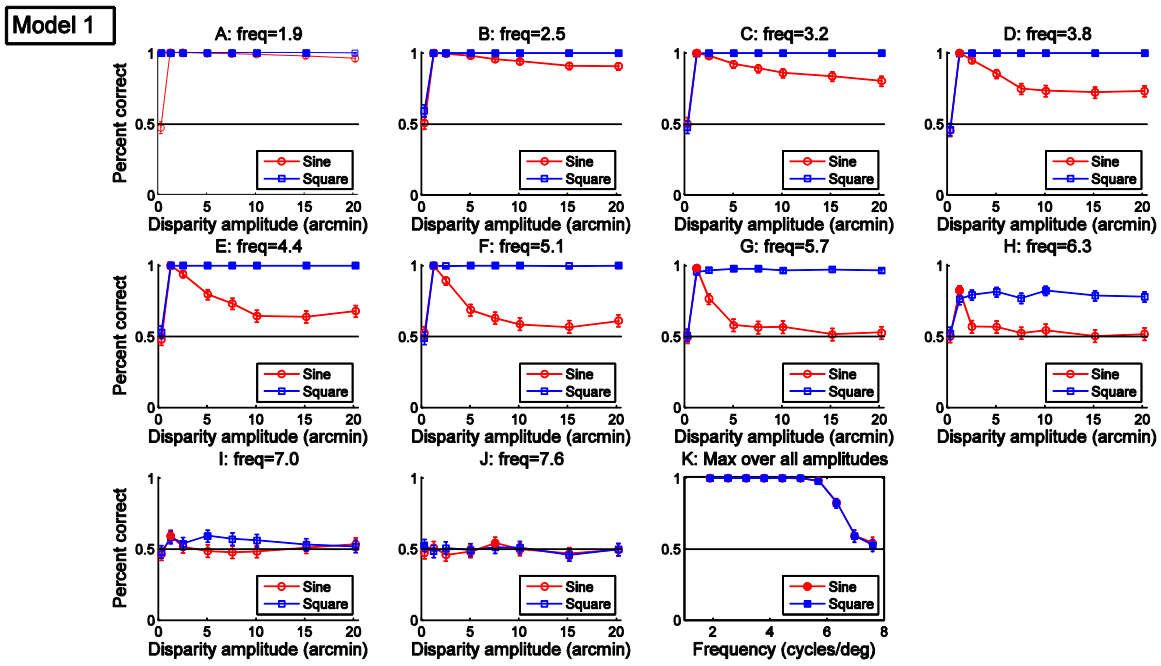


Figure 7: Performance as a function of amplitude and frequency for the model with the decision rule based on autocorrelation and a window with 2*σ = 6 arcmin. The boxed plot (K) shows the maximum performance over all amplitudes for each frequency.

2.3.1.2 Decision rule 2: Template matching

It is important to be clear whether these features of the model performance reflect the low-level, correlation-based encoding of disparity, or whether they are specific to the particular decision rule chosen. In this section, we therefore present results from a more elaborate decision rule. This rule is based on matching the output of the correlation-detector to any one stimulus, to a set of stored template responses to grating patterns. The template set includes responses to both sine- and square-wave gratings, and the decision rule uses whichever matches.

The results for this decision rule are shown in Figure 8, in the same format as in the previous section. The modeling results with the decision rule based on template matching are qualitatively very similar to the results with the autocorrelation based decision rule. The main differences seem to be that, for a given window size, performance starts dropping at slightly lower frequencies and that for the lowest frequencies performance for the sine-waves remains high up to the highest amplitude tested. The reason for the higher performance for high amplitude sine-waves may be that the template matching rule requires accurate disparity detection at a smaller percentage of y-positions to identify a grating; high correlation in small regions close to the peaks of the sine-waves (see Figure 13) may be enough since the relevant template has the same pattern. The drop in performance for the lowest amplitude happens only to a lesser degree for the template matching rule than for the autocorrelation based rule. This is because the template matching rule uses the outputs from all the correlation detectors and not just the one that has the strongest response at each y-position. However, critically, both decision rules show the same key features highlighted at the end of the previous section. In particular, as disparity amplitude increases, performance remains high for the square-wave gratings and declines for the sine-wave. The alternative template-matching approach mentioned in the Methods also showed this behavior (results not shown). Thus, this key behavior is not dependent on any particular decision rule. As explained in the previous section, we attribute it to the properties of the initial disparity encoding performed by correlation-detectors tuned to uniform disparity.

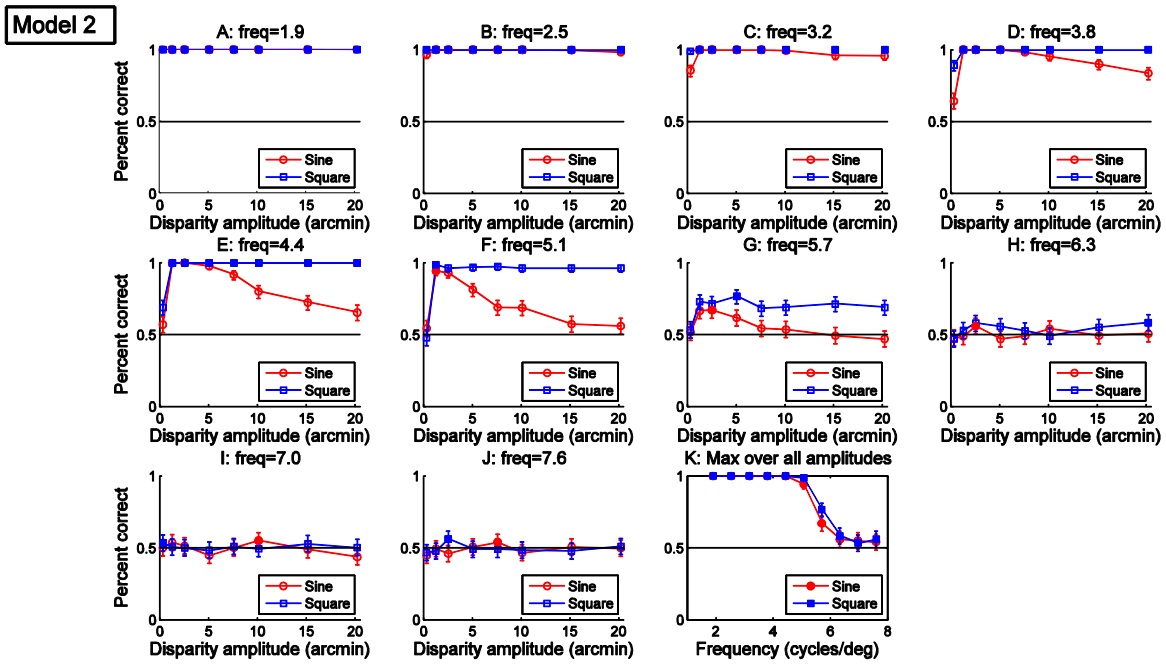


Figure 8: Performance as a function of amplitude and frequency for the model with the decision rule based on template matching and a window with $2\sigma = 6$ arcmin. The boxed plot (H) shows the maximum performance over all amplitudes for each frequency.

2.3.2 *Psychophysics*

We now examine the performance of human subjects, in order to compare it with the predictions of the model. Figure 9 shows the data for four subjects. The boxed panels summarize the results by plotting the maximum performance over all amplitudes for each frequency.

In striking agreement with the predictions of the correlation-based model of Banks and colleagues, we find that the best performance reached at a given frequency is the same for both waveforms. The boxed panels (FLRX) in Figure 9 show this best performance obtained at any disparity amplitude, plotted as a function of the frequency. None of our 4 subjects shows any significant difference in best performance between sine- and square-wave gratings.

However, when we examine the graphs showing performance against disparity amplitude for individual frequencies, we find a significant departure from the model predictions. After initially rising to a peak value, performance then declines as disparity amplitude increases further. The model shows this decline only for sine-wave, not for square-wave gratings. However, for humans, the rate of decline is extremely similar for both sine-wave and square-wave gratings. Where significant differences do exist (e.g. subject PFA, discussed in more detail below), performance is better for the sine grating, not the square-wave as predicted by the model.

Ultimately, of course, the performance of any realistic system must decline, as the disparity of the stimulus moves beyond the range to which its detectors respond. This effect was not included within our model (previous section), which contained an equal number of detectors for all disparities used. However, we do not believe that this omission can account for the difference between model and human performance we observe. We could force the model's performance down for large-amplitude square-wave gratings by reducing the range of disparity detectors. However, the resulting reduction in performance would not be specific to square-wave gratings, but also affect sine gratings. It thus could not reconcile the model with human performance. It would also be unrealistic, because the disparity amplitudes used here are very small, well below Dmax (Glennerster 1998; Read and Eagle 2000) and perfectly detectable in other contexts. This is clear from our own data. Disparity amplitudes which our subjects find easy at low frequencies become impossible at higher frequencies. For example, at a frequency of 1.9 cpd, subject ISP performs at virtually 100% out to amplitudes as large as 10 arcmin, the largest examined. Yet at a frequency of 5.1 cpd, he is at chance for this amplitude, for both sine and square-wave gratings. This cannot be because he lacks neuronal mechanisms capable of encoding disparities of 10 arcmin, since he perceived 10 arcmin perfectly at the lower frequency. Equally, it cannot be because 5.1 cpd is too high a frequency compared to the window-size of his correlation-detectors, because he reaches 80% correct for both grating profiles when the amplitude is smaller, 2-3 arcmin. His poor performance can only be due to the particular combination of frequency and amplitude.

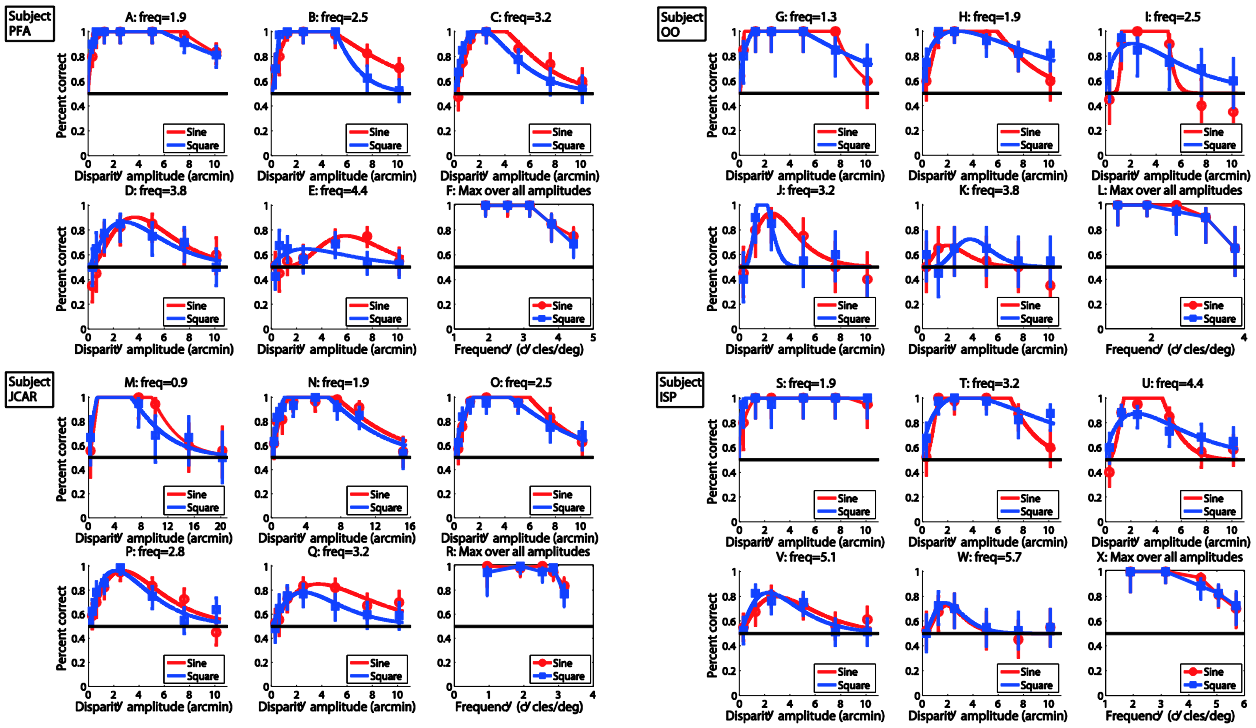


Figure 9: Performance as a function of amplitude and frequency for each subject. The squares and circles are data points and the lines are fits. The boxed plots (FLRX) show, for each frequency, the maximum performance obtained at any amplitude.

2.3.3 *Frequency analysis*

This suggests that the correlation-based model may fail to capture some aspects of human depth perception. We now examine another influential approach to human perception, the Fourier or frequency-based analysis pioneered in the luminance domain by Campbell & Robson (1968), and later applied to disparity (Tyler 1975; Schumer and Ganz 1979; Cobo-Lewis and Yeh 1994; Grove and Regan 2002).

In Fourier analysis, a square-wave grating can be decomposed into a sum of sine-wave gratings: a sine-wave of the same period as the square-wave but with $4/\pi$ its peak-to-trough range, plus successive lower-amplitude sine-waves. As the grating period decreases to the limit of detectable frequencies, a point is reached where the fundamental frequency is still above threshold, but the third harmonic is already below threshold. Sine- and square-wave gratings thus become indistinguishable. Most of our data falls within this domain, since for most subjects the highest frequency tested was just at the threshold of discriminability,

whereas the lowest frequency tested was more than one-third of this value. This means that even at the lowest frequency tested, the third harmonic distinguishing the square-wave from the sine-wave grating would be nearly undetectable if presented alone. Thus if the linear theory is correct, if we plot performance as a function of the amplitude of the fundamental, instead of the whole-waveform amplitude used so far in this chapter, performance should become the same for square-wave and sine-wave gratings.

This is examined in Figure 10. This figure shows the same data as Figure 9, but now plotted as a function of the amplitude of the fundamental. The sine (red) data is thus unchanged, while the square (blue) data and fits are shifted to the right by a factor of $4/\pi$. To assess whether this manipulation brings performance for the two waveforms closer together, we used the curves fitted to each set of data. For each frequency, we computed the integral of the absolute difference between the curves for the sine-waves and the square-waves, first for the original data and then for the adjusted data. If this integral was smaller for the adjusted data, this indicated that the shift to fundamental amplitude had brought the results closer together. This is indicated with a + symbol at the bottom-left of the panels in Figure 10; a – symbol indicates that the shift to fundamental amplitude brought the fits further apart. Bootstrap resampling was used to estimate the significance of any change. The asterisks in Figure 10 indicate $p<0.05$ (two-tailed test), while NS indicate that the adjustment had no significant effect either way.

For subject PFA, who performed the most repetitions per condition, plotting performance as a function of fundamental amplitude brings the curves closer together at every frequency. Plotted as a function of peak-to-trough amplitude (Figure 9), PFA often performed slightly better for the sine-waves. When the data is adjusted so performance is a function of the fundamental amplitude (Figure 10), this effect is almost totally abolished, and the two sets of data overlap almost perfectly. However, this improvement was significant for only frequency, 2.5cpd. For the other subjects, there is little evidence of any systematic effect one way or the other. Thus, our results provide little support for the linear Fourier analysis of disparity. Subjects perform very similarly for high-frequency sine-wave and square-wave gratings, but their performance does not seem to be set by the amplitude of the fundamental.

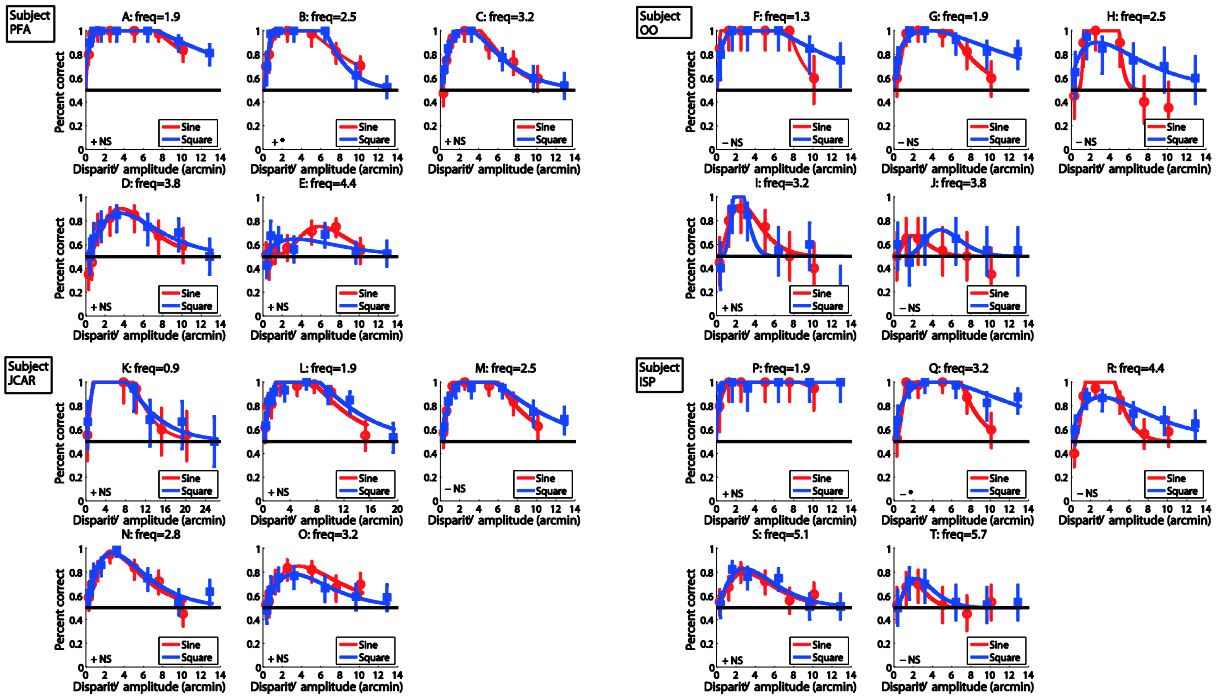


Figure 10: Performance as a function of amplitude and frequency for each subject, with the square-wave data plotted against the amplitude of the fundamental frequency component of the waves. The symbols at the bottom left of each panel show whether this has improved (+) or worsened (-) agreement between the sine- and square-wave results, and whether this is significant at the 5% level (*) or not (NS).

2.3.4 *Disparity gradient limit*

Several previous studies have suggested that stereopsis may be limited by the disparity gradient, rather than disparity per se (Tyler 1975; Burt and Julesz 1980; Kanade and Okutomi 1994; McKee and Verghese 2002; Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). To examine this, in Figure 11 we plot the performance of all subjects on sine-wave gratings of all frequencies plotted against the amplitude of the gratings (ACEG) as well as the maximum disparity gradient in the gratings (the product of frequency and amplitude, BDFH). In order to test whether plotting against disparity gradient brings the curves closer together we used the fits after extending them to end at the same point and cutting them to only include the portion of the curve after the peak. The standard deviation of the set of y-positions that the different curves passed through was computed at each x-position and the mean of this standard deviation over all x-positions was used as a measure of how closely the curves were superimposed. Bootstrap resampling was used to estimate

the significance of any difference between the two ways of plotting the data. The curves were found to be significantly more superimposed ($p<0.05$) when plotting against disparity gradient for three out of four subjects (PFA, OO and ISP). For the fourth subject no significant difference either way was found. Thus, our data are consistent with the idea that performance at high amplitudes is limited by the highest perceivable disparity gradient.

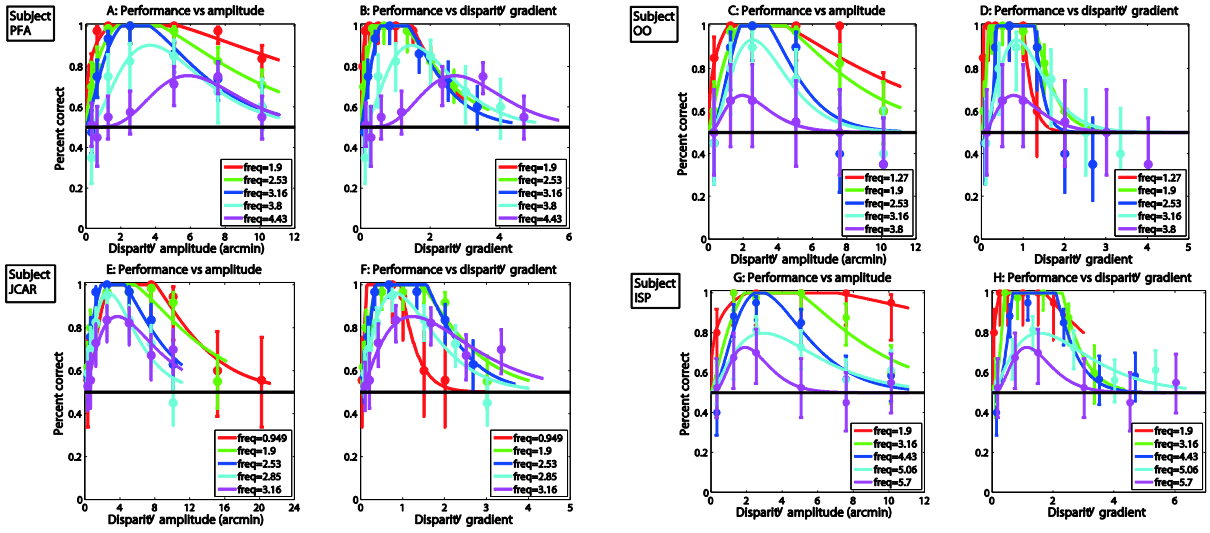


Figure 11: Performance plotted against amplitude (ACEG) and maximum disparity gradient (BDFH) for sine-waves of all frequencies, for each of the 4 subjects.

Figure 12 shows the same plots for the model with the auto-correlation decision rule (similar results were obtained for the model with the template-matching decision rule, not shown). Again, performance on sine-waves of different frequencies is plotted against either the amplitude (A) or the maximum disparity gradient (B) of the grating. To test whether the model results superimposed better when plotted against disparity gradient we used the parts of the curves from the peak to the last data point for the lowest frequency. As for the human data, the standard deviation of the set of y-positions that the different curves passed through was computed at each x-position and the mean of this standard deviation over all x-positions was used as a measure of how closely the curves were superimposed. Bootstrap resampling was used to estimate the significance of any difference between the two ways of plotting the data. No significant difference was found for the results with either of the decision rules. Thus, for the model results, the curves do not superimpose any better when the data is plotted against disparity gradient. Rather, the performance of the model depends separately on frequency and amplitude, and not simply on disparity gradient (amplitude × frequency). This is not surprising given that the model has no mechanisms which specifically detect disparity gradient. The observed dependence of frequency and amplitude may be because the correlation output from the first stage of the model has the highest correlation in the regions close to the flat parts of the sine-wave (see Figure 13). Thus, performance may be limited by the size of the regions that are flat enough to generate high correlation, rather than by the maximum disparity gradient in the stimulus.

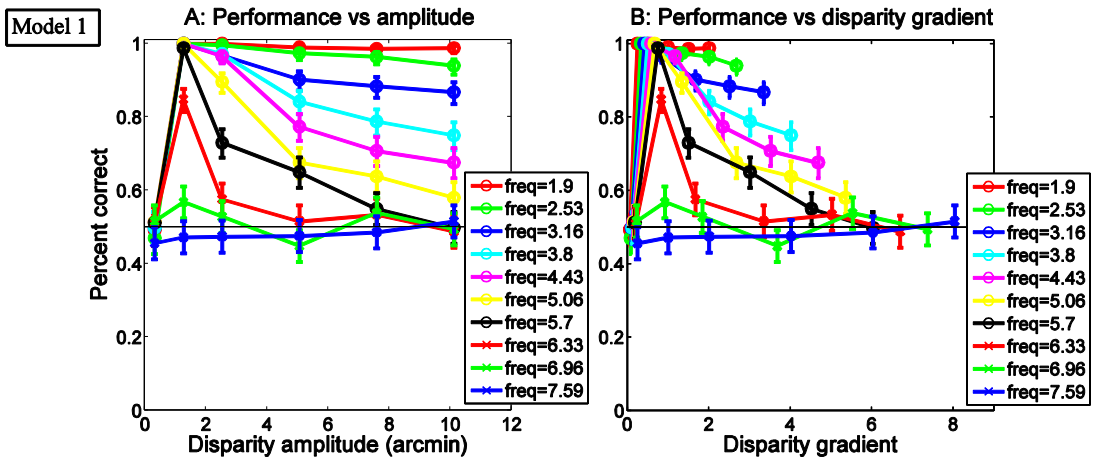


Figure 12: Performance plotted against amplitude (left) and maximum disparity gradient (right) for sine-waves of all frequencies for the model with the decision rule based on autocorrelation

2.4 Discussion

In recent years, many models of human stereopsis have proposed that the initial encoding of disparity occurs in primary visual cortex, V1, by disparity-selective neurons whose major properties are captured by the stereo energy model (Ohzawa, DeAngelis and Freeman 1990; Qian 1994; Qian and Zhu 1997; Cumming and DeAngelis 2001; Read 2005). The neurophysiological evidence suggests that V1 neurons respond optimally to disparity which is constant across their receptive field (Nienborg, Bridge, Parker et al. 2004). In higher brain areas, neurons are found which respond best to particular patterns of varying disparity (Janssen, Vogels and Orban 1999; Sakata, Taira, Kusunoki, Murata, Tsutsui, Tanaka, Shein and Miyashita 1999; Sugihara, Murakami, Shenoy, Andersen and Komatsu 2002; Nguyenkim and DeAngelis 2003). However, current models propose that these higher-level neurons are built by combining the outputs of uniform-disparity V1 neurons (Thomas, Cumming and Parker 2002; Bredfeldt and Cumming 2006; Bredfeldt, Read and Cumming 2009). Thus, Banks and colleagues (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009) have argued that the initial piecewise-frontoparallel encoding of disparity imposes a fundamental limit on stereo resolution. In this view, the high-frequency limit for perceiving disparity gratings is imposed right down in V1, by the receptive field size of disparity-selective neurons.

This piecewise-frontoparallel theory of disparity encoding is quite different from the Fourier or frequency-based analysis pioneered in the luminance domain by Campbell & Robson (Campbell and Robson 1968), and later extended to disparity (Tyler 1975; Schumer and Ganz 1979; Cobo-Lewis and Yeh 1994; Grove and Regan 2002). In that picture, the quantity of interest (disparity or luminance) is initially encoded by a set of frequency channels. The basic "unit" in which the quantity is represented is the sine-wave (or a local version of it, like a Gabor), not a constant-value patch as in the piecewise-frontoparallel theory. In linear Fourier theory, square-wave and sine-wave gratings with the same fundamental amplitudes should become equally detectable at high frequencies, once the third harmonic of the square-wave has passed above the frequency threshold. In contrast, if the piecewise-frontoparallel theory is correct, it should be easier to perceive a square-wave

disparity grating than a sine-wave grating, because the square-wave grating consists of locally frontoparallel regions of disparity, and so should drive V1 neurons more strongly. It is of course possible that there are frequency channels which are built by combining the outputs of uniform-disparity V1 neurons. If so, the initial fronto-parallel representation of disparity will still set limits on performance, even though the later processing needed to construct the frequency channels may limit performance further.

We tested the behavior of the piecewise-frontoparallel model by running simulations. We verified that the model does indeed find it easier to detect square-wave gratings, which are piecewise-frontoparallel, than sine-wave gratings, which everywhere have a non-zero disparity gradient. In particular, for square-wave gratings the model was able to perform well out to high amplitudes (limited only by the range of preferred disparities included within the model neuronal population), whereas for sine-wave gratings, performance declined at high amplitudes. This behavior is what we expected given the structure of the model, cf Figure 2. We confirmed that it does not depend critically on the particular details of the model implementation; for example, we obtained the same behavior with two quite different decision rules. Rather, it reflects the initial stage of local cross-correlation. Figure 13 shows the output of this stage for both sine- and square-wave gratings, at low and high amplitudes, for a relatively low frequency, 1.9 cpd. At low amplitudes, the piecewise-frontoparallel model can successfully track the disparity of both grating profiles (Figure 13AB). In contrast, at high amplitudes (Figure 13CD), only the very peaks of the sine-wave grating remain visible (where the disparity gradient is briefly zero), while the square-wave grating remains just as clear as at low amplitude. Thus, our simulations confirm our intuitions about the behavior of models based on piecewise-frontoparallel disparity encoding.

However, to our surprise, our psychophysical results were quite different. There was no evidence that performance was ever significantly better for square-waves than sine-wave gratings. Like the model, the maximum performance possible at a given frequency was indistinguishable for the two wave-forms. However, after initially rising to a peak, human performance declines as a function of amplitude for both sine- and square-waves. This is quite different from the behavior of the piecewise-frontoparallel model, where performance

declines only for sine-wave gratings and remains high for the square-wave gratings out to large amplitudes. The decline in the performance of human observers occurs for disparity amplitudes which are clearly detectable at lower frequencies. This shows that the poor performance is caused by the frequency of disparity alternation, not the intrinsic detectability of the disparities present in the stimulus.

Thus like Banks and colleagues, we find that the piecewise-frontoparallel model based on local cross-correlation does an excellent job of capturing human performance on sine-wave gratings. However, the discrepancy with square-wave gratings indicates that the model is incomplete as a model of human stereo vision.

A limitation of this model is that it only includes the initial encoding of disparity in V1, not the higher-level neurons which respond to varying disparity (Janssen, Vogels and Orban 1999; Sakata, Taira, Kusunoki et al. 1999; Sugihara, Murakami, Shenoy et al. 2002; Nguyenkim and DeAngelis 2003). Prominent among these are the class of disparity-edge detectors in V2 (von der Heydt, Zhou and Friedman 2000; Bredfeldt and Cumming 2006). There is considerable psychophysical evidence suggesting that "edges" or discontinuities in disparity are particularly salient for stereo vision (Andrews, Glennerster and Parker 2001; Gillam, Blackburn and Brooks 2007; Serrano-Pedraza, Phillipson and Read in press), presumably reflecting the activation of these neuronal disparity-edge detectors. Square-wave gratings contain sharp disparity edges, whereas sine-wave gratings do not. This is probably why disparity thresholds are consistently better for square-wave than for sine-wave gratings at low frequencies (below 2cpd) (Serrano-Pedraza and Read 2009). Thus, the model's failure to include known mechanisms of edge-detection should, if anything, bring square-wave performance *closer* to sine-wave. This deficiency, therefore, also cannot explain the discrepancy between model and human results.

Our psychophysical results did not provide compelling evidence that disparity is encoded within a set of independent frequency channels. A linear frequency analysis would suggest that, at high frequencies, performance on the two types of grating should become more similar when the amplitude of the grating was expressed as the amplitude of the fundamental, rather than as half the peak-to-trough distance. This was the case for only one

of our four subjects. In contrast, our results were more clearly consistent with previous work indicating that disparity gradient is critical to perception (Tyler 1975; Burt and Julesz 1980; McKee and Verghese 2002).

If neither the piecewise-frontoparallel model, nor a linear frequency analysis, seems capable of fully explaining our results, how should we proceed in order to achieve an accurate model of human stereo depth perception? It may be necessary to invoke further processing happening after the cross-correlation stage. Alternatively, it may be possible to modify the cross-correlation model so as to reconcile it with our results. For example, our current model contains equal numbers of sensors with different disparity tuning, whereas V1 neurons are tuned predominantly to near-zero disparities (Prince, Cumming and Parker 2002). It also assumes that the "window" size used for cross-correlation is constant, whereas V1 neurons tuned to larger disparities tend to have larger receptive fields (Prince, Cumming and Parker 2002), reflecting the size/disparity correlation deduced from psychophysical results (Tyler 1975; Smallman and MacLeod 1994; McKee and Verghese 2002; Tsirlin, Allison and Wilcox 2008). Incorporating such sophistications into our model may help it account for human performance with sine-wave and square-wave gratings.

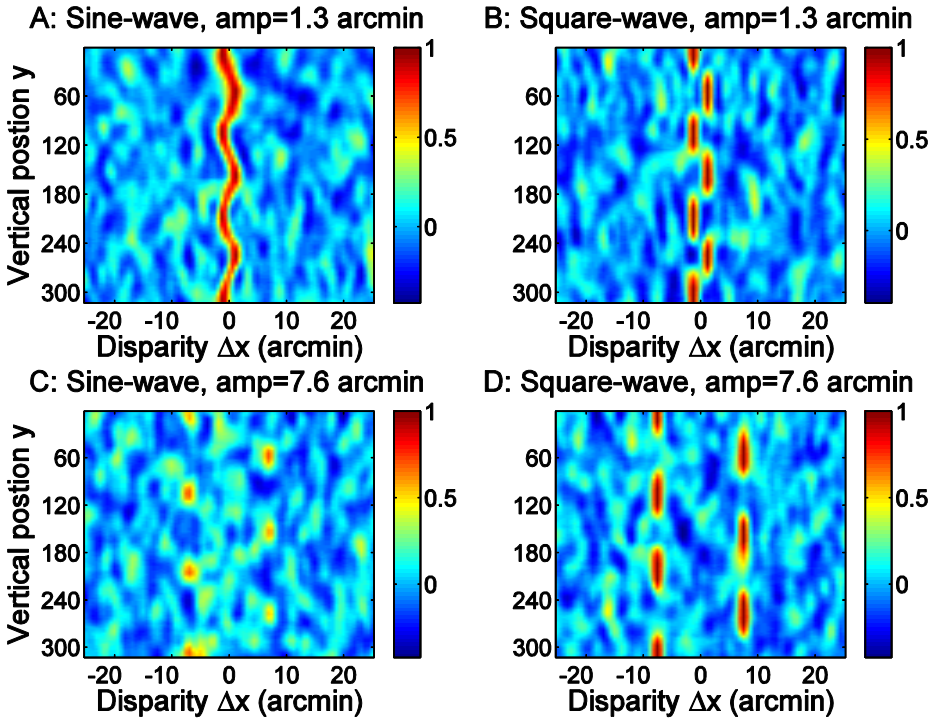


Figure 13: Examples of output from the cross-correlator for square-waves and sine-waves at low and high amplitudes. The quality of the correlation image remains high for the high-amplitude square-wave but drops for the sine-wave, with high correlation only near the peaks. These results are for a frequency of 1.9 cpd and a Gaussian window with $2\sigma = 6$ arcmin.

2.5 Conclusion

Piecewise-frontoparallel local cross-correlation successfully captures many aspects of human stereo vision. However, at least as currently implemented, it predicts that humans should be better at detecting square-wave disparity gratings than sine-wave gratings, when the frequency and amplitude of the gratings are high. In fact, humans perform almost equally well on both grating profiles. In particular, human performance declines as a function of amplitude for both square- and sine-wave gratings, whereas the model predicts a region where performance is independent of amplitude for square-wave gratings. We conclude that the model needs to be refined in order to capture this aspect of human depth perception and we examine how to do this in the next chapter.

Chapter 3. Spatial stereoresolution for depth corrugations may be set in primary visual cortex

3.1 Introduction

In the previous chapter we saw that human subjects perform equally well at detecting sinusoidal and square-wave disparity gratings. We attempted to model this using a local cross-correlation model closely based on the correlation model used in the modeling work by Banks et al. (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). We found that this model captured human performance on sinusoidal gratings well but predicted to high performance for high amplitude square-wave gratings. We concluded that the model would need to be modified in order to explain the new human results. In the discussion we suggested a few modifications that could be made to the model that would make it more physiologically realistic and which might potentially also help account for the human results on sine- vs. square-waves. Among these suggestions we mentioned a modified model incorporating a size/disparity correlation. In this chapter, we examine precisely such a modified version of the model, where larger disparities are detected using larger correlation windows. There is considerable psychophysical evidence for such a size/disparity correlation (Tyler 1973; Tyler 1974; Tyler 1975; Smallman and MacLeod 1994; McKee and Verghese 2002; Tsirlin, Allison and Wilcox 2008), and some physiological evidence has also been found in favour of it (Prince, Cumming and Parker 2002). We show that this new version of the model can capture human performance on both sine- and square-wave depth corrugations.

3.2 Methods

3.2.1 *Model*

3.2.1.1 Stimuli and task

The stimuli and task are as in the previous chapter. Briefly, the stimuli used were random-dot stereograms depicting horizontal sine-wave and square-wave disparity gratings (i.e. modulations in disparity as a function of vertical position in the image). For humans, the disparity gratings are readily visible at low grating frequencies, but as the frequency increases, it becomes impossible to detect the distinct bars of the corrugation, and the dots either appear to be distributed throughout the space between the front and back limits of the stereogram or they appear to be distributed over two planes at the front and back limits, depending on the waveform and amplitude of the grating. Disparity gratings at frequencies beyond the limit of stereoresolution thus remain readily distinguishable from planes of constant disparity or from binocularly uncorrelated dot patterns, but the surface structure cannot be perceived. Accordingly, to probe stereoresolution, we asked subjects to distinguish disparity gratings from disparity noise patterns containing the same range of disparities. Each trial consisted of two intervals. Observers were shown one stereogram depicting a sine- or square-wave grating and one stereogram of the corresponding noise pattern, and had to judge which stereogram contained the grating.

In the psychophysics experiments described in the previous chapter, sine- and square-wave gratings were interleaved so that human observers did not know which sort of grating to look for on any given trial. Disparity grating amplitude and phase were also randomly interleaved, but different frequencies were run in blocks. The computer simulations reflected the human experiments as closely as possible, so the model observer had no prior knowledge of grating waveform, amplitude or phase. The images presented to the model were preprocessed by blurring and scaling to simulate the optics of the human eye, as in the model of Banks et al. (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009) and in the previous chapters.

3.2.1.2 Encoding disparity using cross-correlation

After the preprocessing, the images were presented to a population of cross-correlators tuned to different vertical locations along the grating and to different disparities between left and right eyes. Each cross-correlator had two windows, one in each eye's image. Both windows for a given cross-correlator had the same vertical position. In our model, the left-eye window was always at the same horizontal position. The right-eye window was in one of a range of horizontal positions on either side of the left-eye window. The correlation between contents of the two windows was calculated and recorded for every combination of window-positions. The definition of correlation that was used was:

$$C(y,\Delta x) = \frac{\mathrm{cov}(L_w, R_w)}{\sqrt{\mathrm{cov}(L_w, L_w)\,\mathrm{cov}(R_w, R_w)}}$$

Equation 1

where L_w and R_w are the pixel-values in the left and the right image, multiplied by the window function, and cov is the covariance. We used Gaussian window functions that were cut off at two standard deviations from the centre. That is, if the left window is centered on position (x,y) and $I_L(i,j)$ represents the left eye's image at position (i,j), then

L_w is the set of values $\{I_L(i,j)\exp([-(i-x)^2-(j-y)^2]/2\sigma^2)\}$ for all (i,j) satisfying $|i-x|<2\sigma$ and $|j-y|<2\sigma$, and

R_w is the set of values $\{I_R(i,j)\exp([-(i-x-\Delta x)^2-(j-y)^2]/2\sigma^2)\}$ for all (i,j) satisfying $|i-x|<2\sigma$ and $|j-y|<2\sigma$.

We refer to the standard deviation σ as the size of the window for that cross-correlator. The function $C(y,\Delta x)$ represents a population of neuronal units tuned to different disparities Δx and vertical image positions y. The preferred disparities used were in the range from -13 to 13 arcmin with a step of 0.6 arcmin (1 pixel in the scaled images), except in the section on "Size-disparity correlation and the disparity gradient limit", where we included window disparities up to 140 arcmin, again with a step size of 0.6 arcmin, in order to examine

performance down to lower frequencies. The step size in the range of y-positions was also 1 pixel in the scaled images.

The innovative feature here is that cross-correlators tuned to larger disparities, i.e. with larger separations between the centers of their left-eye and right-eye windows, had larger windows. Psychophysical evidence for a different sort of size-disparity correlation was provided by Smallman and MacLeod (1994). These authors investigated the optimal disparity at which subjects could perform a front back discrimination task with stereograms based on narrow-band filtered noise. They obtained linear fits between optimal disparity and the center spatial frequency of the noise on a loglog scale. Assuming that cells processing higher luminance frequencies have smaller receptive fields, this provides evidence for a correlation between disparity tuning and receptive field size. It has been pointed out however that there is a possibility that the correlation Smallman and MacLeod obtained could also arise a direct consequence of properties of the stimulus rather than telling us anything new about human depth perception (Prince and Eagle 1999). This is because both D_{min} and D_{max} depend on the center spatial frequency, because fine luminance detail (high frequencies) is required for detection of small disparities while a higher central frequency (less fine detail) means less potential for false matches. We have based the form of the size-disparity correlation we use in our model on Smallman and Macleods results. The fits they obtained for the data from their two different subjects had loglog slopes of approximately -1 and -0.5, corresponding respectively to a linear and a quadratic relationship between size and disparity. Motivated by this, we have examined a second order polynomial as well as a linear function as the relationships between window size and preferred disparity in our model:

$$\sigma = 3 + 0.032 * (\Delta x)^2$$

Equation 2

$$\sigma = 3 + 0.27 * |\Delta x|$$

Equation 3

where σ is the standard deviation of the Gaussian window and Δx is the disparity of the window, both measured in arcmin. We have also explored an exponential size disparity relationship. Although the very long run-time of the simulations made it impossible to perform systematic optimization, or to fit the model results to the data of individual subjects, the size/disparity relationships given in Equation 2 and Equation 3 gave the best match to human performance of those we examined.

The cross-correlator output can be visualised as a two-dimensional image showing correlation as a function of the horizontal disparity, Δx, between the windows as well as the vertical position of the windows, y (see Figure 3). This cross-correlation performs the initial encoding of disparity within the model. Physiologically, we envisage this as occurring in primary visual cortex. The cross-correlation calculated for a given window position, size and disparity represents, in idealised form, the combined activity of several disparity-selective neurons in primary visual cortex, all tuned to the same retinal position and disparity. Each row in Figure 3 represents the activity of a group of V1 neurons tuned to the same retinal location but to a range of horizontal disparities. The black lines indicate how the vertical extent of the window increases with the horizontal disparity to which they are tuned.

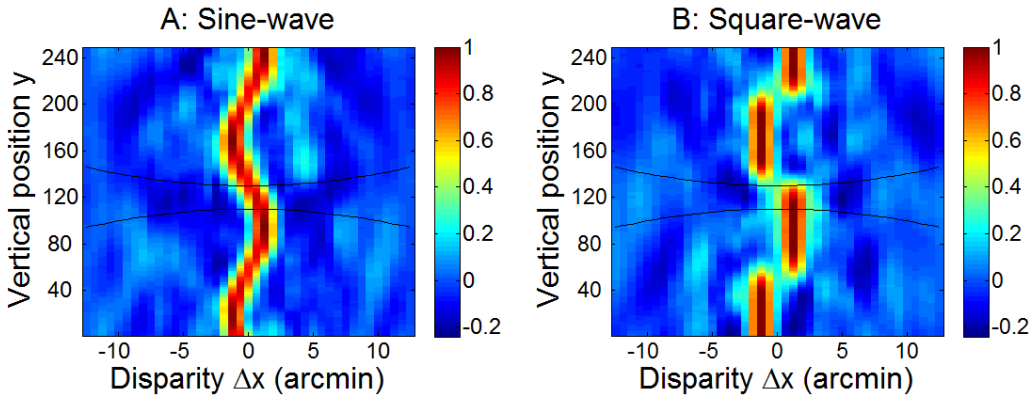


Figure 14: Examples of output from the cross-correlator for one sine-wave and one square-wave disparity grating, both with a frequency of 1.3 cpd. A Gaussian window with σ = 3+0.032*(Δx) 2 arcmin was used. The black lines shows the extent of the correlation window, taken to be the 1SD contour of the Gaussian.

3.2.1.3 Making a perceptual judgment

In order to compare our model to human observers, we needed to take the correlator output from each interval, and use it to make a judgment regarding which interval contained the grating. Physiologically, this process presumably occurs in extra-striate areas, but little is known about how it is achieved. We therefore have little to go on in modelling this process other than some plausible assumptions. In this chapter, we shall ultimately conclude that spatial stereoresolution is fundamentally limited by the initial encoding of disparity in V1, not by the nature of this perceptual read-out process. It is therefore important to demonstrate that our results are qualitatively the same independent of the precise assumptions made regarding read-out. To this end, we have examined three different decision models incorporating specific decision rules, aiming to span a range of possible approaches and assumptions.

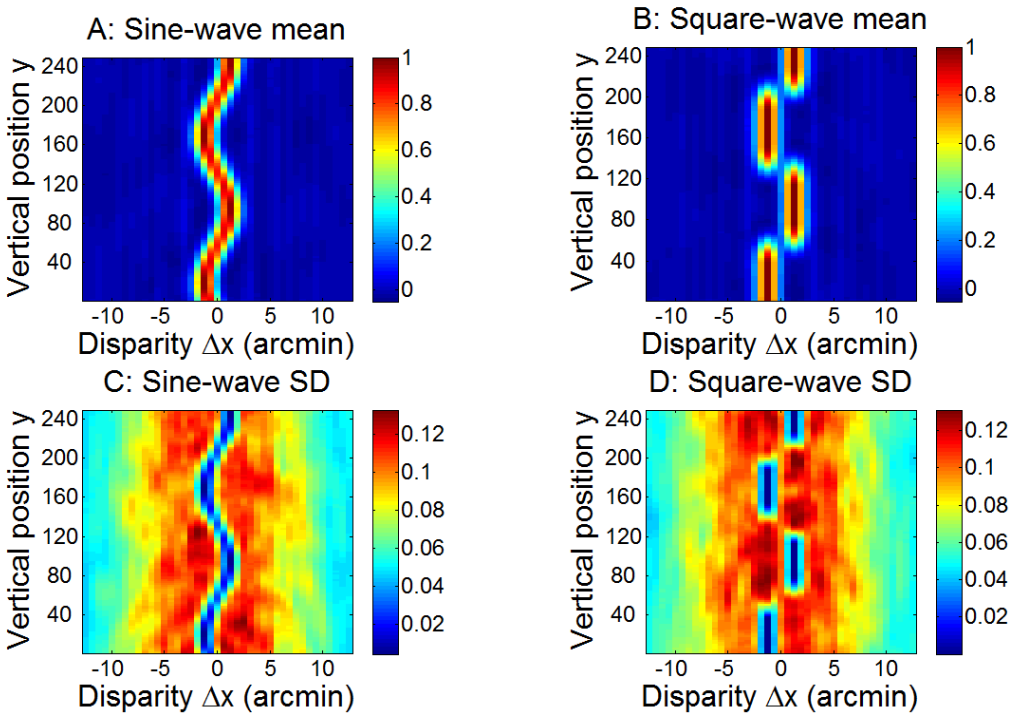


Figure 15: Examples of templates for sine-waves (left) and square-waves (right) with a frequency of 1.3 cpd. The upper row shows the mean and the lower row shows the standard deviation for cross-correlators tuned to vertical position y and disparity Δx, estimated from 100 different random-dot disparity gratings. Figure 3 showed analogous results for a single grating.

The results presented in section 3.3.2 are based on the assumption that the model observer knows the frequency of the grating it is trying to detect, though not the disparity amplitude, waveform (sine vs square) or phase. This is realistic since frequency was blocked in the psychophysical experiments whose results we are trying to reproduce, while amplitude, waveform and phase were interleaved. Avoiding the need to search for frequency speeds up the simulations, but is not critical to our results. In section 3.3.4, we show that very similar results are produced by a model which does not know the frequency.

This method used a set of templates of the correlator output, representing the brain's prior knowledge of the average V1 activity caused by different stimuli. This is closely based on the approach taken by Tsai & Victor (Tsai and Victor 2003). We assume that the brain knows (or is able to reconstruct) the activity expected in response to all the different stimuli used in our experiment, both gratings and noise, based on prior experience. This assumption is discussed further in the Discussion.

The template for each type of stimulus was generated by making 100 different random dot stereograms, preprocessing them with the same preprocessing steps that were used in the main model, and then passing them to the cross-correlator. The mean and standard deviation for each position y and disparity Δx were then calculated based on the resulting set of 100 correlation images (see Figure 15). This process was repeated for gratings of different frequencies, amplitudes, phases and waveforms (sine vs square). The phase of the disparity gratings was varied in steps of 10°. When testing the model, the phase was randomly chosen at each trial to be one of the 36 different phases represented in the set of templates. The disparity amplitudes were 0.3, 1.3, 2.5, 5.1, 7.6, and 10.1 arcmin. Thus there were 432 grating templates per frequency, reflecting 36 phases × 6 amplitudes × 2 grating waveforms. Noise templates were by their nature independent of frequency and phase, so there were 12 noise templates in total, reflecting 6 amplitudes × 2 waveforms.

To simulate an experiment, we assumed that the frequency was known, so the model was using the 432 grating templates for the correct stimulus frequency, as well as the 12 noise templates. In each interval, the correlator output from this stimulus was compared to each

of the 432 grating templates, by calculating the Pearson correlation coefficient between the correlator output and each different grating template (Read 2010). The quality of the match to the best-fitting grating was taken to be

$$M_{grating} = \max_n \frac{\sum((C(\Delta x, y) - \mu_C)(T_n(\Delta x, y) - \mu_{T_n}))}{\sqrt{\sum(C(\Delta x, y) - \mu_C)^2}\sqrt{\sum(T_n(\Delta x, y) - \mu_{T_n})^2}}$$

Equation 4

where C is the correlator output, T_n is the nth grating template, μ_C and μ_{T_n} are the means over all disparities Δx and all y-positions of the correlator output and template T_n respectively, and the sums are over all Δx and all y. The maximum is taken over all values of n, from 1 to 432.

We then calculated the difference ($M_{grating}$-M_{noise}) for each interval, and judged the grating to be in the interval for which this difference was greater.

3.3 Results

3.3.1 *Cross-correlation can be obtained from energy-model units*

The cross-correlation coefficient used in here as well as by Banks et al. differs in a number of ways from the cross-correlation implemented by the energy model. First, it is normalized to lie between 1 (for perfect interocular correlation) and -1 (for anti-correlated stimuli). Second, it operates on the retinal images directly, not the images after filtering by a bandpass receptive field. Finally, the multiplication of the two images is performed first, followed by integration over space, unlike the energy model where the images are integrated over space first and the results are then multiplied together. This has the consequence that the cross-correlation model used here depends more critically on the exact relative positioning of visual features in the two images compared to an energy model unit of the same window-size, and that its disparity tuning is finer and independent of window size. Given that we are claiming our results show that disparity resolution is limited by activity in primary visual cortex, it is important to be clear how the idealized cross-correlation computed in our model relates to more realistic models of individual neurons.

To this end, we begin the Results section by showing that the output of a Banks-style cross-correlator can be approximated by suitably combining the responses of many complex cells tuned to different orientations and frequencies.

In the standard energy model the response of a stereo energy unit is described by the equation:

$$E = (S_{L1} + S_{R1})^2 + (S_{L2} + S_{R2})^2$$

where

$$S_{L1} = \int dxdy I_L(x,y) \exp\left(-\frac{\left((x-x_L)^2 + (y-y_L)^2\right)}{2\sigma^2}\right) \cos(k_x x + k_y y + \phi_L) \text{ and}$$

$$S_{L2} = \int dxdy I_L(x,y) \exp\left(-\frac{\left((x-x_L)^2 + (y-y_L)^2\right)}{2\sigma^2}\right) \sin(k_x x + k_y y + \phi_L)$$

and I_L is the left eye's image, the wavenumbers k_x and k_y together specify the spatial frequency and orientation of the cells receptive field, x_L and y_L specify the position of the center of the left eye's receptive field, ϕ_L is the phase of the receptive field, and σ is the standard deviation of the Gaussian envelope of the receptive field. S_{R1} and S_{R2} are defined analogously. We assume that, due to adaptation at lower levels of the visual system, the image is defined relative to the overall mean luminance, so that averaged across the whole image, $\int dxdy I_L(x,y) = \int dxdy I_R(x,y) = 0$.

Let us assume there are also monocular complex cells which compute

$L = S_{L1}^2 + S_{L2}^2$ and $R = S_{R1}^2 + S_{R2}^2$.

The response of the energy model unit can be split into a binocular part B and monocular parts L and R:

$$E = B + L + R$$

where

$$B = 2S_{L1}S_{R1} + 2S_{L2}S_{R2}$$

Now we compute the total response of all cells at this location which have phase disparity zero and position disparity Δx, summing over cells tuned to a range of spatial frequencies and orientations. In Appendix 1, we show that integrating B in this way over all spatial

frequencies and orientations gives us

$$B_{\text{int}} = 2\int dx'dy' \exp\left(-\frac{\left((x'-x)^2+(y'-y)^2\right)}{2\sigma^2}\right) I_L(x',y') \exp\left(-\frac{\left((x'-x-\Delta x)^2+(y'-y)^2\right)}{2\sigma^2}\right) I_R(x',y')$$

Approximating the integrals with a sum over pixels, and using L_W to represent the image after multiplication by the window function, this is

$$B_{\text{int}} = 2\sum_{i,j} L_w(i,j) R_w(i,j).$$

This is simply the covariance of the weighted image-patches, plus a term reflecting the average pixel-value within the window:

$$B_{\text{int}} = 2n\left[\text{cov}(L_w, R_w) + \bar{L}_w \bar{R}_w\right]$$

where *n* is the total number of pixels included in the sum. Similarly, integrating the monocular terms over all spatial frequencies and orientations, we obtain

$$L_{\text{int}} = n\left[\text{cov}(L_w, L_w) + \bar{L}_w{}^2\right] \text{ and } R_{\text{int}} = n\left[\text{cov}(R_w, R_w) + \bar{R}_w{}^2\right]$$

Now we use the monocular terms to normalise the binocular term (Tsai and Victor 2003; Read and Cumming 2006; Read 2010):

$$C_{\text{int}} = \frac{2B_{\text{int}}}{\sqrt{L_{\text{int}} R_{\text{int}}}} = 2\frac{\text{cov}(L_w, R_w) + \bar{L}_w \bar{R}_w}{\sqrt{\left[\text{cov}(L_w, L_w) + \bar{L}_w{}^2\right]\left[\text{cov}(R_w, R_w) + \bar{R}_w{}^2\right]}}$$

The normalisation ensures that C_{int} remains between +1 (for units tuned to the stimulus disparity, where L_w=R_w) and -1 (for anti-correlated stimuli, where L_w=-R_w).

For random-dot patterns where the correlation window is large compared to the dot-size, the average pixel-value within each eye's window will be very nearly the same as the average pixel-value across the whole eye's image, which is zero by definition. For such images, C_{int} reduces immediately to C as defined in Equation 1. For natural scenes or other images where the luminance undergoes large-scale changes across the image, this would not be the case, and C_{int} would not be zero for binocularly uncorrelated images. Real neurons have not been studied with such images, so it is not possible to say whether C_{int} or C as defined in Equation 1 would be more appropriate in that case.

This analysis shows that the key features of the Banks model – units sensitive to the precise location of features within the window, isotropic windows, disparity tuning curves whose width is independent of window size – can be produced within a more physiologically-realistic model, simply by combining the outputs of energy-model units tuned to many spatial frequencies and orientations. Essentially, the Banks model is a computational short-cut which enables us to approximate the properties of a much larger population of energy-model units at vastly reduced computational cost. This is somewhat analogous to how the energy-model itself uses a quadrature pair of units with 0 and $\pi/2$ phase to approximate the output of a large number of subunits tuned to a range of phases. This derivation gives us confidence that the encoding stage of our model, while clearly highly idealised, is nevertheless consistent with the physiology of early visual cortex.

One important limitation of the analysis performed in this section is that this analytical proof only works if sigma does not depend on (luminance) spatial frequency or orientation. This limitation is discussed further in section 3.4.5 where various other limitation of the model are also discussed. It is quite possible that the equivalence between combined energy model units and local cross-correlation may still hold to good approximation even with a certain dependence between window size and spatial frequency, at least there is no immediately obvious reason why this could not be the case. In section 5.2 there is a brief discussion of how this could be tested with simulations.

We now move on to examine how the model performs when its outputs are used to perform our psychophysical task, under various different decision models.

3.3.2 *Size-disparity correlation makes sine- and square-wave gratings equally detectable*

Figure 16 shows the results of the model. Panels A-H show the model's performance (percent correct judgments) as a function of disparity amplitude for different grating frequencies and the final panel shows the maximum performance, i.e. that at the optimal disparity amplitude for each frequency, as a function of frequency. Red circles show results for sine-wave gratings; blue squares those for square-wave gratings. Throughout, error bars show 95% confidence intervals. Critically, the results are now very similar for both sine-

and square-wave disparity gratings – like human observers and unlike the original model (Figure 16). Like human observers, as disparity amplitude increases beyond its optimal value, performance for both grating waveforms decays back to chance.

Similar figures are given in section 3.3.4 and section 3.3.5 for alternative decision models (Figure 18 and Figure 20). Unsurprisingly, there are quantitative differences between the results from different decision models, especially in the percent correct at the lowest disparity amplitude. This amplitude, 0.3 arcmin, is below the step size of 0.6 arcmin in the range of correlation detectors, and the decision models vary in how efficient they are at extracting information at this sub-step-size disparity. Similarly, the decision models vary somewhat in the frequency at which peak performance first starts to decline. We know in principle how to match human performance on both of these. Capturing sensitivity to small disparity amplitudes would require the right minimum spacing in the population of cross-correlators, plus the addition of noise to limit the ability to discriminate tiny disparities. Capturing the correct frequency at which performance declines would require us to tweak the minimum window-size, i.e. the value of the first term in Equation 3, as done by Banks et al (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). Given the long simulation run-time and the fact that these issues are solved in principle, we have not here attempted to chase down these parameters further.

In Figure 20, showing results for a decision model based on auto-correlation, there are a couple of frequencies where performance starts dropping for the sine-waves at slightly lower amplitudes than for the square-waves. Interestingly, 2 of our 4 human observers also displayed this tendency (Figure 9 in chapter 2), while neither humans nor model ever displayed an earlier drop for square-waves than for sine-waves.

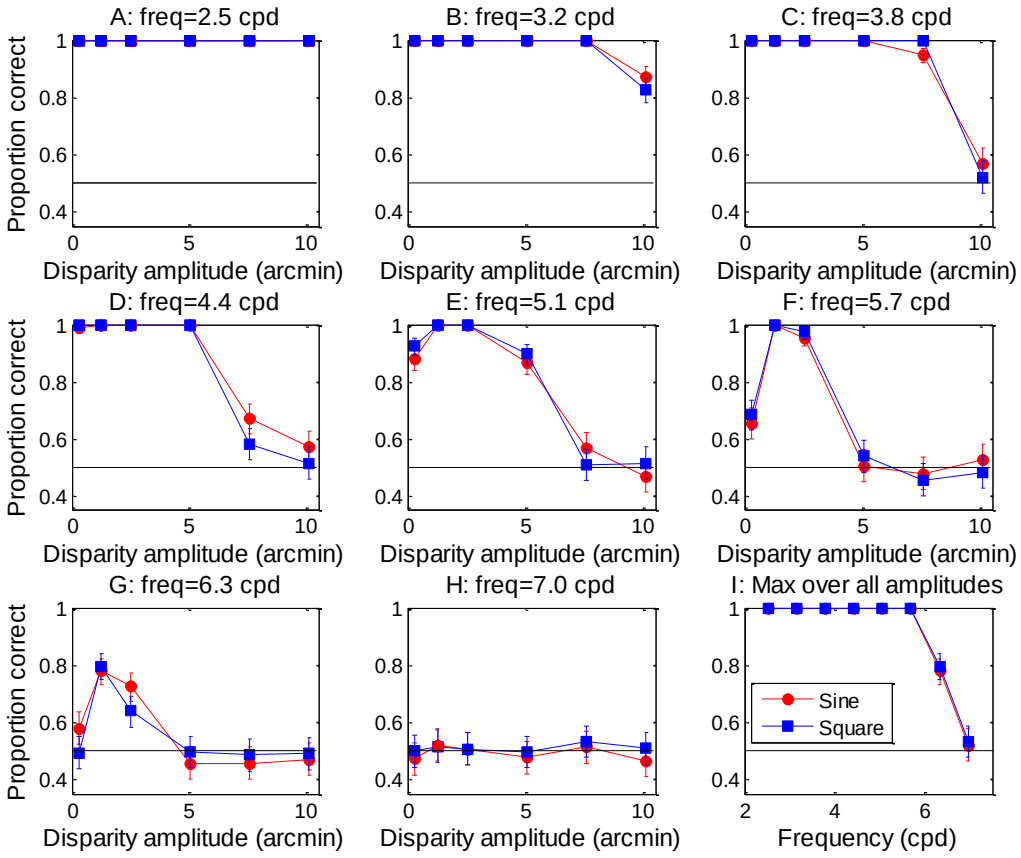


Figure 16: Model performance on the grating detection task as a function of amplitude and frequency. The last plot (I) shows the maximum performance over all amplitudes for each frequency. This is for the model with the template matching decision model with known frequency and a quadratic size-disparity relationship (Equation 2).

3.3.3 *Form of the size-disparity correlation is not critical*

The results in Figure 16 assumed a quadratic relationship between a correlator's window-size and its preferred disparity. The psychophysical data suggests there may be noticeable inter-subject variation in the relationship between spatial scale and disparity correlation, with Smallman & McLeod's two subjects showing linear and quadratic relationships respectively. However, all our subjects showed near-identical performance on sine- and square-wave gratings (Allenmark and Read 2010). We therefore wanted to check that the precise form assumed for the size-disparity correlation was not critical for our results. To this end, we also tested the model with a linear size/disparity correlation (Equation 3). The

results (Figure 17) are similar to those obtained with the second order polynomial size/disparity correlation (Equation 2), and in particular the key result holds: differences between the sine-wave and square-wave results remain negligible. This suggests that several different forms of the size/disparity correlation may be consistent with the human data in the previous chapter.

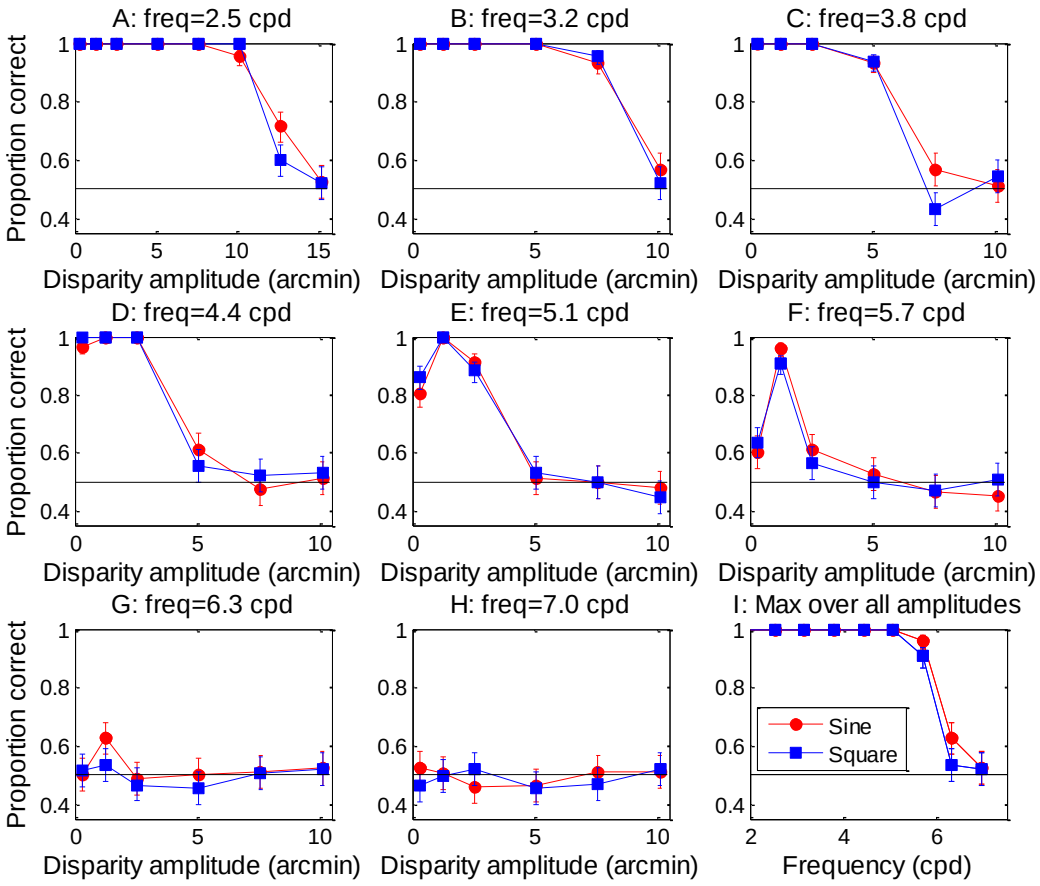


Figure 17: As for Figure 16 but with a linear size-disparity relationship (Equation 3).

3.3.4 *Decision model using template matching with unknown frequency*

This method was the same as that described in section 3.2.1.3 on making a perceptual judgement, except that $M_{grating}$ was calculated for templates of all frequencies, including two frequencies (1.9 cycles/degree and 7.6 cycles/degree) for which no results are shown (because the model performed either perfectly or at chance), not just the 432 with the correct stimulus frequency. The results are shown in Figure 18.

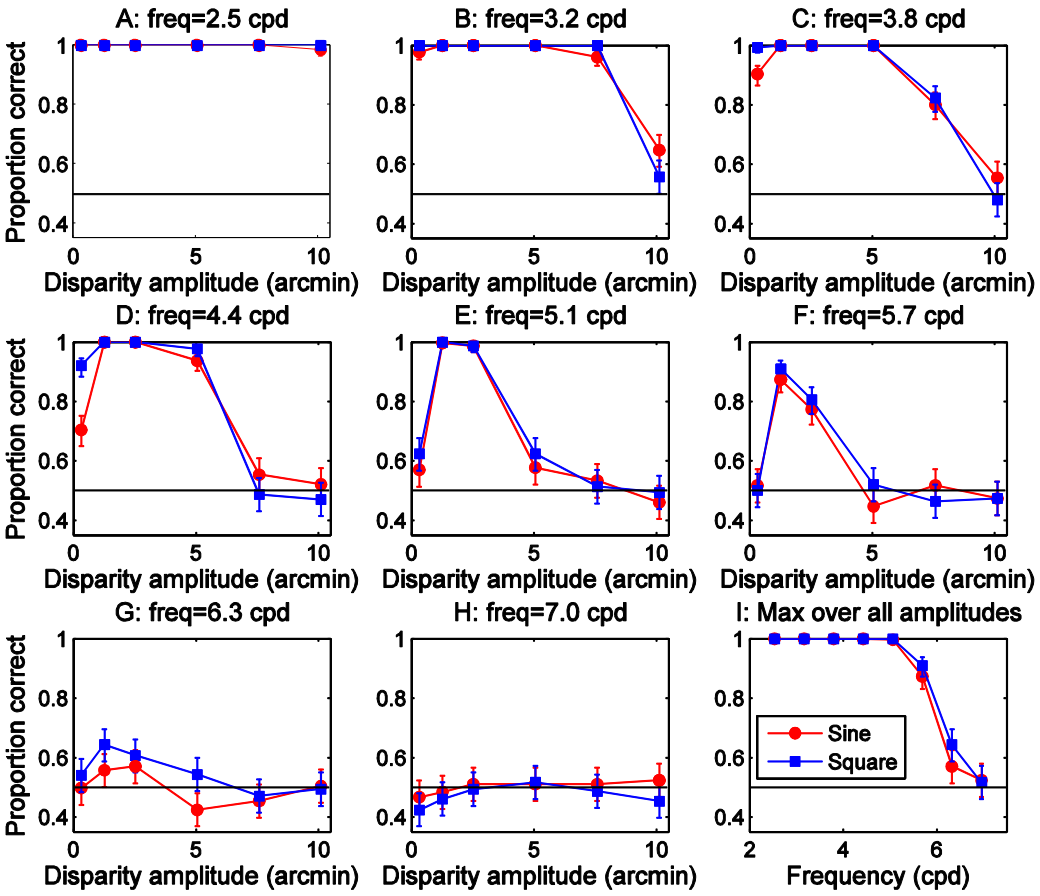


Figure 18: Model performance on the grating detection task as a function of amplitude and frequency. The boxed plot (I) shows the maximum performance over all amplitudes for each frequency. This is for the model with the template matching decision model with unknown frequency and a quadratic size-disparity relationship (Equation 2 in the main document).

3.3.5 *Decision model using autocorrelation*

Here we show results from a decision model which does not use template-matching at all, but detects the grating from the autocorrelation of the disparity map. In this decision model, we start by estimating disparity at each point on a vertical line down the image. We do this by finding the peak correlation on each row of the correlation images that were the output from the cross-correlator. The disparity at which the peak correlation was found at each row was recorded as an estimate of the horizontal disparity at the corresponding vertical position:

$$\Delta x_{est}(y) = \arg\max(C(y, \Delta x))$$

Equation 5

The result was a curve of estimated disparity as a function of vertical position (see Figure 19AB). The next step was to calculate the autocorrelation of this curve as:

$$a_n = \frac{\sum_{i=1}^{N-n} (\Delta x_{est}(y_i) - \mu)(\Delta x_{est}(y_{i+n}) - \mu)}{((N-n)*\sigma^2)}$$

Equation 6

where μ is the mean and σ is the standard deviation of Δx_{est}, the sum is over all vertical positions *i* and n is the lag. Figure 19CD shows two examples of auto-correlograms. The last step was to fit a sine-wave and a triangular wave, the auto-correlation function of a square-wave, with the same frequency used in the stimulus to the auto-correlogram and record the r^2-value of the best fit. Only the frequency of the gratings was given to the model. The amplitude was acquired by choosing the amplitude which gave the best fit and the model did not need to know the phase since the autocorrelation function is largely independent of phase. Letting the model know the frequency of the gratings was motivated because we had kept the frequency constant in each block of trials in the psychophysics experiments. The decision on which of the two image pairs given to the model in any trial contained the grating was then made by choosing the image pair which gave the highest r^2-value.

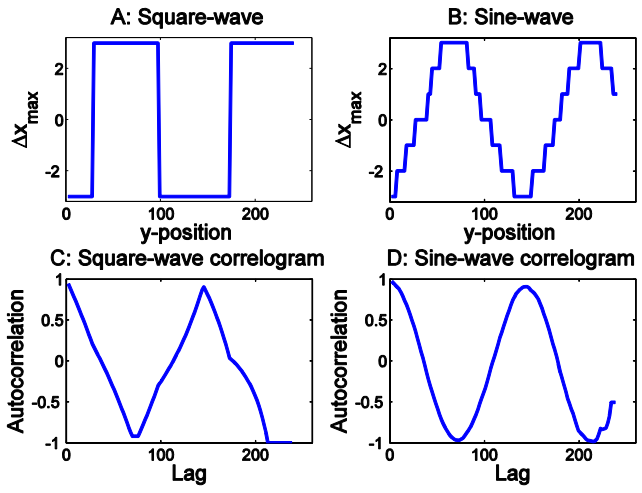


Figure 19: Examples of estimated disparity curves and their autocorrelograms for one square-wave and one sine-wave both with a frequency of 1.3 cpd. A Gaussian window with $2*\sigma = 6+0.063*(\Delta x)^2$ arcmin was used. The estimated disparity curve for the sine-

grating is quantized because the model only included detectors tuned to integer disparities (in pixels).

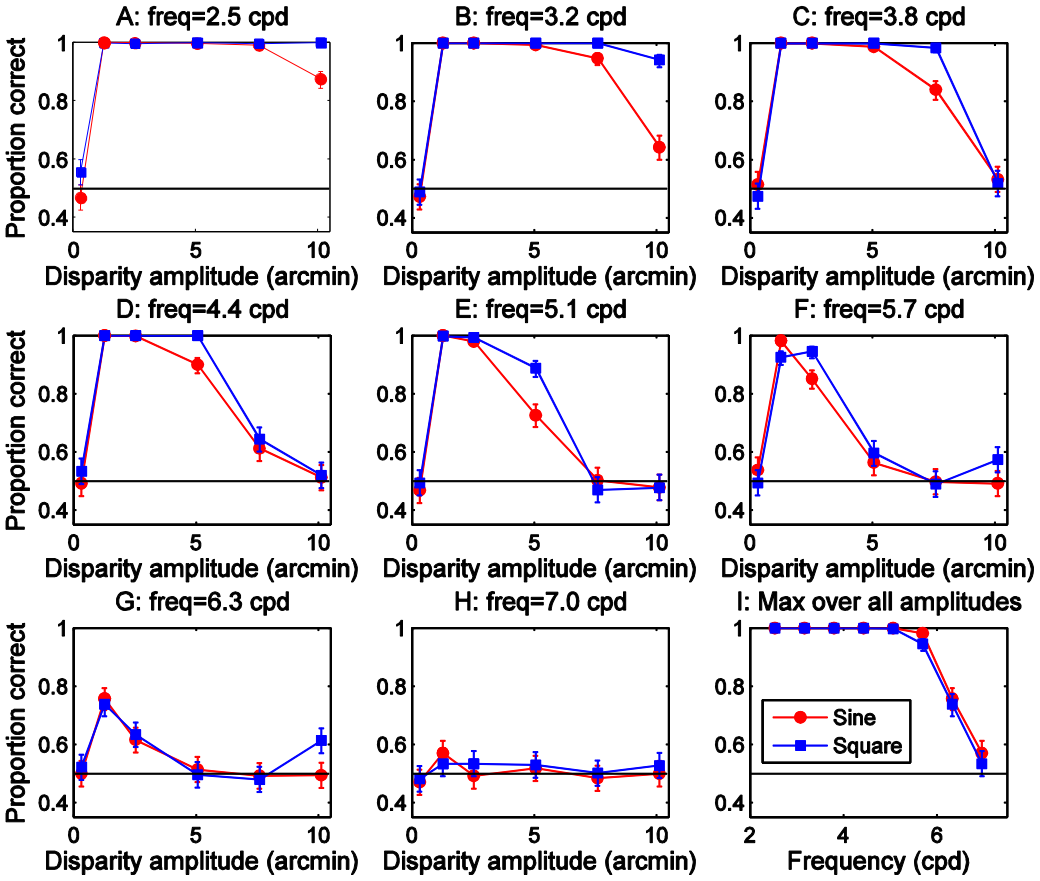


Figure 20: Model performance on the grating detection task as a function of amplitude and frequency. The boxed plot (I) shows the maximum performance over all amplitudes for each frequency. This is for the model with the autocorrelation based decision model and a quadratic size-disparity relationship (Equation 2).

3.3.6 *Model with size-disparity correlation explains disparity gradient limit for sine and square-wave gratings*

Many previous studies have suggested that human depth perception is limited in the disparity gradients it can detect (Tyler 1975; Burt and Julesz 1980; Kanade and Okutomi 1994; McKee and Verghese 2002; Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). For example, Tyler found that, for sinusoidal disparity gratings, the highest disparity amplitude which can be perceived is inversely proportional to grating frequency (i.e. lies on

a line with a slope of minus one in log-log coordinates (Tyler 1975); black symbols in Figure 21), as if perception is limited by the maximum gradient present in the grating. This observation does not require a size-disparity correlation; for example, Filippini & Banks (Filippini and Banks 2009) successfully reproduced it with their local cross-correlation model which incorporates no relationship between size and disparity tuning of detectors (Figure 21A). However, Tyler also found the same relationship between upper depth limit and frequency in square-wave disparity gratings. He argued that this does imply a size-disparity correlation. No computational model has yet reproduced this observation. To examine this, we re-ran our simulations using a larger range of correlation detectors, including detectors tuned to disparities up to 140 arc min. This enabled us to probe the model's upper depth limit even at frequencies <1 cpd, where performance remains perfect up to tens of arc min.

The coloured symbols in Figure 21 shows the upper limit of disparity amplitude, defined as the maximum amplitude for which performance exceeds 80% on our grating detection task, as a function of grating frequency. For comparison, Tyler's results are replotted in black. Figure 21A shows our results with the original, constant window-size model. For sinusoidal disparity gratings, the upper limit falls as a power-law with frequency, replicating the finding of Filippini & Banks. However, the model fails completely for square-wave gratings. No results are shown since the model has no upper depth limit for square-wave gratings; performance remains optimal at all amplitudes up to Panum's fusional limit, with no trade-off between upper depth limit and frequency. This is inconsistent with Tyler's data showing that, for human subjects, the upper depth limit for square-waves falls with increasing frequency in the same way as it does for sine-waves (Tyler 1975), as well as with our own data (Allenmark and Read 2010).

Figure 21B shows the results of the new model using a linear size/disparity correlation (Equation 3). For both square-wave and sine-wave gratings, the upper depth limit is inversely proportional to frequency, in agreement with the human data. However, in the model results the sine- and square-wave curves overlap almost perfectly while they are offset by a constant amount in Tyler's data. Tyler's data were obtained using a different stimulus, line stereograms rather than random dot stereograms, and while similar results

have also been obtained with random dot stereograms for sine-waves (Tyler 1974), to our best knowledge the frequency dependence of the upper depth limit for square-waves has only been measured with line stereograms, making it hard to say whether this difference reflects a real problem with the model or if it is just a consequence of using a different stimulus. In the human data presented in the previous chapter, some subjects seem to show a difference in the same direction as Tyler, though smaller, while others show almost no difference. But there we only looked at high frequencies and the experiments were not designed specifically to test the upper disparity limit. Clearly, more data on the upper disparity limit for sine- vs. square-wave disparity gratings in random dot stereograms would be needed to test whether the lack of an offset between the sine- and square-wave results reflects a remaining problem with the model.

Figure 21C shows the results of the new model using a quadratic size/disparity correlation (Equation 2). The results for sine-waves and square-waves are again very similar, but now the upper depth limit rises less steeply as frequency is reduced, or put another way, the highest frequency detectable for a given amplitude decreases at an accelerating rate as the amplitude increases.

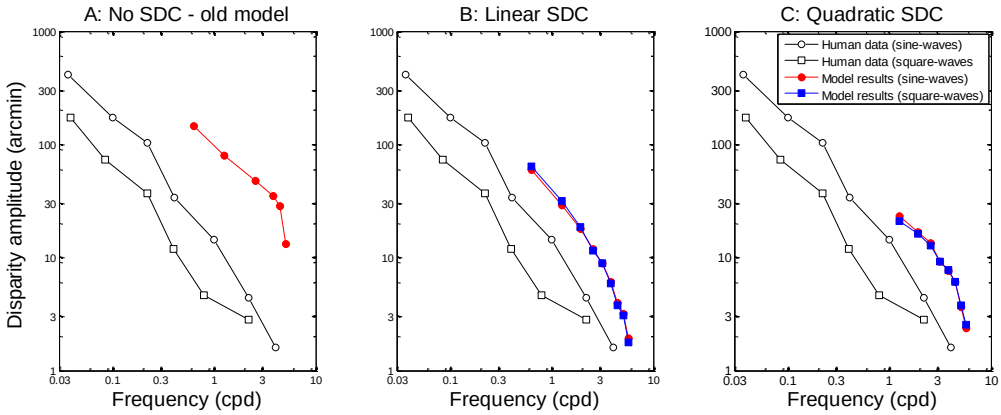


Figure 21: The maximum amplitude at which sine- and square-wave disparity gratings can be detected with >80% accuracy, as a function of frequency. The black squares and circles show human data for square- and sine-waves replotted from Tyler (Tyler 1975). The red circles show model results on sine-waves and the blue squares show model results on square-waves. A: Results with the old constant window-size model. No square-wave results are shown because the constant window-size model does not have an upper depth limit for square-waves. B: Model results using a linear size/disparity correlation in the encoding population (Equation 3). C: Model results using the same decision model but a quadratic size/disparity correlation (Equation 2).

3.4 Discussion

The idea of primary visual cortex as a cyclopean retina goes back to Julesz (1971). Recently, the suggestion has emerged that certain key aspects of human depth perception, notably the low spatial resolution for stereo depth, are set by the initial encoding of disparity in primary visual cortex (V1). This suggestion has been quantified with models closely based on known physiology, in which disparity is encoded via a local cross-correlation of the two eye's images, within a finite window (Kanade and Okutomi 1994; Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). In chapter 2 we identified a problem with the current implementation of this model. The model predicts a difference between the detectability of sine- vs square-wave gratings which is not observed in humans. The model predicts that, for sine-wave gratings, performance should decline from its peak value as disparity amplitude increases, while for square-wave gratings, performance should remain high. In humans, performance declines for both types of gratings. Clearly, the model needed to be altered to account for these observations.

This then raised the question of what sort of modifications were needed. Potentially, the discrepancies might reflect the model's failure to include more elaborate disparity processing in extra-striate cortex. For example, some extra-striate areas contain neurons that are tuned to disparity-defined edges, slant and curvature (Janssen, Vogels and Orban 1999; Sakata, Taira, Kusunoki et al. 1999; von der Heydt, Zhou and Friedman 2000; Sugihara, Murakami, Shenoy et al. 2002; Nguyenkim and DeAngelis 2003; Bredfeldt and Cumming 2006). These are not included in the model. If such extra-striate mechanisms turn out to play a critical role in setting spatial stereoresolution, this would undermine the claim that stereoresolution is limited by the initial encoding of disparity performed in striate cortex. However, current models also ignore many known features of primary visual cortex, partly for practical reasons (simulation runtimes rapidly become unmanageable if one attempts to include all known variations) and partly for theoretical ones (insight is gained by abstracting out the key features which are responsible for a particular behaviour). Thus, it seemed to us that the first line of inquiry should be to explore whether a more realistic

representation of the initial disparity encoding stage could reconcile the model with human behaviour.

One obvious property neglected by the current model is the tuning of neurons in early visual cortex to luminance spatial frequency and orientation. Rather, as we have shown in the first section of the Results, the model's idealised, isotropic cross-correlators represents the combined output of many such tuned neurons (as for example in (Read and Cumming 2006)). For the broad-band random-dot patterns used here, we believe that this simplification is adequate, and unlikely to affect the model's performance on the particular tasks under consideration. We therefore chose to address, instead, another property ignored by current models, namely the size/disparity correlation. Much previous psychophysical work has indicated a correlation between the spatial scales over which disparity is extracted, and the amplitude of the disparity itself (Tyler 1975; Smallman and MacLeod 1994; McKee and Verghese 2002; Tsirlin, Allison and Wilcox 2008). Physiologically, this implies that a population of neurons tuned to low spatial frequencies would encode disparities over a larger range than a population with tuned to high spatial frequencies. There is some physiological evidence supporting this (Prince, Cumming and Parker 2002). In the correlation model, spatial frequencies are not explicitly represented, but the integration implicitly includes all spatial frequencies with the same weighting (a limitation we discuss further below). Thus it is difficult to incorporate a relationship between disparity and spatial frequency tuning. However, it is easy to incorporate a relationship between disparity and receptive field size. We believed that such a size/disparity correlation could potentially account for the poor human performance on square-wave gratings. Our reasoning was that square-wave gratings present a greater magnitude of disparity, averaged across a cycle, than sine-wave gratings of the same amplitude. Thus, their disparity should be encoded by cross-correlators with larger average window-size than sine-wave gratings. When the window-size associated with the largest disparity in the grating is comparable to or larger than half the spatial period of the grating this effect will tend to reduce performance on square-wave gratings relative to sines, although the piecewise-frontoparallel nature of square-wave gratings will tend to enhance performance relative to sines. We wondered whether, with an appropriate relationship between window-size and

disparity magnitude, these two effects could cancel out and thus account for the very similar human performance on both types of gratings.

Here, we have shown that our intuition was correct. Introducing a size/disparity correlation into the initial stage of disparity encoding, such that larger disparities are detected using larger correlation windows, solves both of the problems we identified with earlier version of the model. We have investigated various decision models, and shown that the model's performance does not depend critically on the particular decision model used. Rather, it reflects the information available at the initial encoding stage, for the reasons we now discuss.

3.4.1 *Why a size-disparity correlation reconciles the model with human performance on square-wave gratings*

Correlation-based models are built of disparity detectors which respond maximally, i.e. with correlation output 1, to uniform stimulus disparity at their preferred value. Stimulus disparities away from the preferred value cause a decline in the reported correlation output. In this type of model, the rate of the decline is ultimately limited by the point-spread function of the eye, with an SD of around 2 arcmin.

In the old, fixed-window-size model, the quality of the correlator output declines with increasing amplitude for the sine-waves, but not for the square-waves. Figure 22 shows examples of the old model's correlator output for sine- and square-waves with low and high amplitude, for a frequency of 3.8 cpd. The white lines show which disparity was actually presented at each vertical position. The black lines show the extent of a correlation window, which for purposes of discussion we will take to be the 1SD contour of the Gaussian. For the low amplitude gratings (Figure 22AB), the correlator output is of high quality for both waveforms. It is maximal at the front and back surfaces of each waveform, where the range of stimulus disparities within the correlation window is smallest. In this example, the grating half-period is 7 arcmin, so for the square-wave, detectors positioned at the center of the grating's front and back surfaces experience uniform stimulus disparity everywhere within their 6-arcmin correlation window. Detectors tuned to the stimulus disparity will therefore respond close to their maximum possible value of 1. Even at the

edges of the square-wave the window will only experience two disparities, each covering half the window, allowing the correlation to be relatively high (close to 0.5) for detectors tuned to either of these two disparities. For the sine-wave, the stimulus disparity is constantly varying. However, detectors positioned at the peak and trough of the gratings experience only a small (0.8-arcmin) range in disparity within their correlation window, so the response is still high at the front and back surfaces. Even detectors at the centre of the grating (zero disparity) experience a range of only 2.4-arcmin disparity, and so give a clear, though reduced, response.

For the high-amplitude sine-wave grating, Figure 22C, the situation is very different. Detectors at the centre of the grating now experience a 14-arcmin range of stimulus disparities. There is thus almost no visible response to the slanting regions of the grating which can be distinguished from chance responses to particular random dot patterns within the stimulus. Detectors centred on the peaks and troughs of the sine-wave experience a lower disparity range of 4.8 arcmin, and periodic blobs of higher activation are still just visible here. Thus overall, the high-amplitude sine-wave grating is barely visible in the correlator output. For the high-amplitude square-wave, Figure 22D, little is changed compared to the low-amplitude case, Figure 22B. Detectors in the center of the grating's front and back surfaces still experience uniform disparity, and so their response is undiminished. Detectors at the edges of the square-waves still only experience two disparities. That these are now further apart makes no difference, each disparity is still seen by half the window allowing correlations of about 0.5 even close to the edges. This is why the old model performed so much better with high-amplitude square-waves than with sines (Figure 7 and Figure 8).

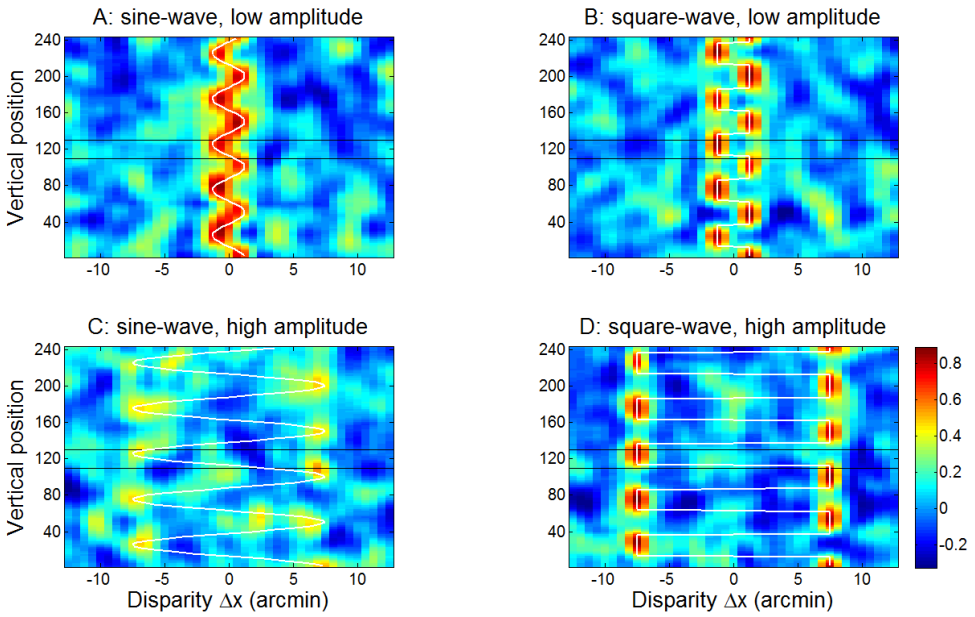


Figure 22: Examples of output from the cross-correlator for the old model at a frequency of 3.8 cpd. The top row shows output for a sine-wave (A) and a square-wave (B) with low amplitude (4 pixels = 1.3 arcmin) while the bottom row show output for a sine-wave (C) and a square-wave (D) with high amplitude (24 pixels = 7.6 arcmin). Notice that the quality of the correlator output remains high for the high amplitude square-wave (D) while only the regions close to the peaks are visible in the output for the high amplitude sine-wave (C).

How does the size/disparity correlation change matters? Figure 23AB shows correlator output for our new model, for high amplitude sine- and square-waves at 3.8 cpd, the same frequency that was used in Figure 22. For the low amplitude gratings, the correlator output remains almost exactly the same as shown in Figure 22AB, since the window-size remains close to that used in the fixed- window-size model. For high-amplitude gratings on the other hand, considerably larger windows will be used to detect the large disparities, as indicated by the black lines. For sine-wave gratings, this has relatively little effect. Detectors at the peaks and troughs of the grating now have a window-size of 2σ=10 arcmin. The range of disparity they experience within their correlation window is therefore larger, at 10.7 arcmin as compared to 4.8 in Figure 22C. The correlation output in Figure 23A is therefore somewhat reduced compared to the old model, Figure 22C (note slightly different colorscale), but the grating is still visible in the periodic "blobs" of higher correlation. For the square-wave, on the other hand, the increase in window-size has a more serious effect.

The window now exceeds the grating half-period, meaning that correlation detectors at the middle of the front or back surfaces no longer sample only their preferred disparity, but also some disparities 15 arcmin away from their preferred value. Detectors at different vertical positions now vary only in the proportion of dots which are at their preferred disparity. Accordingly, not only are the "blobs" marking each front and back surface now lower in amplitude, but critically, they are no longer separated by clear regions of low activation (compare Figure 23B vs Figure 22D).

This is very damaging to the model's performance. Recall that, in order to assess spatial resolution, observers were asked to discriminate stimuli in which disparities were arranged as a periodic function of position (gratings) from those in which the same disparities were scattered at random (noise). Figure 23CDEF shows the mean correlator output for both types of stimuli: that is, the grating templates for this frequency and amplitude (Figure 23CD), and the noise templates for this amplitude (Figure 23EF). The model's task, then, is essentially to decide whether the output to a given stimulus, Figure 23A and B, is a better match to the grating templates in Figure 23CD or to the noise templates in Figure 23EF. These are distinguished only by their periodicity.

For the square-wave grating, the periodicity was perfectly clear with the fixed-window-size model (Figure 22CD), and is much less obvious with the size-disparity correlation model (Figure 23AB), thanks to the larger window sizes at the relevant disparities. In the new model, both the sine-wave and the square-wave output is now hard to distinguish from the noise patterns. This is why all our decision rules gave similar results for both square-wave and sine-wave gratings. For the frequency and amplitude used in this example, the template matching decision rule with known frequency performed at about 80% correct for both.

In order to gain a better understanding of what happens when the model fails we have also approached this question in a different way. The best matching template for the grating interval was recorded at each trial using both sine-waves and square-waves and two different combinations of frequency and amplitude: 2.5 cpd and 4 arcmin where the model performed at 100% correct and 6.3 cpd and 4 arcmin where the model performed at roughly 80% correct. This was done for 100 trials of each of these combinations. At the lower

frequency the model always picked the right template. At the higher frequency the model always picked a template with the right waveform and amplitude but only got the phase right roughly 15% of the time. However, it picked a phase within 30° of the correct one roughly 60% of the time and a phase within 80° of the correct one roughly 95% of the time. It seems then that what happens when the task starts to get harder is that the model's estimate of the phase of the grating gradually gets worse. This makes sense intuitively when looking at Figure 23. The responses of the model to gratings of different wave-form look quite different even at a high frequency. The same is true of the model's responses to gratings of different amplitude. However, as the correlator output starts to get more similar to the noise templates, which of course are independent of phase, it naturally also starts to become more similar for gratings with different phases.

The grating and noise templates shown in Figure 23CDEF also allow us to make a prediction about how the model would perform on a discrimination task where it has to discriminate between sine- and square wave disparity gratings. Since both the high frequency grating templates and the noise templates look quite different for the sine- and the square-waves it seems likely that the model would be able to discriminate between sine- and square-wave gratings up to any frequency. At high frequencies the gratings would of course become indistinguishable from noise, but if the two types of noise are distinguishable from each other, which is suggested by the dissimilar sine- and square-wave noise templates, then this would not be a problem. This is also supported by the results presented in the previous paragraph, where the model always picked a template with the correct waveform for the grating interval, even when it was only performing at 80% and starting to get the phase of the template wrong. However, this ability of the model to discriminate sine- and square-wave gratings at any frequency is quite different from what would be predicted by the frequency analysis view discussed in section 1.2.1.1. The frequency analysis view predicts that sine-waves and square-waves should become indistinguishable when the second harmonic of the square-waves is no longer individually detectable which of course means that sine- and square-wave gratings should be indistinguishable at high frequencies. It therefore seems like the model is clearly inconsistent with the frequency analysis point of view (as applied to the disparity domain). Based on my experience with the stimuli it seems like it is possible for humans to

distinguish the two types of noise stimuli as well. This would of course need to be confirmed by further experiments before any more certain conclusions can be drawn, but if it is true it would mean that human perception is consistent with the behavior of the model and inconsistent with the frequency analysis point of view.

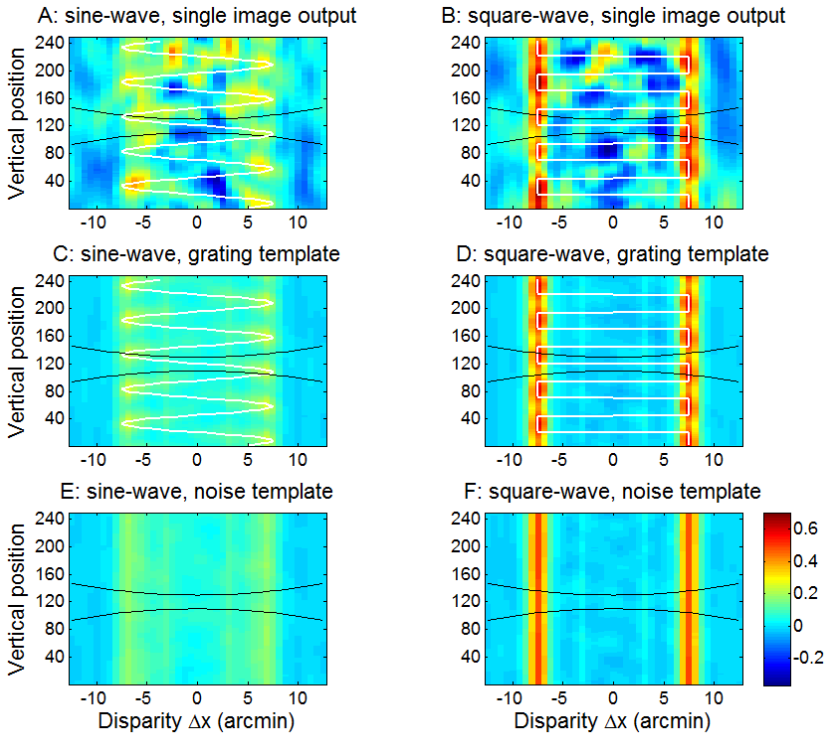


Figure 23: The top row shows examples of output from the cross-correlator for the new model at a frequency of 3.8 cpd for sine-waves (A) and square-waves (B). The middle row shows grating templates at the same frequency for sine-waves (C) and square-waves (D). The bottom row shows noise templates for sine-waves (E) and square-waves (F). The correlator output matches the grating templates better than the noise templates.

3.4.2 *Initial encoding not decision rule is critical*

Although we have concentrated on the template-matching decision model when explaining why the size-disparity correlation has the effect it does, qualitatively similar results were obtained from all four decision models examined. We conclude that stereoresolution is limited by the initial encoding of disparity, not by the particular read-out we have adopted. Similar conclusions were reached by Banks (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009) and Harris et al (Harris, McKee and Smallman 1997).

3.4.3 *Size-disparity correlation and the disparity gradient limit*

Previous studies have suggested that our perception of depth patterns containing a large range of disparities may be limited by disparity gradient rather than the large disparities as such (Burt and Julesz 1980; Kanade and Okutomi 1994; McKee and Verghese 2002; Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). In particular a study by Tyler (Tyler 1975) found that the maximum depth limit, the disparity amplitude at which depth differences are no longer perceived in sinusoidal and square-wave disparity gratings, depends on corrugation frequency in a way that approximately corresponds to a straight line with slope -1 in log-log coordinates. Banks et al. (Filippini and Banks 2009) had previously shown that a constant window size local cross-correlation model performed in a qualitatively similar way when tested with sinusoidal disparity gratings. Here, we have replicated this finding and shown that when a size/disparity correlation is incorporated into the model it performs in the same way for square-wave disparity gratings, consistent with Tyler's results. The model achieves this despite lacking any sensors tuned to non-zero disparity gradients. Banks et al. suggested that the disparity gradient limit was a by-product of using local cross-correlation to estimate disparity (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). However, as Tyler (1975) recognized, this alone cannot explain why the frequency dependence of the upper depth limit exists for square-waves as well as for sine-wave gratings. We have found that incorporating a size/disparity correlation into a correlation-based model makes it perform consistently for random-dot patterns depicting both square-wave and sine-wave disparity gratings. This supports Tyler's conclusion (1975) that the disparity gradient limit reflects a size/disparity correlation, rather than being solely a by-product of local cross-correlation.

3.4.4 *Relationship to previous models*

Models of stereopsis based on cross-correlation of local patches of the two eyes' images have a long history (Hannah 1974; Panton 1978; Kanade and Okutomi 1994; Steingrube, Gehrig and Franke 2009). They are widely used in computer vision as a fast and relatively reliable approach of achieving stereo correspondence. They have often been used to model human vision (Cormack, Stevenson and Schor 1991; Harris, McKee and Smallman 1997; Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). Local cross-correlation is closely related to the "stereo energy" computation performed by cells in primary visual

cortex (Ohzawa, DeAngelis and Freeman 1990; Fleet, Jepson and Jenkin 1991; Qian and Zhu 1997; Qian and Mikaelian 2000), although cells spectrally filter the local image patches before cross-correlating them. Models based on stereo energy units have also been used as models of human vision (Qian 1994; Fleet, Wagner and Heeger 1996; Qian and Zhu 1997; Tsai and Victor 2003; Read and Cumming 2006; Read and Cumming 2007). All these implementations have recognized that useful disparity estimates require the outputs of many stereo energy units to be combined in some way. For example, models have estimated disparity by combining the outputs of stereo energy units with different spatial locations (Qian and Zhu 1997; Read and Cumming 2004), or different spatial frequencies and/or orientations (Fleet, Wagner and Heeger 1996; Read 2002; Read and Cumming 2006). As we have shown in this chapter, combining stereo energy units tuned to many different spatial frequencies and orientations can produce something which is formally identical to local cross-correlation of the unfiltered image.

Stereo energy units based on phase disparity (Deangelis, Ohzawa and Freeman 1991; Fleet, Jepson and Jenkin 1991) naturally incorporate a size-disparity correlation. In this type of disparity encoding, the unit's preferred disparity Δx is roughly $\Delta\phi/2\pi f$, where $\Delta\phi$ is its preferred phase and *f* its preferred spatial frequency. If the largest phase disparity and bandwidth are the same for all spatial scales, then the largest preferred disparity is inversely proportional to frequency and thus proportional to size. Tsai & Victor (Tsai and Victor 2003) used stereo energy units with phase disparity which therefore incorporated a size-disparity correlation. They showed that this model, with template-matching, was able to account for stereoacuity as a function of frequency in sine-wave luminance gratings (NB these are luminance gratings at a constant depth, not random-dot patterns depicting sinusoidal depth modulation as have been used in this thesis). Our model uses position disparity, in which size-disparity correlation does not arise naturally, but has been built in by design. This leads to an important difference between the two implementations. Our size-disparity correlation links disparity to the size of the window across which disparities are sought, but not to spatial frequency. Our correlation-based model includes information from all spatial frequencies, independent of window size. Thus, the meaning of "size-disparity correlation" is somewhat different in the two cases.

3.4.5 *Limitations of the model*

Our model suffers from many limitations, most of which were forced on us by the difficulty of running simulations with large numbers of neurons. Most previous studies have either used stimuli with a uniform disparity profile, meaning that it suffices to model neurons at only one location in the visual field (Tsai and Victor 2003; Read and Cumming 2006; Read 2010), or have modelled neurons at several locations but with only one spatial frequency and orientation (Qian 1994). In order for the model to detect gratings that vary in depth, we needed to compute responses in many locations in the visual field. It would have been very costly also to model the responses of stereo energy units tuned to many different spatial frequencies and orientations. We therefore used the cross-correlation technique (Qian 1994; Harris, McKee and Smallman 1997; Banks, Gepshtein and Landy 2004; Filippini and Banks 2009) as a convenient short-cut to approximate the responses of many stereo energy units tuned to all possible frequencies and orientations.

Our analysis showing how local cross-correlation can be implemented exactly by stereo energy units is clearly idealized. Most notably, we integrated the response over all spatial frequencies, while keeping the receptive field size constant. Extending the integration to infinite spatial frequency is obviously unrealistic, although in practice will not greatly affect the results, since unrealistically high spatial frequencies will be removed from the images by the optical blurring and pre-processing. Keeping the receptive field size constant is a more serious limitation. Of course, primary visual cortex contains cells with a range of receptive field sizes. We have included only one window-size (receptive field size) at each preferred disparity. Once again, this was for reasons of computational economy. We regard the window-size within our model as representing the smallest receptive field sizes which contribute significantly to disparity detection. Ideally, we would have included a range of window-sizes at every disparity, with the smallest window-size at each disparity increasing as a function of disparity. However, since stereoresolution is limited by the smallest windows present, we would not expect this to alter our results substantially.

Keeping the receptive field size constant corresponds to postulating that bandwidth declines with spatial frequency, as it does in the macaque (Devalois, Albrecht and Thorell 1982). Assuming Gabor receptive fields, a Gaussian envelope with standard deviation 3 arcmin

implies a bandwidth of 0.5 octaves at 15cpd; at 5cpd the bandwidth ranges from 1.5 octaves (sine phase) to 2.0 octaves (cosine phase), while at 0.5cpd the bandwidth is 1.8 octaves for sine phase (cosine-phase cells are low-pass). These values are consistent with those reported in macaque (Devalois, Albrecht and Thorell 1982). At a given frequency, the bandwidth will be narrower for large RFs than for small ones.

As mentioned in the previous section, for each window-size, our correlation-based model includes information from all (luminance) spatial frequencies. In our model there is therefore no dependence between (luminance) spatial frequency and window-size. This is a consequence of the mathematical trick we have used to integrate over frequencies. In fact, several lines of evidence suggest that larger disparities are detected predominantly by mechanisms tuned to lower spatial frequencies in the luminance domain (Kulikowski 1978; Schor and Wood 1983; Smallman and MacLeod 1994). Thus, it would be more realistic to include a weight term in the integration over luminance spatial frequency, weighting the integral towards lower frequencies at the larger disparities/window-sizes, and towards higher frequencies at the smaller disparities/window-sizes.

We have not included any neuronal noise within our model, nor have we attempted to reproduce human stereoacuity for gratings: the smallest disparity amplitude detectable at each frequency. In principle, it would be simple to add this. Stereoacuity is limited by the spacing of disparity detectors, and by neuronal and stimulus-dependent noise (random correlations between non-corresponding parts of the dot pattern, for example).

We have only modeled the detection of horizontally-oriented disparity gratings. Humans find these easier to detect than vertically-oriented gratings (Bradshaw and Rogers 1999; Bradshaw, Hibbard, Parton, Rose and Langley 2006; Serrano-Pedraza and Read 2010; van der Willigen, Harmening, Vossen and Wagner 2010). It is currently unclear what model features would be required to match this feature of stereo vision. However, a clue may be that the disparity tuning surfaces of real cortical neurons are extended horizontally and are relatively narrow vertically (Cumming 2002). In any stereo algorithm, the choice of window-size represents a trade-off between resolution and accuracy. Large windows collect support over a wider region of the image, enabling greater accuracy and robustness against

false matches. However, they also lose the ability to track rapid changes in depth. For this reason, disparity steps are detected most accurately by windows which are elongated parallel to the edge and narrow orthogonal to the edge (Kanade and Okutomi 1994). Thus, the horizontally-elongated disparity tuning surfaces of real neurons would be expected to give greater sensitivity to changes in depth along a vertical direction in the image, as observed in humans. Further modelling work is required to examine whether models which incorporate this known anisotropy in V1 neurons can reproduce the anisotropy in human depth perception.

A great deal is now known about how disparity is encoded within V1. Much less is known about how this activity is read out in higher areas to result in depth perception and judgments on tasks such as our grating detection (Parker 2007). Thus, our model is necessarily much more speculative here. Is it realistic to assume that our brains have access to "templates" representing the expected V1 output for different stimuli? Physiologically, these templates could be represented as the synaptic weights between V1 and "grating detector" units in a higher visual area (see (Read 2010) for a more detailed account). While neurons specifically tuned for disparity gratings have not been reported, "grating detector" units would also respond preferentially to disparity curvature and slant, and such neurons are known to exist in areas IT and MT (Janssen, Vogels and Orban 1999; Nguyenkim and DeAngelis 2003). Alternatively, such neurons might be constructed as required. In areas such as LIP, neurons quickly adapt their responses to the particular task requirements at hand (Snyder, Batista and Andersen 2000). In this view, participants may be able to construct adequate templates simply from the few disparity gratings they are shown as demonstration stimuli.

3.5 Conclusions

Local cross-correlation within a fixed window has been postulated as a model of human stereo vision. This model accounts for stereoresolution when depth is modulated sinusoidally, but gives incorrect predictions for square-waves. We have shown that introducing a size/disparity correlation, such that larger disparities are detected within coarser windows, reconciles the local cross-correlation model with human stereoresolution

on both square- and sine-wave disparity gratings. This supports the original conclusion of Banks et al. (2004) that the limit on spatial stereoresolution is set by the smallest receptive field size of V1 neurons, which respond best to locally frontoparallel surfaces (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). There is thus no need to invoke further limits imposed by cells in extrastriate cortex tuned to more complicated aspects of disparity such as slant and curvature. Such cells can be created by combining the outputs of V1 neurons with different preferred disparities, but in this view, they inherit a fundamental limit on stereoresolution, set in primary visual cortex.

Chapter 4. Conjunctions between motion and disparity are encoded with the same spatial resolution as disparity alone

4.1 Introduction

In chapter 2 we measured human spatial resolution for disparity-defined depth and, consistent with previous results (Tyler 1974; Bradshaw and Rogers 1999; Banks, Gepshtein and Landy 2004; Filippini and Banks 2009), we found it to be much worse than for luminance information. This is believed to be because spatial resolution for disparity is limited by the overall sizes of receptive fields in primary visual cortex, whereas spatial resolution for luminance is limited by the size of their ON/OFF subregions (Banks, Gepshtein and Landy 2004; Nienborg, Bridge, Parker et al. 2004; Filippini and Banks 2009). Thus, information about the fine detail of disparity, potentially available within the photoreceptor activations, is lost at an information bottle-neck in V1. In the previous chapter we modeled disparity-selective cells in V1 using a local cross-correlation model and showed that this model can account for human performance on the detection of disparity gratings of different waveforms. This result lended further support to the idea of a disparity processing bottle-neck in V1. We wondered if other information bottle-necks at subsequent levels of cortical processing could be revealed by their effect on perception. To this end, we examined the spatial resolution with which humans can detect conjunctions between horizontal motion and disparity.

Disparity and motion are linked in natural viewing because objects closer to and farther than the plane of fixation appear to move in opposite directions when you move your head. Because this link is generated by observer self-motion, it applies across the entire visual field. Thus, conjunctions between motion and disparity arising from self-motion should not need to be encoded with very fine resolution. Cells which respond well to specific conjunctions of motion and disparity have been found in cortical area MT (Bradley, Qian and Andersen 1995; DeAngelis and Uka 2003), where receptive fields are around 10 times larger than those in V1 (Gattass and Gross 1981). If conjunctions between motion and

disparity are detected by specialized MT cells, we expect such conjunctions to be encoded with much lower resolution than disparity alone.

To examine this, we designed a task which requires the observer to detect conjunctions between motion and disparity. We introduced a "joint motion/disparity grating", a random-dot pattern in which the pairing between horizontal motion and disparity alternated as a function of vertical position. That is, in alternate horizontal strips, near dots moved left while far dots moved right, or near dots moved right while far dots moved left (Figure 24A). This is different from either a pure disparity grating built from moving dots, Figure 24B, or a pure motion grating built from two depth planes, Figure 24C, both of which we also used for comparison. In each case, we asked subjects to discriminate the "signal" grating from "noise", shown in Figure 24D. Figure 25 represents the stimuli in disparity/velocity space. To a system which detects only disparity, or to one which detects only motion, the joint motion/disparity grating is indistinguishable from noise. Thus, this task requires mechanisms which extract both motion and disparity and the correlations between them (Qian and Andersen 1997; Anzai, Ohzawa and Freeman 2001; Read and Cumming 2005c; Qian and Freeman 2009).

In the same subjects, we probed the spatial resolution for each of these three types of gratings, using correlation thresholds to equalize task difficulty, and obtain an unbiased estimate of spatial resolution. Using a signal-detection theory model, we extracted estimates of the receptive field size and internal noise with which the brain detects each type of grating.

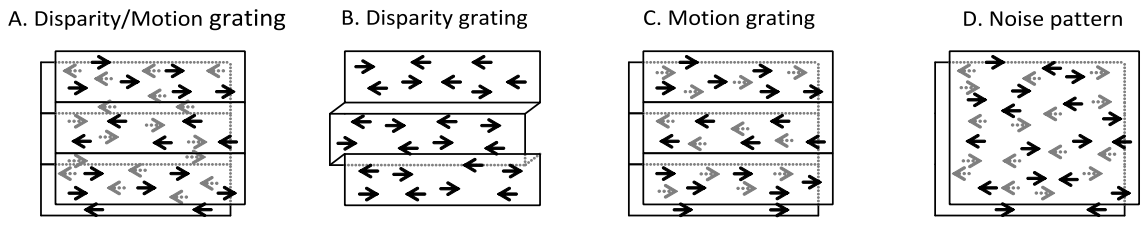


Figure 24: Sketches of the different types of stimuli used. Notice that in every case the same speeds and disparities were present. The "pure disparity" grating is built from moving dots; there are leftward and rightward dots everywhere in the stimulus, but the depth of the dots alternates as a function of vertical position. Similarly the "pure

motion" grating contains two transparent depth planes, but the direction of motion of dots in the two planes alternates.

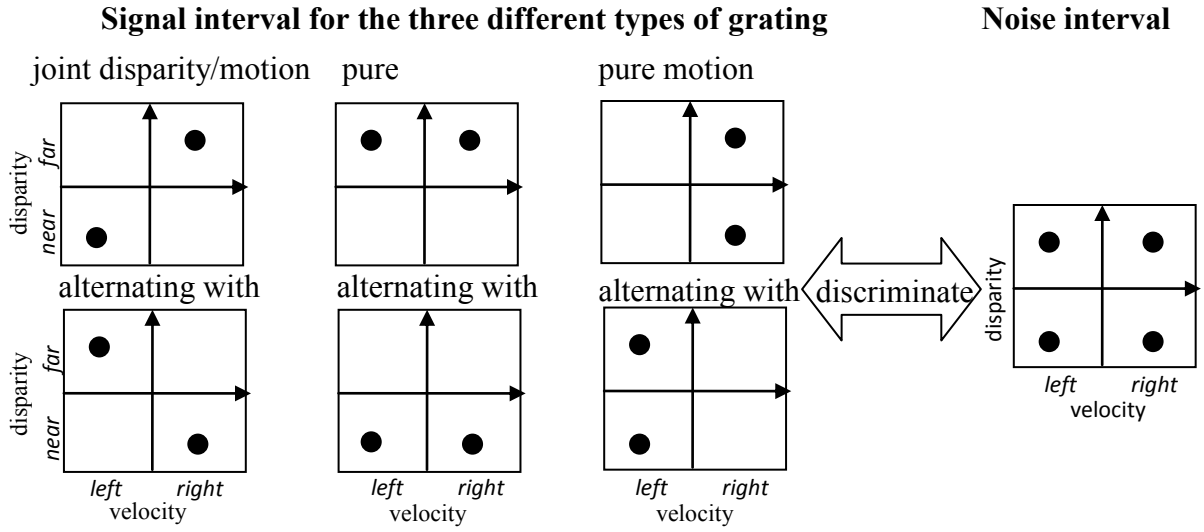


Figure 25. The task in our grating discrimination experiment, sketched in disparity/velocity space. In the gratings, the dot disparities and velocities alternate as a function of vertical position in the image. The noise contains the same velocities and disparities, but without the spatial structure.

4.2 Methods

4.2.1 *Equipment*

The experiments were performed in a dark room. Stimuli were projected on a projection screen (300 x 200 cm, Stewart Filmscreen 150, www.stewartfilm.com, supplied by Virtalis, Manchester), which the observers viewed from a distance of 160 cm. The subject's head was stabilized using a chin rest (UHCOTech HeadSpot). Two projectors, projecting through polarizing filters, were used to separate the two eye's images. The interocular cross-talk was less than 2%. White had a luminance of 4 cd/m^2 and black had a luminance of 0.07 cd/m^2. The projected image was 71 x 53 cm subtending 25° x 19°. The stimuli were presented in the central region of the image and had a size of 500 x 500 pixels (9° x 9°). The dot size was 2 x 2 pixels (2.1 x 2.1 arcmin).

4.2.2 *Stimuli*

Stimuli were presented using Matlab (The Mathworks, Natick, MA, USA; www.mathworks.com) with the Psychophysics Toolbox (Brainard 1997; Pelli 1997; Kleiner, Brainard and Pelli 2007). The stimuli used were random-dot stereograms with equal numbers of dots moving to the left and to the right with equal speed, depicting either a grating or a noise-pattern. Three different kinds of gratings were used. The first type of grating had two transparent depth planes and was made up of horizontal strips of equal width, where in each strip all dots moving to the left were in one depth plane and all dots moving to the right were in the other and where the direction of movement in the different depth planes was alternated between adjacent strips (see Figure 24A). The second type of grating was a horizontal square-wave in depth made up of equal numbers of dots moving in both directions (see Figure 24B). The third type of grating consisted of two transparent planes in depth with horizontal strips, where all dots in a single strip moved in the same direction and the direction of motion alternated between adjacent strips (see Figure 24C). The noise patterns consisted of two transparent depth planes with an equal number of dots moving in both directions in both planes (see Figure 24D). Any individual monocular frame of any stimulus was simply a structureless random-dot pattern with 150 dots per degree^2.

A problem with comparing resolution for different grating types is that one task may be harder than another. For example, detecting a joint motion/disparity grating requires information from two visual modalities to be combined, and thus arguably requires a more challenging judgment than, say, detecting a motion grating. This could lead to erroneous conclusions regarding resolution. For example, consider the toy example sketched in Figure 26. Figure 26A shows the internal signal for two hypothetical tasks, red and blue. These both have the same resolution, in that the signal is maximal for DC (0), and falls to zero at the same frequency. However, the red task is "harder", in that, at any frequency, its signal is lower than the blue signal by a constant factor. Now suppose there is some non-linearity converting this signal into perceptual judgments. In particular, there is a "floor" (when the signal falls below this level, perceptual performance on the relevant task is chance) and a "ceiling" (when the signal falls above this level, performance is perfect). Figure 26B shows the resulting performance. Performance falls at much lower frequencies for the red task,

despite the fact that the dependence of the underlying signal on frequency is the same in both cases.

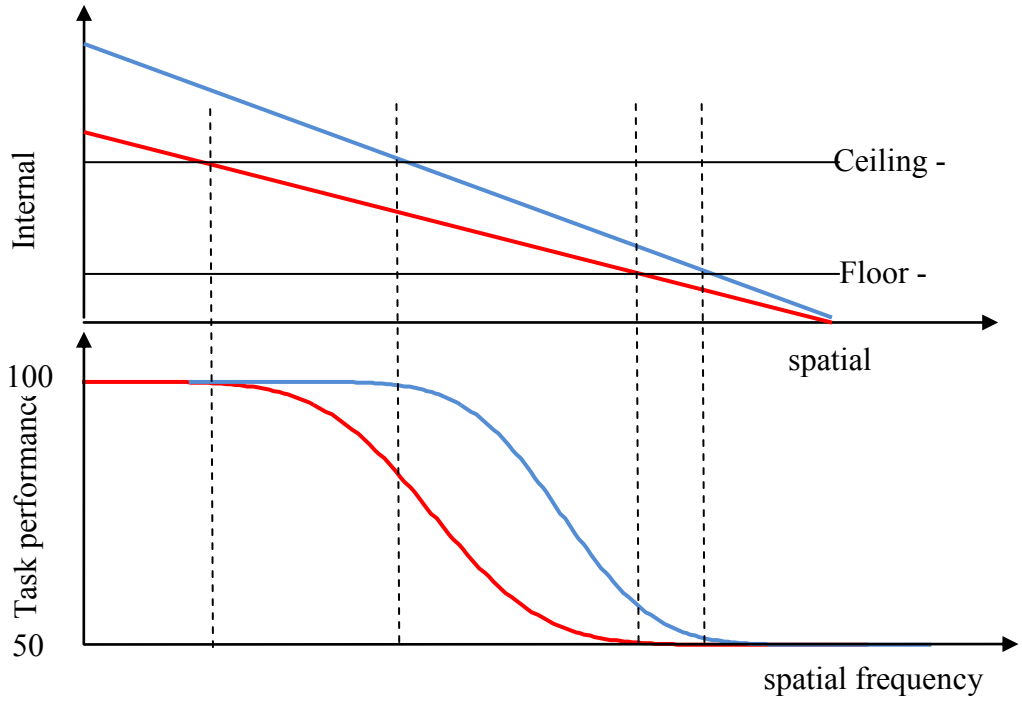


Figure 26: Cartoon of a possible relationship between internal signal and performance for two different tasks, represented in red and blue, which could lead to erroneous conclusions about spatial resolution. See text for details.

To avoid this problem, we used decorrelation to reduce the strength of the internal signal available for each task. This removed the ceiling effect, at least: if the internal signal was above ceiling, so that performance was perfect, we simply decreased correlation until the performance fell to 82%. In this way, we ensured that the difficulty of each task was equal. For motion, "decorrelation" means reducing motion coherence; for disparity, it means reducing interocular correlation. Thus for the pure motion gratings, we measured the motion coherence threshold at each frequency. The motion coherence was varied by, at each frame, giving each dot a probability p of being randomly repositioned rather than displaced in its direction of motion. The coherence level is defined as 1-p, such that for example a coherence level of 0.6 means that at any frame each dot had a 40% probability of being randomly repositioned.

For the pure disparity gratings, we measured the interocular correlation threshold at each

frequency. The interocular correlation was varied by, in the first frame of the stimulus, giving each dot a probability p of being positioned randomly in both eyes, instead of randomly in one eye and then offset horizontally by the desired disparity in the other eye. In subsequent frames, interocularly uncorrelated dots moved smoothly with the specified motion until they vanished off the edge of the stimulus. For the joint motion/disparity gratings, we measured both correlation and coherence thresholds.

4.2.3 *Observers*

10 observers participated in the experiments: one of the authors and nine inexperienced observers. Observer CB was unable to perform the interocular correlation threshold parts of experiment two.

4.2.4 *Tasks*

To obtain the speed and disparity amplitude for which the subjects could best detect the joint motion/disparity gratings at high frequencies (experiment 1) we used a one interval task as well as a two interval task. Amplitude is defined as half the peak-to-trough range of the waveform, (max-min)/2. For the one interval task, in each trial either a grating or a noise pattern was presented and the task was to report, by a button press, whether a grating had been presented or not. The subjects were allowed to view the stimuli for as long as they desired before making a decision. For the two interval task, one interval contained a grating and the other a noise pattern, and the task was to report, by a button press, which interval contained the grating. The interval length was 750 ms with a 200 ms blank between intervals. Subject PFA was tested with the one interval task and all other subjects with the two interval task. Once the optimal speed and disparity amplitude had been determined for a subject, that speed and disparity amplitude was used in all further testing of that subject. To obtain coherence and interocular correlation thresholds once the optimal speed and amplitude had been determined we used adaptive QUEST staircases (Watson and Pelli 1983) converging to 82% correct with a two-interval forced choice task where one interval contained a grating and the other interval contained a noise pattern and the task was to report, by a button press, which interval contained the grating. The interval length was either 500 or 750 ms with a 200 ms blank between intervals. The 500 ms interval length

was used for subject PFA, who is an author and an experienced psychophysical observer, and the 750 ms interval length was used for all other subjects. Each staircase was repeated three times in the same session.

4.3 Results

4.3.1 *Experiment 1: Obtaining optimal stimulus parameters for each subject*

In this chapter, we wanted to detect the finest resolution with which motion and disparity information is represented. Obtaining 4 correlation/coherence thresholds at many different spatial frequencies was a long and demanding experiment, and it was not feasible to also examine dependence on speed and disparity amplitude at each frequency. We therefore began by measuring each subject's performance as a function of speed and disparity only for a single, high frequency. In this way we aimed to identify a pair of values where the subject is able to perform well.

Figure 27 shows performance on the joint motion/disparity grating detection task as a function of disparity amplitude and speed for all subjects, for perfectly correlated stimuli. In each case there is a region of high performance surrounded by a region where performance was lower. The amplitudes and speeds used in experiment 2 were chosen for each subject individually in order to be approximately in the center of the region of high performance for that subject (white crosses in Figure 27). Table 1 shows the values used for each subject in the subsequent experiments.

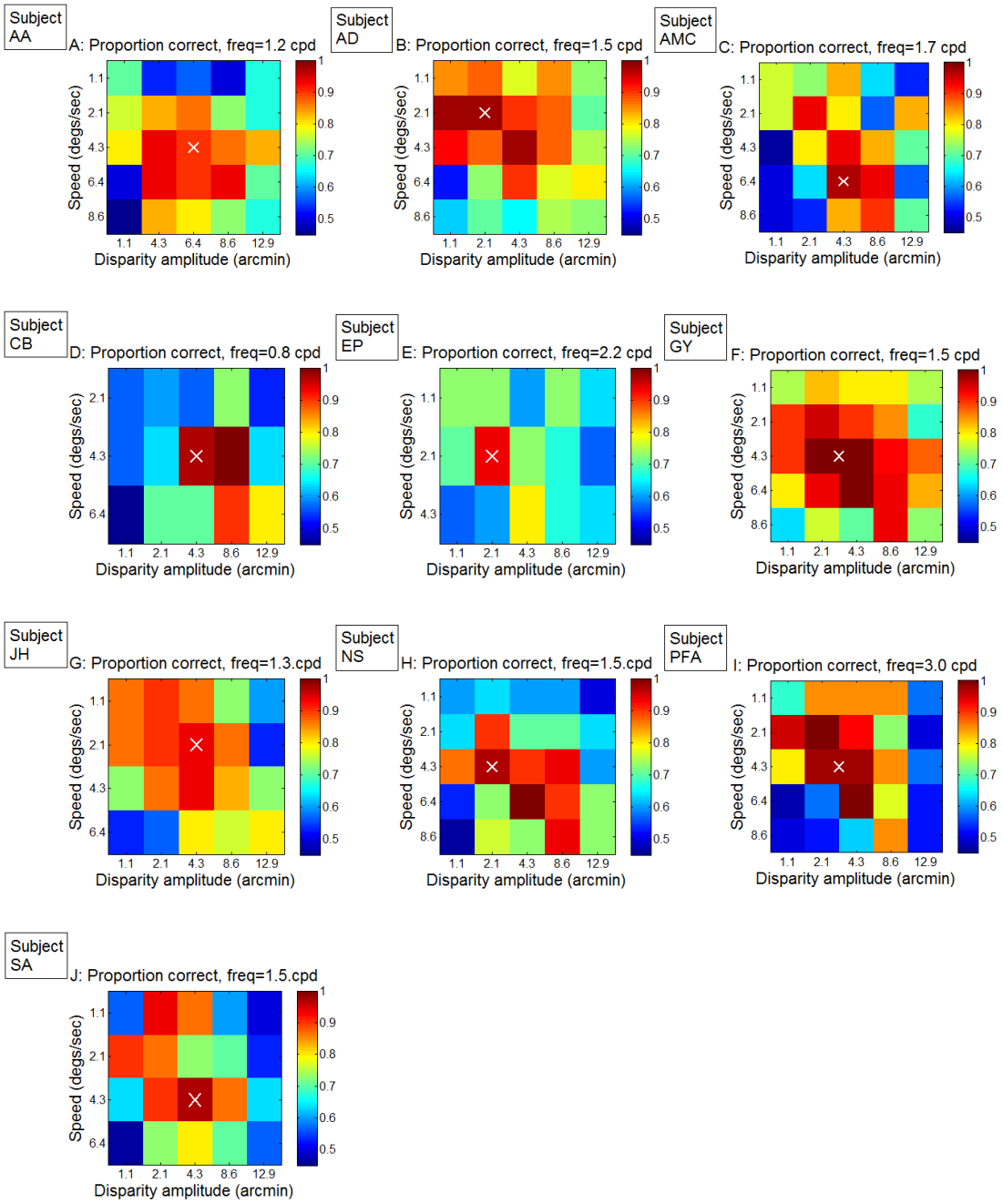


Figure 27: Performance on the 100%-correlated joint motion/disparity grating as a function of speed and disparity amplitude for all subjects. The white crosses show the values used in the subsequent experiments (see also Table 1). The proportion correct shown is based on 30 trials per combination of speed and amplitude for all subjects except subjects AD, GY and PFA who performed 40 trials per combination.

	AA	AD	AMC	CB	EP	GY	JH	NS	PFA	SA
Speed (degrees/s)	4.3	2.1	6.4	4.3	2.1	4.3	2.1	4.3	4.3	4.3
Disparity amplitude (arcmin)	6.4	2.1	4.3	6.4	2.1	3.3	4.3	2.1	3.3	4.3

Table 1: Speed of dot motion and disparity amplitude used in experiment 2 chosen based on the results of experiment 1

4.3.2 *Experiment 2*

We now proceeded to measure coherence and correlation thresholds for the three different types of gratings. Figure 28 shows the motion coherence thresholds measured at different frequencies for both the motion/disparity gratings and the pure motion gratings. The error-bars show ±1 standard error based on the three repetitions of each staircase. At low frequencies, subjects are able to perform the tasks at relatively low coherence; as the frequency increases, subjects require progressively more coherence in order to be able to reach threshold. All subjects can detect motion gratings even at very low coherences, down to 20% at the lowest frequencies. For some subjects there is little difference between the thresholds for the two types of gratings at low frequencies; PFA, for example, is equally good at detecting both sorts of grating. However, for some subjects, such as AMC in Figure 28, the coherence thresholds are far higher for the joint motion/disparity grating, even at the very lowest frequencies. This indicates that for this subject, detecting joint motion/disparity gratings is a genuinely harder task than detecting motion gratings, irrespective of their respective spatial resolutions. Thus, without the use of a coherence threshold, one could seriously misestimate the relative resolution in this subject (see Figure 26).

At higher frequencies the thresholds become increasingly different for all subjects. The pure motion gratings can be detected up to frequencies where the joint motion/disparity gratings are invisible, even at 100% motion coherence.

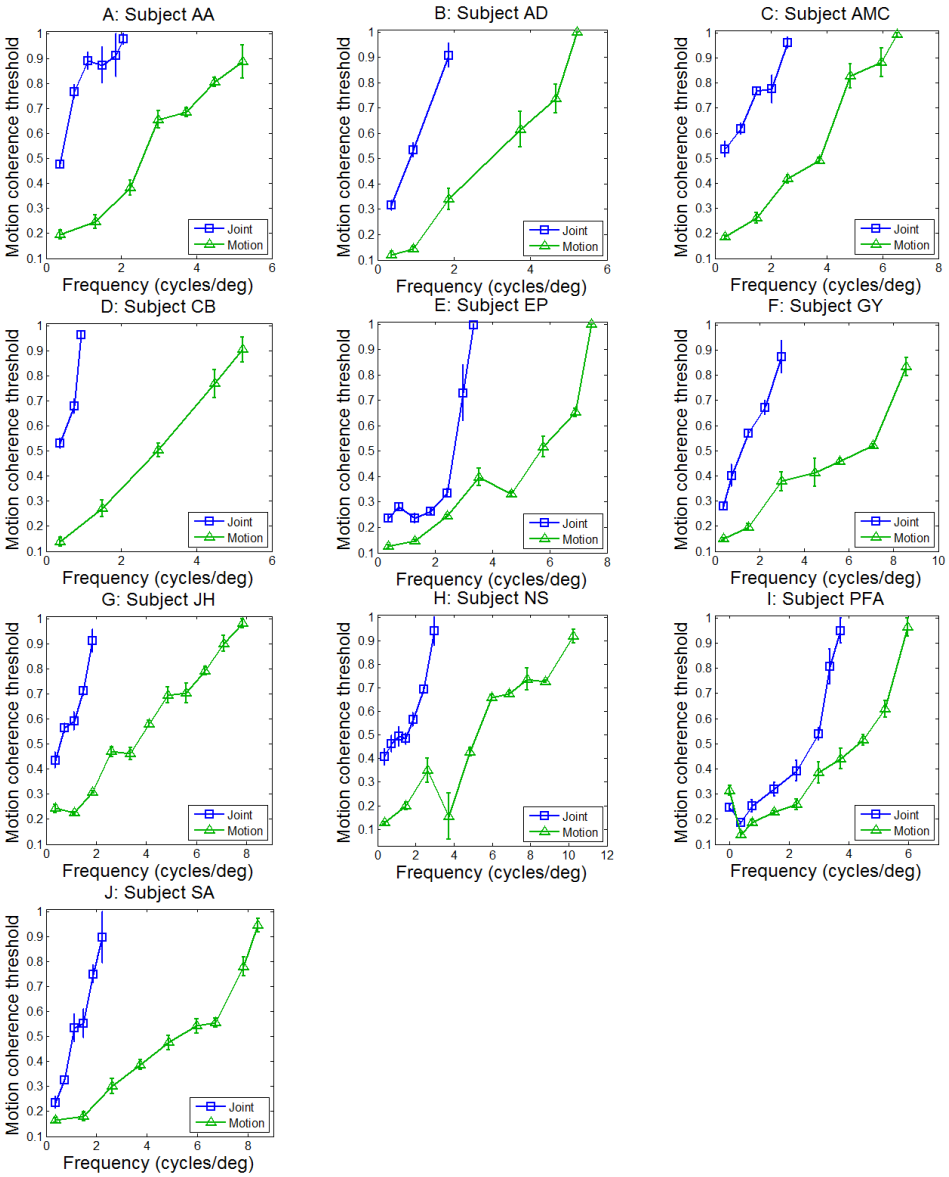


Figure 28: Motion coherence threshold as a function of frequency for the motion/disparity and pure motion gratings for all subjects. The speed and disparity amplitudes used for the gratings were set individually for each subject; values in Table 1.

Figure 29 shows the interocular correlation thresholds measured at different frequencies for both the motion/disparity gratings and the pure disparity gratings. Here, there is much less difference between the thresholds for the two different types of the gratings at low frequencies. For some subjects, this remains true at high frequencies, while for others, such as JH, there is a large difference at the highest frequencies.

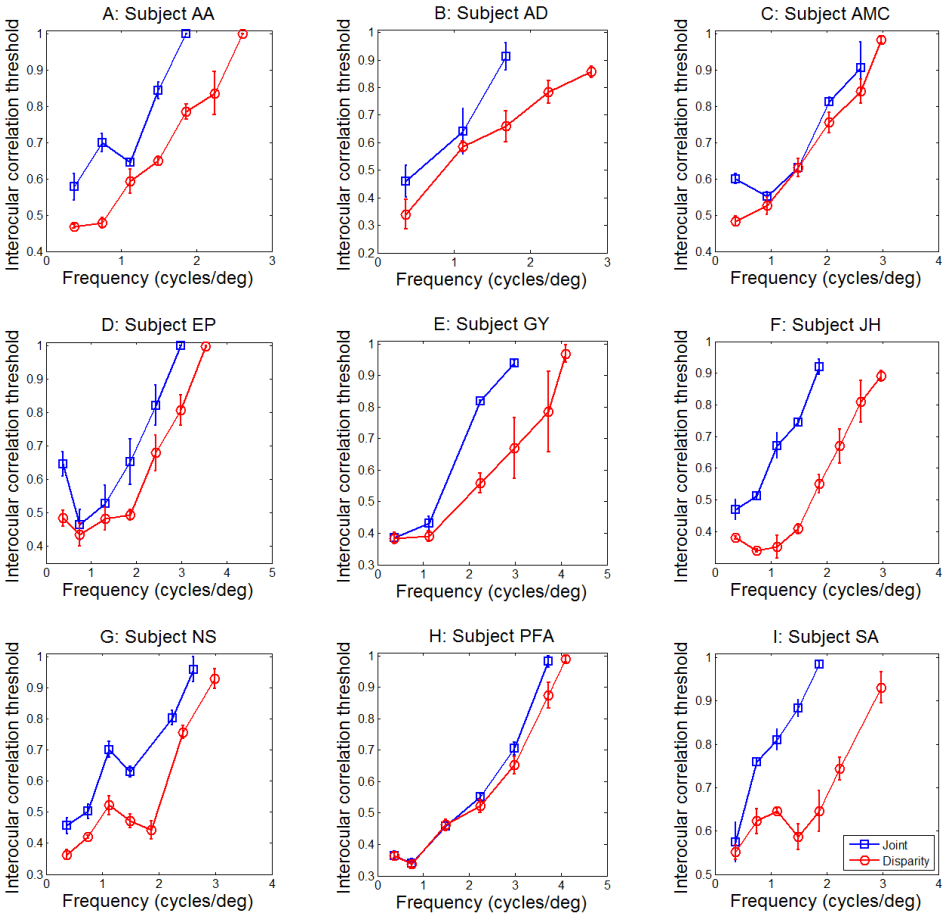


Figure 29: Interocular correlation threshold as a function of frequency for the motion/disparity and pure disparity gratings for all subjects.

In the figures above, we have presented two different types of threshold for the joint motion/disparity gratings: interocular correlation and motion coherence thresholds. Figure 30 compares these two threshold measurements. For some subjects, the thresholds are comparable in the two cases, but where there is a systematic difference such that the thresholds all differ in the same direction at least up to some frequency close to the highest one tested (as for subject PFA in Figure 30) it is the interocular correlation thresholds that are higher. This suggests that despite their conceptual similarity, the two manipulations are not equivalent perceptually, with reduction in interocular correlation having a more disruptive effect than reduction in motion coherence.

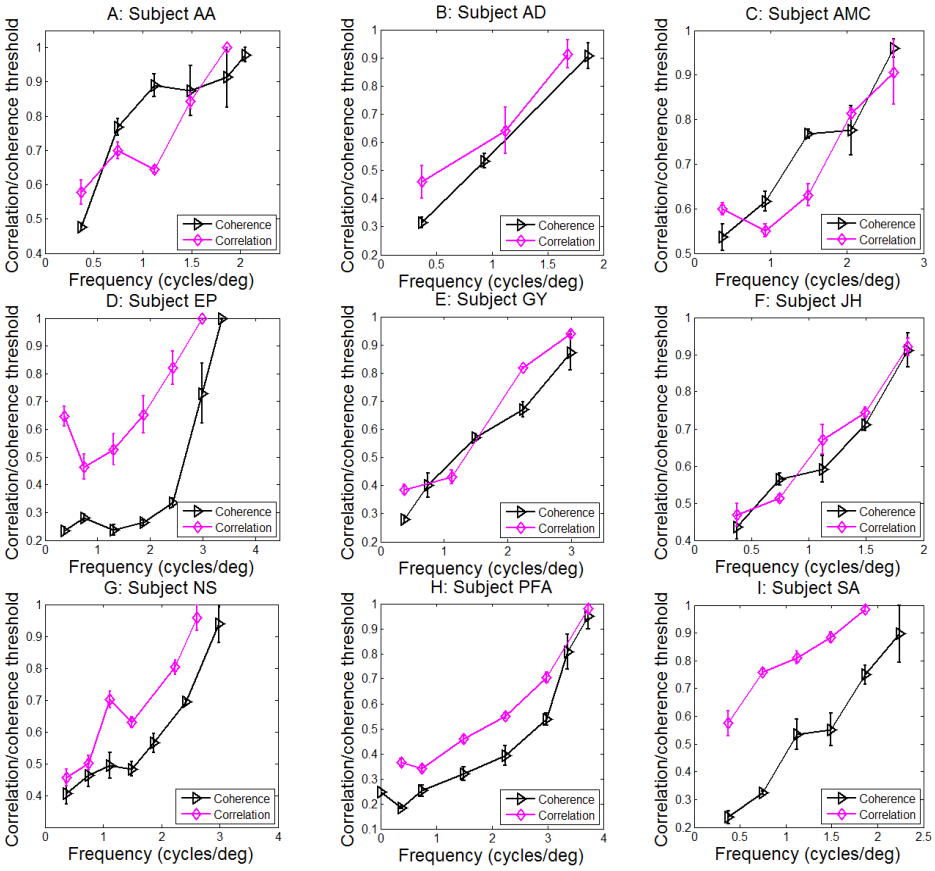


Figure 30: Interocular correlation thresholds and motion coherence thresholds for the motion/disparity gratings

4.3.3 *Data analysis*

In order to turn these measurements of coherence and correlation thresholds into a quantitative estimate of receptive field size, we used a model based on signal detection theory. We assumed that, for 100% correlated stimuli, the internal signal was proportional to the RMS of the unit-amplitude grating waveform after convolution by a Gaussian with standard deviation σ. Recent work has suggested this is a good model for the detection of disparity gratings (Serrano-Pedraza and Read 2010). This signal could be computed by a population of energy-model-like disparity-selective cells with Gaussian receptive field envelopes of diameter 2σ. Figure 31 shows how the RMS of the convolution between the Gaussian and the square-wave varies as a function of the ratio between the SD of the Gaussian and the wavelength λ=1/f of the square-wave. We write this function RMS(σ/λ). This function depends only on σ normalized by λ and not on σ and λ independently of each other, which is why this curve can be fit to threshold data at any frequency.

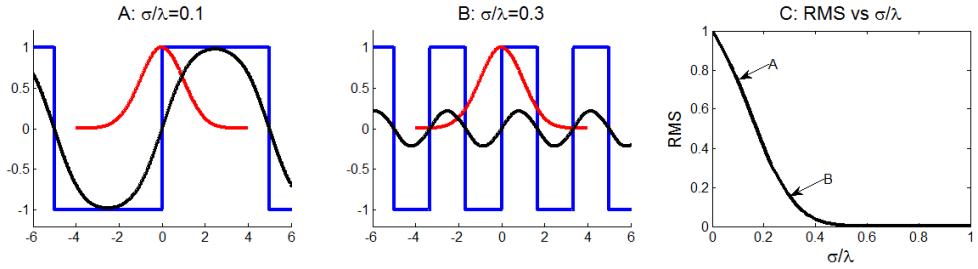


Figure 31: The resulting curves (black) when a square-wave (blue) is convolved with a Gaussian (red) for a lower frequency square-wave (A) and a higher frequency square-wave (B) of the same amplitude, and a plot of the RMS of these resulting curves for a range of ratios between the SD of the Gaussian and the wavelength of the square-wave (C).

Reducing interocular correlation or motion coherence must reduce this internal signal. As we have seen, interocular correlation thresholds were in general higher than motion coherence thresholds for the joint motion/disparity task at low frequencies. This suggests that, at least for some observers, a decrease in interocular correlation increases task difficulty more than the same decrease in motion coherence. At 100% the interocular correlation and motion coherence versions of the stimulus were exactly the same and at 0% of either coherence or correlation there was no signal, so there could only be a difference in difficulty at intermediate values of correlation/coherence. We therefore chose to model the different effects of changing interocular correlation and motion coherence by assuming that the signal depended on the correlation/coherence level raised to some power, κ, allowing different values of κ for the correlation and coherence. We refer to κ as the "decorrelation parameter", since it describes how seriously the available signal is degraded by decoherence/decorrelation. κ=1 means that the signal degrades linearly with decoherence/decorrelation; κ>1 gives faster degradation. With these assumptions, the internal signal available for performing the task is

$$sig = C^{\kappa} * RMS(\sigma/\lambda)$$

We then used signal detection theory to predict performance on the task. Since a two interval task was used, the signal detection theory prediction is that:

$$PC = 0.5 + 0.5 * \mathrm{erf}\left(\frac{sig}{\sqrt{2}N}\right)$$

where PC is the proportion of correct answers, erf is the error function, sig is the signal and N is the internal noise. At the 82% threshold, this yields:

$$0.82 = 0.5 + 0.5 * \text{erf}\left(\frac{C_{thresh}(f)^{\kappa} * RMS(\sigma/\lambda)}{\sqrt{2}N}\right)$$

from which we obtain:

$$C_{thresh}^{-1}(f) = \left(\frac{\text{RMS}(\sigma/\lambda)}{\sqrt{2}\text{N} * \text{erf}^{-1}(0.64)}\right)^{\kappa^{-1}}$$

Equation 7

However, we can notice an odd property of this equation. The right hand side can clearly drop below one at large frequencies but the left hand side is an inverse correlation/coherence threshold which cannot be smaller than one. Therefore one might think that in order to get the best possible fits to the data, the right hand side of the equation should flat-line when it reaches one. However, it is not the case that the psychophysical threshold reaches one and then stays at one at higher frequencies. Clearly the performance of a subject will keep getting worse as the frequency is increased further, such that the performance no longer reaches the threshold level of 82% correct even at the highest possible level of correlation/coherence. This further decrease in performance would not be captured by the flat-lining model which assumes that the level of performance remains constant after the threshold reaches one. The model described by Equation 7 on the other hand does capture this further decrease in performance even if it does so in a way that it is hard to make sense of. When the quality of the fits are evaluated this model but not the flat-lining model will introduce an extra penalty when it has gone past the point where the threshold reaches one and into the region where performance is below threshold level even at the highest possible level of correlation/coherence at a frequency where the threshold in the human data is lower than one (i.e. where it is possible to obtain a threshold). This seems quite reasonable and may arguably be a reason to prefer Equation 7 over the flat-lining version. Therefore Equation 7 will be used in this chapter. However, since there is still something odd about letting the curves that are supposed to model inverse correlation/coherence thresholds drop below one, the analysis has also been performed with the flat-lining model and these results are presented in Appendix 2.

A scaled version of the RMS curve from Figure 31 can be fitted to the coherence and correlation thresholds from Figure 28 and Figure 29 by finding appropriate values of σ, N

and κ, giving us estimates of the receptive field diameter (2σ) and internal noise levels (N) relevant to each task. Within a subject, the RF and noise parameters for the two different sets of data on the joint motion/disparity task, i.e. the interocular correlation thresholds and the motion coherence thresholds, were assumed to be the same. Similarly, within each subject, the decorrelation parameter was kept the same for both motion coherence data sets, and for both interocular correlation data sets. Therefore there were, for each subject, eight parameters in total: RF diameters and noise parameters for the motion, disparity and joint motion/disparity data and decorrelation-parameters for the interocular correlation and motion coherence thresholds. We fitted these parameters to the experimentally measured values of C_{thresh}, by minimizing the sum of squared errors over all four fits plus an additional term $\max(\kappa, \kappa^{-1})$ for each of the two decorrelation parameters. The additional term was included to keep either decorrelation parameter from growing too small/large. We used resampling to obtain error bars on the parameters by repeating the fitting 10,000 times, each time simulating a new repetition of each staircase by running a new staircase with a simulated observer with the experimentally measured threshold.

We can also use Equation 7 to estimate f_{max}, the highest grating frequency at which the task could be performed. At this frequency, performance is only threshold even when the stimulus is perfectly coherent/correlated, i.e. $C_{thresh}(f_{max})=1$. Thus f_{max} is given by the solution of

$$\mathrm{RMS}(f_{max}\,\sigma) = \sqrt{2}N * \mathrm{erf}^{-1}(0.64)$$

Figure 32 and Figure 33 show the inverted coherence and correlation thresholds along with the fits. The fits are generally good, validating the assumptions used in producing our model. The percentage of variance explained was at least 70% and at average 85% for the motion fits, at least 78% and at average 90% for the disparity fits and at least 40% and at average 84% for the joint fits. Note that each parameter affects more than one curve, so fits are not necessarily optimal for any individual curve.

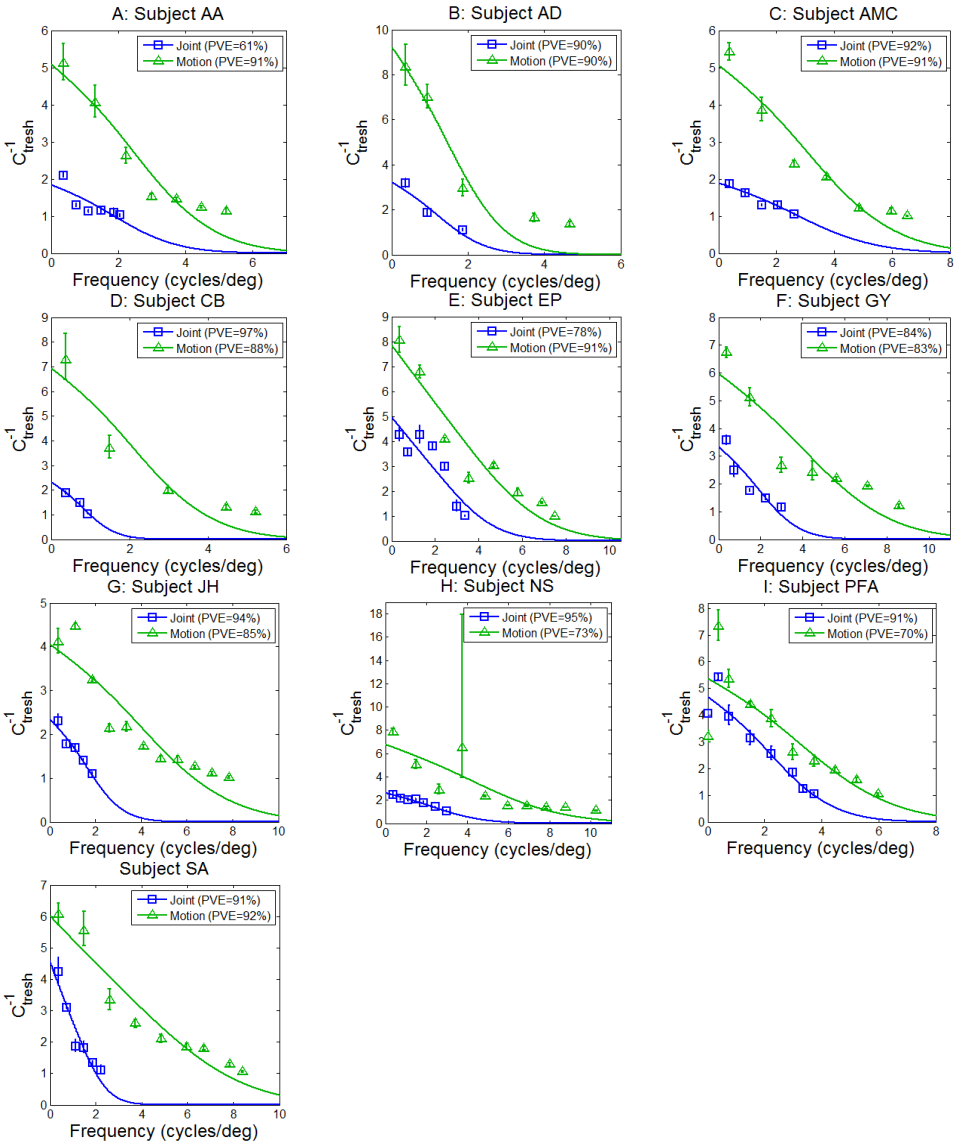


Figure 32: Inverted motion coherence thresholds as a function of frequency for the pure motion gratings (green) and joint motion/disparity gratings (blue) and model fits (see text) for all subjects.

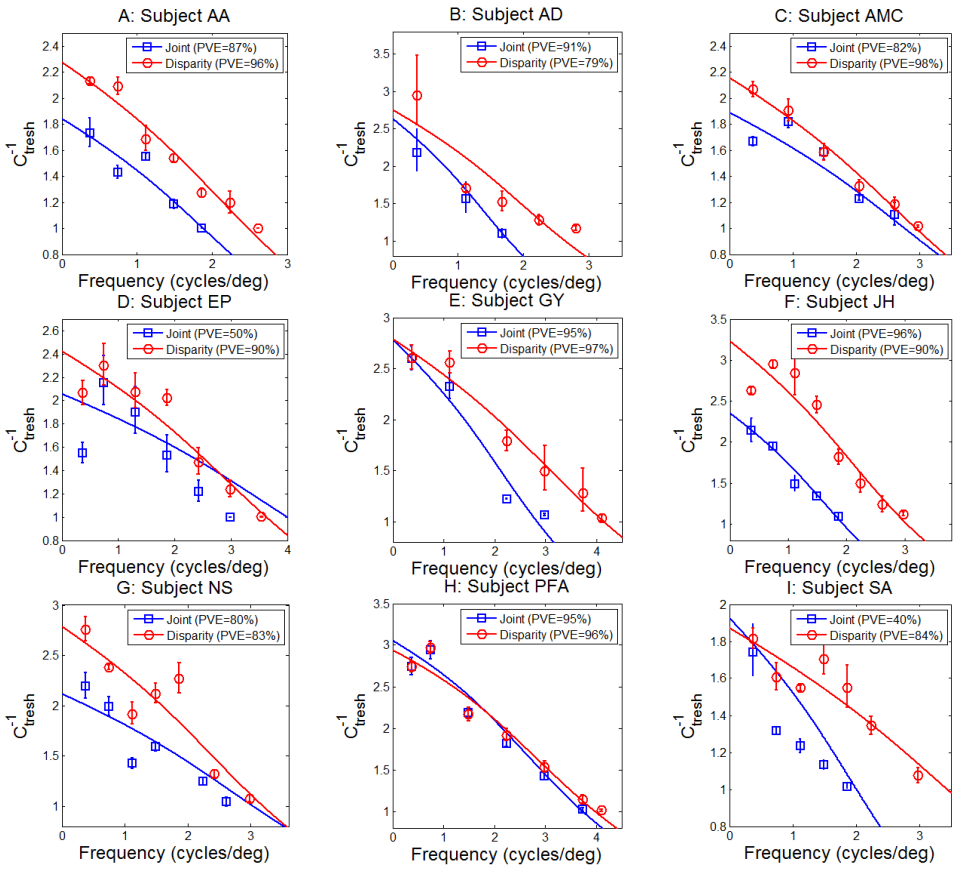


Figure 33: Inverted interocular correlation thresholds as a function of frequency for the pure disparity gratings (red) and joint motion/disparity gratings (blue) and model fits (see text) for all subjects.

Table 2 and Figure 34 show the parameters that gave the best fits for each subject. The receptive field sizes limiting detection are estimated at around 6 arcmin for the pure motion task and 8 arcmin for the pure disparity, similar though slightly larger than the 6 arcmin previously estimated by Banks et al (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009). Figure 35 shows the RF diameters and noise parameters from Figure 34AB after normalizing them to be 1 for the pure motion data. We see immediately that the RF diameter and neuronal noise estimated for the pure motion task are both smaller than for either the pure disparity or the joint motion/disparity task. This statement holds for all subjects individually, apart from subject AD where the motion fit is poor (see Figure 32B). At a population level, the RF diameter for pure motion is significantly smaller than for pure disparity ($p<0.05$, paired t-test, n=9, comparing σ_{motion} to $\sigma_{disparity,}$ i.e. triangles vs circles in Figure 34A) and for joint motion/disparity ($p<0.01$, paired t-test, n=10, comparing σ_{motion} to $\sigma_{joint,}$ i.e. triangles vs squares in Figure 34A). Similarly, the noise affecting pure motion

judgements is significantly smaller than for pure disparity ($p<0.01$, paired t-test, n=9, comparing N_{motion} to $N_{disparity,}$ i.e. triangles vs circles in Figure 34B) or for joint motion/disparity ($p<0.01$, paired t-test, n=10, comparing N_{motion} to $N_{joint,}$ i.e. triangles vs squares in Figure 34B). These two effects, smaller receptive fields and lower noise, combine to make motion gratings detectable up to higher frequencies than gratings defined by disparity. All subjects including AD can detect motion gratings up to higher frequencies than either pure disparity or joint motion/disparity gratings ($f_{max}^{motion} > f_{max}^{disparity}$, $f_{max}^{motion} > f_{max}^{joint}$). Thus, our results show clearly that motion is encoded with higher resolution than disparity information, and also that it is affected by less neuronal noise.

In contrast, there is no such clear difference between spatial resolution for pure disparity as compared to resolution for conjunctions between motion and disparity. Pure disparity gratings remain detectable up to slightly higher frequencies than joint motion/disparity gratings (3.3 cpd vs 2.5 cpd), but this does not seem to reflect a difference in receptive field size. The relative RF diameters estimated for the joint and for the pure disparity gratings show no consistent difference across our population. At the population level, the mean RF diameter is larger for the joint motion/disparity task than for the pure disparity task, but this difference is not significant either for the raw RFs (Figure 34, $p=0.07$, paired t-test, n=9) or after normalising by the motion RFs (Figure 35, $p=0.15$, paired t-test, n=9). In contrast, the estimated noise level is larger for the joint than for the pure disparity wherever there is a significant difference (5/9 subjects), and this difference is significant on the population level both for the raw noise parameters (Figure 34, $p<0.05$, paired t-test, n=9) and after normalising by the motion noise parameters (Figure 35, $p<0.05$, paired t-test, n=9). Thus, our analysis suggests that pure disparity and joint motion/disparity gratings are encoded with the same spatial resolution. The pure disparity encoding is, however, subject to lower effective noise, meaning that pure disparity gratings can be detected up to somewhat higher frequencies than joint motion/disparity gratings despite the similar RF sizes.

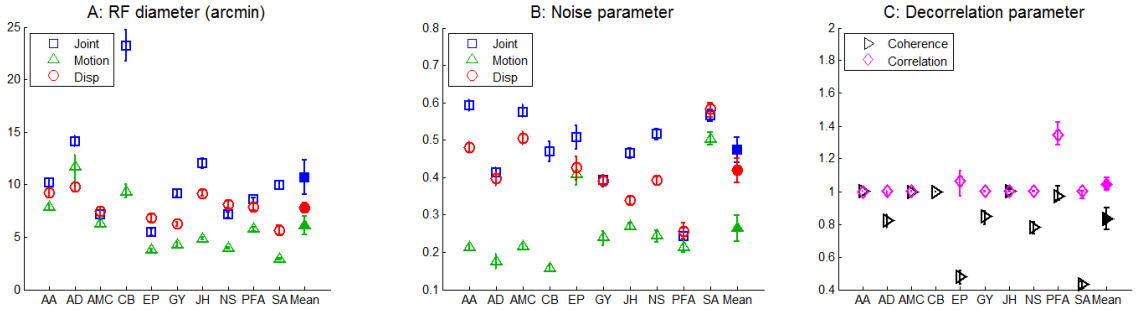


Figure 34: The parameters that gave the best fits to the data. The filled symbols on the right show the averages across subjects. Error-bars on individual subjects' results show the 95% confidence intervals obtained by resampling, as described in the text; error-bars on the population averages show ±1 standard error of the results from individual subjects.

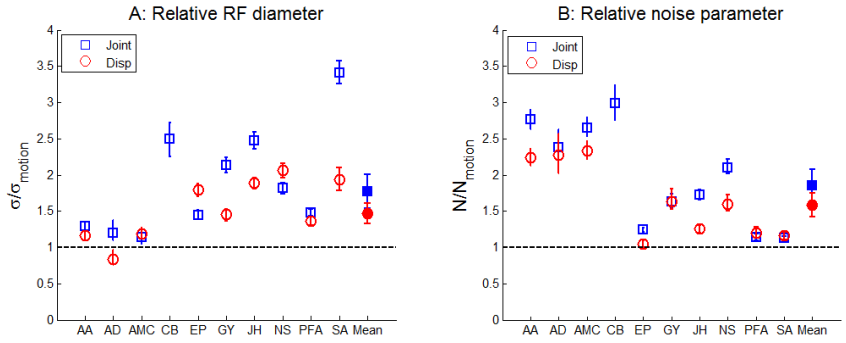


Figure 35: The data from Figure 34AB, normalized to be one for the pure motion data. The filled symbols on the right show the averages across subjects. Error-bars on individual subjects' results show the 95% confidence intervals on these ratios, obtained by resampling as described in the text; error-bars on the population averages show ±1 standard error of the ratios from individual subjects.

	Subject	AA	AD	AMC	CB	EP	GY	JH	NS	PFA	SA	Mean	SD
Fitted parameters	Motion RF diameter (arcmin)	7.9	11.8	6.3	9.3	3.8	4.3	4.9	3.9	5.8	3.0	6.1	2.80
	Disparity RF diameter (arcmin)	9.2	9.9	7.5		6.8	6.3	9.2	8.1	7.9	5.7	7.8	1.42
	Joint RF diameter (arcmin)	10.1	14.2	7.1	23.2	5.5	9.1	12.0	7.1	8.5	10.0	10.7	5.07
	N_{motion}	0.22	0.17	0.22	0.16	0.42	0.24	0.27	0.25	0.21	0.51	0.26	0.11
	$N_{disparity}$	0.49	0.40	0.52		0.43	0.40	0.34	0.40	0.26	0.60	0.42	0.10
	N_{joint}	0.60	0.42	0.59	0.48	0.52	0.40	0.47	0.53	0.24	0.58	0.47	0.11
	κ_{decoh}	1.0	0.83	1.0	1.0	0.48	0.85	1.0	0.78	0.97	0.43	0.83	0.22
	κ_{decorr}	1.0	1.0	1.0		1.1	1.0	1.0	1.0	1.3	1.0	1.04	0.10
Derived quantities	$\sigma_{disparity}/\sigma_{motion}$	1.16	0.83	1.19		1.79	1.46	1.88	2.06	1.36	1.94	1.46 *	
	$\sigma_{joint}/\sigma_{motion}$	1.29	1.2	1.14	2.5	1.45	2.13	2.48	1.82	1.48	3.41	1.77*	
	$N_{disparity}/N_{motion}$	2.23	2.28	2.34		1.04	1.63	1.26	1.60	1.20	1.16	1.59*	
	N_{joint}/N_{motion}	2.76	2.38	2.66	2.99	1.24	1.63	1.72	2.10	1.14	1.13	1.86*	
	f_{max}^{motion} (cycles per degree)	4.22	3.03	5.28	3.94	6.68	7.47	6.32	8.10	5.78	7.55	5.84	1.70
	$f_{max}^{disparity}$ (cycles per degree)	2.49	2.65	2.95		3.64	4.14	3.05	3.20	3.97	3.45	3.28	0.57
	f_{max}^{joint} (cycles per degree)	1.88	1.78	2.77	1.00	4.00	2.83	1.94	3.04	3.73	2.01	2.50	0.94

Table 2. Fit parameters and derived quantities for all subjects. Asterisks * indicate ratios significantly greater than 1 (t-test on the log-ratios, p<0.01). "Mean" is the arithmetic mean except for the 4 rows showing ratios, where it is the geometric mean.

4.4 Discussion

In this chapter, we have examined spatial resolution for disparity judgments with moving dots, motion direction judgments with disparate dots, and a novel disparity/motion conjunction task. The joint motion/disparity grating used in this task cannot be detected by pure disparity sensors or by pure motion sensors alone. If viewed with one eye, removing disparity information, the signal interval containing the grating appears identical to the

noise interval: at all locations in the image, there are dots streaming both leftward and rightward with no spatial structure. If a single frame is viewed in isolation, removing motion information, again both intervals are identical, since they both depict two transparent planes of dots at both near and far disparities. To detect the joint grating requires the observer to extract not only the local motion and disparity in the stimulus, but also the conjunctions between them.

The existing literature suggests that cortical area MT would be ideally suited for this task. MT contains many neurons which are sensitive both to motion and disparity. MT neurons are typically suppressed by motion in opposite directions within the same depth plane (Snowden, Treue, Erickson and Andersen 1991; Qian and Andersen 1994), as in our noise stimulus. However, they respond well to transparent motion in opposite directions in two different depth planes, as in our grating (Bradley, Qian and Andersen 1995). Indeed, the transparent motion/disparity random-dot patterns from which we built our joint motion/disparity gratings were originally introduced to study MT neurons (Bradley, Qian and Andersen 1995; Bradley, Chang and Andersen 1998; Dodd, Krug, Cumming and Parker 2001). Thus, MT neurons should respond more strongly to the signal interval containing the joint motion/disparity than to the noise interval.

If observers perform the task using this difference in the activity of MT, we can make a strong prediction about the resulting spatial resolution. The physiological literature suggests that MT neurons respond best when the conjunction between motion and disparity (e.g. left-near/far-right) is the same all over the receptive field. There is evidence that motion integration in pattern-selective MT cells occurs at a scale that is smaller than the entire receptive field, such that the cells are only pattern-selective if the components that make up the moving plaid overlap and not if they are presented in different parts of the receptive field (Majaj, Carandini and Movshon 2007). However, there is no evidence that MT neurons have subunits tuned to opposite directions and disparities, as would be required to detect motion boundaries. Therefore, if MT is involved in performing our joint motion/disparity task, we would expect the spatial resolution to be low, reflecting the large size of MT receptive fields which are typically around 4° at small eccentricities, (Raiguel, Van Hulle, Xiao, Marcar and Orban 1995). Specifically, it should be much poorer than for

pure disparity gratings, where spatial resolution reflects the much smaller receptive fields found in V1 (Banks, Gepshtein and Landy 2004; Nienborg, Bridge, Parker et al. 2004; Filippini and Banks 2009; Allenmark and Read 2010; Allenmark and Read 2011).

Our results comprehensively disprove this prediction. Our results for pure motion and disparity gratings are similar to previous results (Anderson and Burr 1987; Bradshaw and Rogers 1999; Georgeson and Scott-Samuel 2000; Banks, Gepshtein and Landy 2004; Allenmark and Read 2010), although these workers used disparity gratings built from static dots and motion gratings without disparity. We find that subjects are able to detect pure motion and disparity gratings up to frequencies an order of magnitude lower than for luminance. We find that motion gratings can be detected up to significantly higher frequencies than disparity (mean f_{max} = 5.8 cpd for motion and only 3.3 cpd for disparity). Our analysis suggests that this is partly because receptive fields for motion are smaller than those for disparity (6 arcmin vs 8 arcmin), and partly because motion judgments are subject to less internal noise (effective noise higher for disparity than for motion by a factor of 1.6). However, contrary to the prediction, joint motion/disparity gratings could be detected up to frequencies only slightly lower than disparity itself, at mean f_{max} 2.5 cpd. Our analysis suggests that conjunctions between motion and disparity are detected with the same spatial resolution as disparity itself, with the limit set by sensors around 8 arcmin in diameter. The slightly lower frequency limit for joint motion/disparity gratings reflects slightly higher effective noise. Thus, spatial resolution for motion/disparity conjunctions is limited by spatial resolution for each component in isolation. The effective resolution is therefore that of disparity, the lower-resolution component. Importantly, resolution is not limited further by whatever mechanism detects the conjunction. The physiological arguments laid out above therefore strongly imply that this mechanism is not located in MT.

Indeed, the fine resolution reported for motion gratings already implies that area MT may not be limiting perception here. Physiological studies of area MT in the macaque have failed to find cells selective for the position or orientation of motion boundaries (Marcar, Xiao, Raiguel, Maes and Orban 1995). Similarly, human brain imaging studies have not found any evidence that area MT is involved in the perception of motion boundaries (Orban, Dupont, De Bruyn, Vogels, Vandenberghe and Mortelmans 1995; Reppas, Niyogi,

Dale, Sereno and Tootell 1997), instead identifying a different area, with no clear counterpart in the monkey visual system, as being involved in the processing of motion defined contours (Orban, Dupont, De Bruyn et al. 1995). Since the pure motion gratings could be detected by looking for motion boundaries, and the joint motion-disparity task could involve looking for motion boundaries in a specific depth plane, the apparent lack of involvement of area MT in the processing of motion boundaries suggests that a different area may have been used to perform the task and the area found in the fMRI experiments of Orban et. al. seems like a reasonable candidate since this area was found to be involved specifically in the perception of motion boundaries.

Recently, there has been much debate over whether the ability to detect conjunctions between motion and disparity requires V1 neurons which are specifically tuned to both motion and disparity (Qian and Andersen 1997; Anzai, Ohzawa and Freeman 2001; Qian and Freeman 2009), or whether V1 neurons which are tuned solely to motion or solely to disparity can also contribute, if correlations between their activity are read out subsequently (Read and Cumming 2005a; Read and Cumming 2005b; Neri and Levi 2008). Around 14% of disparity-selective cells in macaque V1 are also selective for direction of motion (Read and Cumming 2005b) and these cells could support performance on the present task. If these cells were solely responsible, it is perhaps slightly surprising that the level of internal noise deduced for the joint task was only 1.12 higher than for the pure disparity task, given the physiological data implying around 7 times as many pure disparity cells as jointly-tuned cells in early visual cortex. Perhaps performance was supported also by cells selective to motion or disparity alone. Such cells would, individually, be blind to the difference between the joint grating and the noise stimulus, but the presence of the grating could be revealed by correlations in their activity (Read and Cumming 2005c). The emerging consensus seems to be that both mechanisms contribute (Neri and Levi 2008), and our results are consistent with that.

Our estimates of receptive field size suggest that the resolution for motion, disparity and for conjunctions between the two are all limited by V1 receptive field sizes. In other words, there is no subsequent information bottle-neck affecting joint motion and disparity; information available in V1 is accurately passed on to perception. This is perhaps

surprising. If conjunctions between motion and disparity were used primarily to deduce self-motion from motion parallax, for example, a very coarse encoding would suffice, since in that case all objects nearer than fixation move in one direction, while all objects further than fixation move the other. Similarly in many visual scenes it would suffice to have an accurate spatial map of motion, or of disparity, alone. The joint motion/disparity detection required for our task is not required to perceive scenes with rapid local variations in both motion and depth, say a crowded street with many people at different distances moving in different directions. Such a scene could be accurately represented by extracting motion alone and disparity alone, and then overlaying the representations of the two quantities. Our results imply the additional ability to represent different motions and disparities at the same point in space. This more subtle ability benefits scenes with transparency, e.g. a flock of birds in flight, or the branches of a tree moving in the wind, or a shoal of fish under the reflective surface of the water. Our remarkable ability to resolve fine conjunctions between motion and disparity information may reflect the importance of such scenes during our evolution.

Chapter 5. Conclusions and future directions

5.1 Conclusions

The psychophysical results on detection of sine-wave vs. square-wave disparity gratings in chapter 2 presented a challenge to local cross-correlation models of disparity detection and therefore also questioned the conclusions that have been drawn based on modeling work using such models. In particular it called into question the conclusion reached by Banks et al. (Banks, Gepshtein and Landy 2004; Filippini and Banks 2009) that spatial stereoresolution is set in area V1.

In chapter 3 it was found that this challenge could be met by a local cross-correlation model that incorporated the known size-disparity correlation. This provides further support to the theory that spatial stereoresolution is set in area V1. In addition to explaining the new results presented in chapter 2 this modified correlation model also explains old human results on the frequency dependence of the upper depth limit (Tyler 1973) for both square-waves and sine-wave disparity gratings while the old model without a size-disparity correlation could only explain the results for sinusoidal gratings.. This supports Tyler's suggestion that the disparity gradient limit is a consequence of the size-disparity correlation.

The small difference between the resolution for joint motion/disparity perception and pure disparity perception found in chapter 4 is inconsistent with what would be predicted if the resolution for the joint motion/disparity perception was limited by the large receptive field sizes in area MT. This suggests that information on joint motion/disparity available in area V1 is read out in an area other than MT with little or no loss in resolution. This information could be readout from V1 cells tuned to both motion and disparity, it could be based on detection of correlations in the activity of cells tuned only to motion and cells tuned only to disparity or it could be a combination of both. As mentioned in the discussion section of Chapter 4 recent evidence (Neri and Levi 2008) suggests that it is a combination of both. However, the most important conclusion that can be drawn from the results presented in

Chapter 4 is that the readout cannot be done by cells in higher areas that integrate over many V1 cells tuned to different positions but with the same motion direction and/or disparity tuning. Since the readout must involve cells in higher areas somehow, and these tend to have larger receptive fields, it seems likely that cells that have subunits with different disparity and/or motion direction tuning must be involved. Such cells could get their receptive field properties by combining the output from many V1 cells with different disparity and/or motion tuning at different positions. This question of how to read out the joint motion/disparity information from area V1 without loss of resolution is part of a more general problem of how conjunctions between any two visual properties can be detected by neurons at a higher level of the visual system without a loss of resolution due to the larger receptive field size of such a higher level neuron. It seems likely that this is done in a similar way in many different cases and a solution to this problem in the joint disparity/motion perception case would therefore potentially be of quite general interest.

5.2 Future directions

The psychophysical data presented in chapter 2 (Figure 9) suggests that the upper depth limit may in general be slightly higher for sine-waves than for square-waves. However, this difference in upper depth limit is not entirely consistent across subjects and frequencies and the data is in principle also consistent with there being no real difference at all in upper depth limit between the two waveforms. Previous work by Tyler (1973) where the upper depth limit as a function of spatial frequency was measured for sinusoidal and square-wave disparity gratings using line stereograms supported a difference in upper depth limit between the two waveforms that was independent of frequency and larger than what is suggested by the data presented in chapter 2. Based on this, a natural way of extending the work presented in chapter 2 would be to perform experiments with the same type of stimuli and task used in the experiments described in chapter 2 but designed specifically to measure the upper depth limit and to include lower frequencies instead of just looking at what happens close to the upper frequency limit where subjects stop being able to do the task at any disparity amplitude. This would provide the answer to the question raised in section 3.3.6 of whether the difference in upper depth limit between the two waveforms found by Tyler may be partially or completely an effect of using line stereograms and may

be reduced or disappear with random dot stereograms. The answer to that question would be of particular interest since the results of the simulations with the modified model presented in section 3.3.6, which were obtained using random dot stereograms, had the same frequency dependence as Tyler's human results but did not show the lower upper depth limit for square-wave gratings compared to sinusoidal gratings.

The modeling work presented in chapter 3 used a local cross-correlation model, while the current model that best captures the known physiology of disparity-selective V1 cells is the stereo energy model. In section 3.3.1 it was shown that the local cross-correlation model used can be thought of as the combined output from energy model units tuned to different (luminance) spatial frequencies and orientations. However, as was discussed in section 3.4.5, the local cross-correlation model is clearly an idealization, most importantly because it assumes integration over an infinite range of spatial frequencies. It would therefore be interesting to confirm that similar results can be obtained using energy model units. Repeating all the simulations presented in chapter 3 with energy model units tuned to a range of different spatial frequencies and orientations would be extremely time consuming, but it would be possible to repeat a small part of the simulations using energy model units and confirm that the results come out the same, and this together with the theoretical proof presented in section 3.3.1 would be sufficient to strongly support the assumption that a simulation with local cross-correlation model used provides a good approximation to a more detailed simulation based on energy model units.

Another approximation made in the modified local cross-correlation model with the size disparity correlation was that only one window-size was included for each preferred disparity. As discussed in section 3.4.5, including only the smallest available RF-size in the modeling should be an acceptable approximation, since spatial stereoresolution is limited by the smallest RF-size and it was necessary to make this approximation because simulations with the full range of window-sizes would have been too time consuming. However, again it would be possible to repeat a small part of the simulations with a more realistic model, having a range of different window-sizes for each preferred disparity, and check that the results come out the same.

In section 3.3.1 it was mentioned that the proof of the equivalence between the combined response of many energy model units with different frequency and orientation and local cross-correlation requires on the assumption that window size is independent of (luminance) spatial frequency. However, this equivalence may still hold approximately with a certain dependence between window size and frequency. This is likely to be hard to prove or disprove analytically, but it could be testing with simulations. This could be done by simulating the responses of energy model units, tuned to the same position and with the same receptive field size, but tuned to different frequencies and orientations, to a few simple patterns of disparity. Then the responses of these energy model units could be combined while giving different weight to units tuned to different frequencies, thereby introducing a dependence between RF size and spatial frequency while avoiding the obvious difficulties involved in combining response of units with different RF size. The combined responses could then be compared to the responses of a local cross-correlator with the same RF/window size to the same stimuli. This could be repeated for a few different RF/window sizes using different patterns of weights in the combination across frequencies for each RF/window size in order to simulate a realistic dependence between RF size and spatial frequency tuning.

All the psychophysical experiments and simulations presented in chapters 2 and 3 were performed with horizontal gratings. Previous psychophysical experiments have shown that vertical sine-wave disparity gratings are harder to detect than horizontal ones (Bradshaw and Rogers 1999; Bradshaw, Hibbard, Parton et al. 2006) and that this difference in detectability is much smaller for square-wave gratings (Serrano-Pedraza and Read 2010). As suggested in section 3.4.5 it may be possible to model this orientation dependence of the detectability of disparity gratings by giving the model neurons windows with greater horizontally than vertical elongation such as what has been found in the receptive fields of real V1 neurons (Cumming 2002). This would favor detectability of horizontal gratings, since for these spatial stereoresolution depends on the smaller vertical elongation of the window, while it depends on the larger horizontal elongation for vertical gratings. However, it is not clear why such a model would predict the much larger difference for the sinusoidal compared to the square-wave gratings. It is therefore possible that a full explanation of the stereo anisotropy would require better models of processing in higher

visual areas. In particular Serrano-Pedraza and Read (2010) suggest that the explanation may be “multiple spatial frequency channels for detecting horizontally oriented modulations in horizontal disparity, but only one for vertically oriented modulations”.

In section 3.4.1 it was discussed how the correlation model that was used in Chapter 3 seems highly likely to perform in a way that is inconsistent with the frequency analysis point of view that is discussed in section 1.2.1.1. This intuition could of course be confirmed by simulating a discrimination experiment, where the model has to discriminate between the two types of noise pattern. If the model is indeed able to perform this discrimination it would be interesting to perform such a discrimination experiment with human subjects. If human subjects can perform the discrimination as well this would lend further support to the model and serve as strong evidence against the frequency analysis point of view (in the disparity domain). If the discrimination cannot be performed by human subjects, this would be consistent with the frequency analysis point of view and would present a major problem for the model.

The modelling performed in chapter 4 used a relatively abstract model. This was sufficient for the purpose of estimating the difference in receptive field size for the different kind of stimuli. However, there are more detailed computational models of area MT such as the one by Qian et al (Qian, Andersen and Adelson 1994). A natural addition to the work presented in chapter 4 would therefore be to confirm that this more detailed model predicts a much lower resolution than what was found in our experiments, as suggested by our more abstract modelling. Finally using fMRI experiments comparing neural activity in different areas with the joint motion/disparity gratings and the noise stimulus used in chapter 4 may be a good way of making further progress in answering the question of which brain areas are involved in joint motion/disparity perception. The brain area, mentioned in section 4.4, which in previous fMRI experiments have been found to be involved in processing of kinetic boundaries (Orban, Dupont, De Bruyn et al. 1995) may be of particular interest in such an fMRI experiment.

There is psychophysical evidence showing that what appears as frontoparallel can be affected by a slanted reference plane, such that surfaces parallel to the reference plane

appear frontoparallel (Mitchison and Westheimer 1984). Various processes in stereo vision have been found to be influence by the presence of such a reference plane (Glennerster and McKee 1999). This suggests that it may not be the actual slant in the sine-wave disparity gratings that make them harder to detect as the amplitude is increased but perhaps rather the local slant relative to the average slant of the entire grating. If this is the case then it should not make much of a difference whether the entire grating is slanted or frontoparallel. This idea is not inconsistent with the model that has been presented in this thesis. However, the psychophysics and modelling presented in this thesis has only dealt with the case where the average slant is zero. In order to incorporate the idea that the "meaning of frontoparallel" can change a more general model may need to adapt to the average slant of the stimulus and seek correlations in a plane with such a slant.

Appendix 1. Computing the binocular term of an energy model unit

The binocular term in the response of a single energy-model complex cell is:

$$B = 2(S_{L1}S_{R1} + S_{L2}S_{R2})$$

$$= 2\int dxdy \int dx'dy' I_L(x,y) I_R(x',y') \exp\left(-\frac{\left((x-x_L)^2 + (y-y_L)^2\right) + \left((x'-x_R)^2 + (y'-y_R)^2\right)}{2\sigma^2} \right)$$

$$\left[\cos(k_x x + k_y y + \phi_L)\cos(k_x x' + k_y y' + \phi_R) + \sin(k_x x + k_y y + \phi_L)\sin(k_x x' + k_y y' + \phi_R)\right]$$

This cell is tuned to a spatial frequency and orientation specified by the wavenumbers k_x and k_y, and has receptive fields centered at (x_L,y_L) and (x_R,y_R), with phases ϕ_L and ϕ_R respectively. We now compute the total response of many such cells tuned to many spatial frequencies and orientations, but all with the same receptive field centers and phases:

$$B_{\text{int}} = \int B dk_x dk_y$$

$$= 2\int dxdy \int dx'dy' I_L(x,y) I_R(x',y') \exp\left(-\frac{\left((x-x_L)^2 + (y-y_L)^2\right) + \left((x'-x_R)^2 + (y'-y_R)^2\right)}{2\sigma_{RF}^{\ 2}} \right)$$

$$\int dk_x dk_y \left[\cos(k_x x + k_y y + \phi_L)\cos(k_x x' + k_y y' + \phi_R) + \sin(k_x x + k_y y + \phi_L)\sin(k_x x' + k_y y' + \phi_R)\right]$$

Doing the innermost integral first, we obtain:

$$\int dk_x dk_y \left\{\cos(k_x x + k_y y + \phi_L)\cos(k_x x' + k_y y' + \phi_R) + \sin(k_x x + k_y y + \phi_L)\sin(k_x x' + k_y y' + \phi_R)\right\}$$

$$= \frac{1}{4}\int dk_x dk_y \begin{Bmatrix} \left[\exp i(k_x x + k_y y + \phi_L) + \exp -i(k_x x + k_y y + \phi_L)\right]\left[\exp i(k_x x' + k_y y' + \phi_R) + \exp -i(k_x x' + k_y y' + \phi_R)\right] \\ -\left[\exp i(k_x x + k_y y + \phi_L) - \exp -i(k_x x + k_y y + \phi_L)\right]\left[\exp i(k_x x' + k_y y' + \phi_R) - \exp -i(k_x x' + k_y y' + \phi_R)\right] \end{Bmatrix}$$

$$= \frac{1}{4}\int dk_x dk_y \begin{Bmatrix} \exp i(k_x x' + k_y y' + \phi_R)\exp i(k_x x + k_y y + \phi_L) + \exp i(k_x x' + k_y y' + \phi_R)\exp -i(k_x x + k_y y + \phi_L) \\ + \exp -i(k_x x' + k_y y' + \phi_R)\exp i(k_x x + k_y y + \phi_L) + \exp -i(k_x x' + k_y y' + \phi_R)\exp -i(k_x x + k_y y + \phi_L) \\ - \exp i(k_x x' + k_y y' + \phi_R)\exp i(k_x x + k_y y + \phi_L) + \exp i(k_x x' + k_y y' + \phi_R)\exp -i(k_x x + k_y y + \phi_L) \\ + \exp -i(k_x x' + k_y y' + \phi_R)\exp i(k_x x + k_y y + \phi_L) - \exp -i(k_x x' + k_y y' + \phi_R)\exp -i(k_x x + k_y y + \phi_L) \end{Bmatrix}$$

$$= \frac{1}{2}\int dk_x dk_y \left\{\exp i(k_x x' + k_y y' + \phi_R)\exp -i(k_x x + k_y y + \phi_L) + \exp -i(k_x x' + k_y y' + \phi_R)\exp i(k_x x + k_y y + \phi_L)\right\}$$

$$= \frac{1}{2}\exp i(\phi_R - \phi_L)\int dk_x dk_y \left\{\exp i(k_x x' + k_y y')\exp -i(k_x x + k_y y)\right\}$$

$$+ \frac{1}{2}\exp i(\phi_L - \phi_R)\int dk_x dk_y \left\{\exp -i(k_x x' + k_y y')\exp i(k_x x + k_y y)\right\}$$

$$= \frac{1}{2}\exp i(\phi_R - \phi_L)\int dk_x dk_y \exp i\left(k_x(x'-x) + k_y(y'-y)\right) + \frac{1}{2}\exp i(\phi_L - \phi_R)\int dk_x dk_y \exp i\left(k_x(x-x') + k_y(y-y')\right)$$

$$= \frac{1}{2}\left(e^{i\phi_R - i\phi_L} + e^{-i\phi_R + i\phi_L}\right)\delta(x-x')\delta(y-y') = \cos(\Delta\phi)\delta(x-x')\delta(y-y')$$

where $\Delta\phi = \phi_R - \phi_L$ is the phase disparity of the cells. Using this result in the equation for the integral of B gives us:

$$B_{\text{int}} = \int B dk_x dk_y =$$

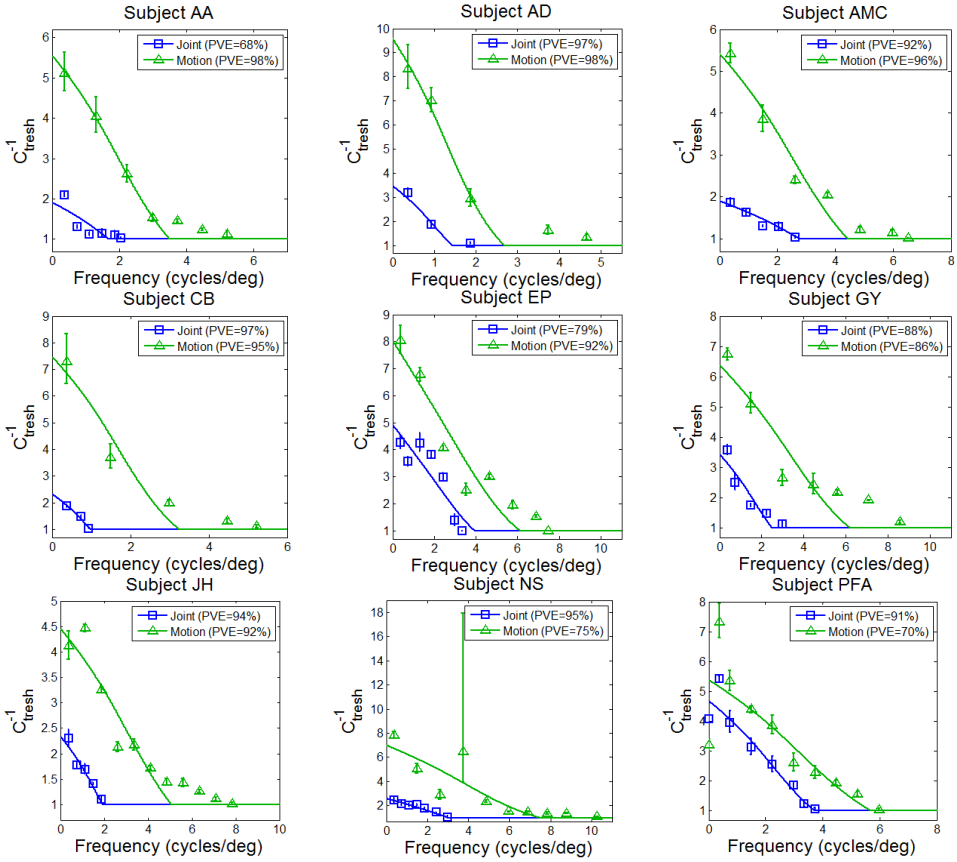
$$2\cos(\Delta\phi)\int dxdy \exp\left(-\frac{\left((x-x_L)^2+(y-y_L)^2\right)}{2\sigma^2}\right) I_L(x,y) \exp\left(-\frac{\left((x-x_R)^2+(y-y_R)^2\right)}{2\sigma^2}\right) I_R(x,y)$$

Appendix 2. Flat-lining model of coherence/correlation thresholds

In this appendix the results of an analysis of the threshold data presented in chapter 4 is presented that is identical to what is described in section 4.3.3 except that the equation for C_{thresh}^{-1} has been changed so that the value can never drop below 1:

$$C_{thresh}^{-1}(f) = max\left(\left(\frac{\mathrm{RMS}(\sigma/\lambda)}{\sqrt{2}\mathrm{N} * \mathrm{erf}^{-1}(0.64)}\right)^{\kappa^{-1}}, 1\right)$$

The results are qualitatively very similar to the results presented in Chapter 4. The estimated RF diameter is the smallest for motion and the largest for joint motion/disparity and the difference between the RF diameters for the joint motion/disparity and pure disparity is much smaller than what would be predicted if joint motion/disparity was processed by cells in area MT. The main conclusions of Chapter therefore still hold if this alternative model is used. The main difference to the results obtained in Chapter 4 is that the estimated RF sizes are slightly larger and therefore differ a bit more from previously obtained estimates for pure motion and pure disparity.



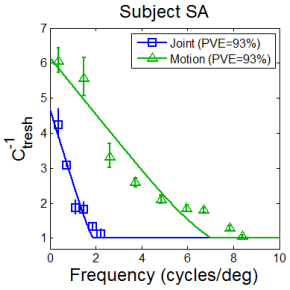


Figure 36: Inverted motion coherence thresholds as a function of frequency for the pure motion gratings (green) and joint motion/disparity gratings (blue) and model fits for all subjects.

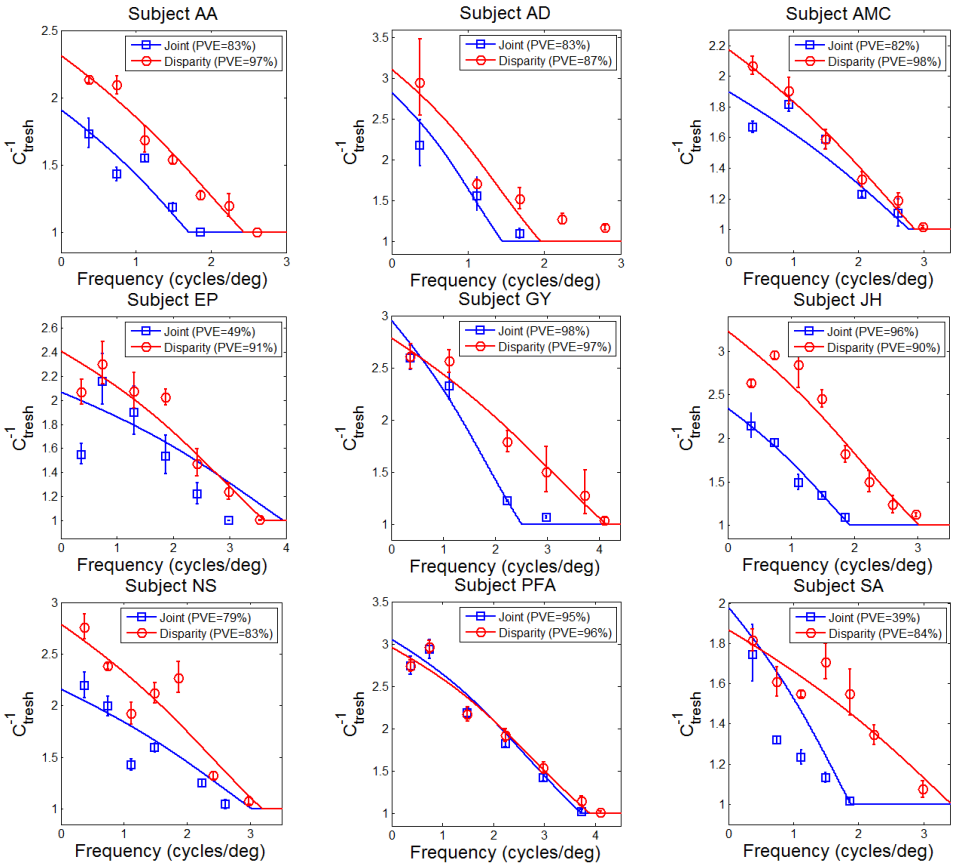


Figure 37: Inverted interocular correlation thresholds as a function of frequency for the pure disparity gratings (red) and joint motion/disparity gratings (blue) and model fits for all subjects.

	Subject	AA	AD	AMC	CB	EP	GY	JH	NS	PFA	SA	Mean	SD
Fitted parameters	Motion RF diameter (arcmin)	9.8	19.5	7.6	11.5	3.8	5.3	6.3	4.4	5.9	3.4	7.8	4.90
	Disparity RF diameter (arcmin)	9.5	15.2	7.7		6.8	6.3	9.2	8.1	8.1	6.0	8.5	2.77
	Joint RF diameter (arcmin)	11.7	14.5	7.1	23.3	5.5	10.6	12.0	7.3	8.6	11.5	11.2	5.05
	N_{motion}	0.20	0.12	0.21	0.15	0.42	0.21	0.25	0.23	0.21	0.47	0.25	0.11
	$N_{disparity}$	0.48	0.30	0.51		0.43	0.40	0.34	0.40	0.26	0.57	0.41	0.10
	N_{joint}	0.58	0.33	0.59	0.48	0.52	0.38	0.47	0.51	0.25	0.53	0.46	0.11
	κ_{decoh}	1.0	0.97	1.0	1.0	0.56	0.87	1.0	0.81	0.98	0.48	0.87	0.19
	κ_{decorr}	1.0	1.2	1.0		1.2	1.0	1.0	1.0	1.4	1.1	1.1	0.14
Derived quantities	$\sigma_{disparity}/\sigma_{motion}$	0.97	0.78	1.01		1.79	1.19	1.46	1.84	1.37	1.76	1.35 *	
	$\sigma_{joint}/\sigma_{motion}$	1.19	0.74	0.93	2.0	1.45	2.0	1.9	1.66	1.46	3.38	1.67*	
	$N_{disparity}/N_{motion}$	2.4	2.5	2.43		1.02	1.90	1.36	1.74	1.24	1.21	1.76*	
	N_{joint}/N_{motion}	2.9	2.75	2.81	3.2	1.24	1.81	1.88	2.21	1.19	1.13	2.11*	
	f_{max}^{motion} (cycles per degree)	3.51	2.71	4.45	3.26	6.14	6.25	5.09	7.49	5.71	7.04	5.17	1.64
	$f_{max}^{disparity}$ (cycles per degree)	2.44	1.96	2.87		3.63	4.15	3.05	3.20	3.90	3.43	3.18	0.70
	f_{max}^{joint} (cycles per degree)	1.71	1.46	2.77	1.00	3.96	2.51	1.95	3.04	3.73	1.87	2.40	0.98

Table 3: Fit parameters and derived quantities for all subjects. Asterisks * indicate ratios significantly greater than 1 (t-test on the log-ratios, $p<0.05$). "Mean" is the arithmetic mean except for the 4 rows showing ratios, where it is the geometric mean.

References

Adelson, E. H. and J. R. Bergen (1985). "Spatiotemporal Energy Models for the Perception of Motion." Journal of the Optical Society of America a-Optics Image Science and Vision **2**(2): 284-299.

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

Allenmark, F. and J. C. Read (2010). "Detectability of sine- versus square-wave disparity gratings: A challenge for current models of depth perception." Journal of Vision **10**(8): 1-16.

Allenmark, F. and J. C. Read (2011). "Spatial stereoresolution for depth corrugations may be set in primary visual cortex." PLoS Comput Biol **7**(8): 1-14.

Anderson, S. J. and D. C. Burr (1987). "Receptive field size of human motion detection units." Vision Res **27**(4): 621-35.

Anderson, S. J. and D. C. Burr (1989). "Receptive field properties of human motion detector units inferred from spatial frequency masking." Vision Res **29**(10): 1343-58.

Andrews, T. J., A. Glennerster and A. J. Parker (2001). "Stereoacuity thresholds in the presence of a reference surface." Vision Res **41**(23): 3051-61.

Anzai, A., I. Ohzawa and R. D. Freeman (1999). "Neural mechanisms for processing binocular information I. Simple cells." J Neurophysiol **82**(2): 891-908.

Anzai, A., I. Ohzawa and R. D. Freeman (2001). "Joint-encoding of motion and depth by visual cortical neurons: neural basis of the Pulfrich effect." Nat Neurosci **4**(5): 513-8.

Banks, M. S., S. Gepshtein and M. S. Landy (2004). "Why is spatial stereoresolution so low?" J Neurosci **24**(9): 2077-89.

Banks, M. S., S. Gepshtein and H. F. Rose (2005). "Local cross-correlation model of stereo correspondence." Human Vision and Electronic Imaging X **5666**: 53-61.

Blakemore, C. (1970). "The range and scope of binocular depth discrimination in man." J Physiol **211**(3): 599-622.

Bradley, D. C., G. C. Chang and R. A. Andersen (1998). "Encoding of three-dimensional structure-from-motion by primate area MT neurons." Nature **392**(6677): 714-7.

Bradley, D. C., N. Qian and R. A. Andersen (1995). "Integration of motion and stereopsis in middle temporal cortical area of macaques." Nature **373**(6515): 609-11.

Bradshaw, M. F., P. B. Hibbard, A. D. Parton, D. Rose and K. Langley (2006). "Surface orientation, modulation frequency and the detection and perception of depth defined by binocular disparity and motion parallax." Vision Research **46**(17): 2636-2644.

Bradshaw, M. F. and B. J. Rogers (1999). "Sensitivity to horizontal and vertical corrugations defined by binocular disparity." Vision Res **39**(18): 3049-56.

Brainard, D. H. (1997). "The Psychophysics Toolbox." Spat Vis **10**(4): 433-6.

Bredfeldt, C. E. and B. G. Cumming (2006). "A simple account of cyclopean edge responses in macaque v2." J Neurosci **26**(29): 7581-96.

Bredfeldt, C. E., J. C. Read and B. G. Cumming (2009). "A quantitative explanation of responses to disparity-defined edges in macaque V2." J Neurophysiol **101**(2): 701-13.

Burt, P. and B. Julesz (1980). "A Disparity Gradient Limit for Binocular Fusion." Science **208**(4444): 615-617.
Burt, P. and B. Julesz (1980). "Modifications of the classical notion of Panum's fusional area." Perception **9**(6): 671-82.
Campbell, F. W. and D. G. Green (1965). "Optical and retinal factors affecting visual resolution." J Physiol **181**(3): 576-93.
Campbell, F. W. and R. W. Gubisch (1966). "Optical quality of the human eye." J Physiol **186**(3): 558-78.
Campbell, F. W. and J. G. Robson (1968). "Application of Fourier analysis to the visibility of gratings." J Physiol **197**(3): 551-66.
Cobo-Lewis, A. B. and Y. Y. Yeh (1994). "Selectivity of cyclopean masking for the spatial frequency of binocular disparity modulation." Vision Res **34**(5): 607-20.
Cormack, L. K., S. B. Stevenson and C. M. Schor (1991). "Interocular Correlation, Luminance Contrast and Cyclopean Processing." Vision Research **31**(12): 2195-2207.
Cumming, B. G. (2002). "An unexpected specialization for horizontal disparity in primate primary visual cortex." Nature **418**(6898): 633-636.
Cumming, B. G. and G. C. DeAngelis (2001). "The physiology of stereopsis." Annu Rev Neurosci **24**: 203-38.
Deangelis, G. C., I. Ohzawa and R. D. Freeman (1991). "Depth Is Encoded in the Visual-Cortex by a Specialized Receptive-Field Structure." Nature **352**(6331): 156-159.
DeAngelis, G. C. and T. Uka (2003). "Coding of horizontal disparity and velocity by MT neurons in the alert macaque." J Neurophysiol **89**(2): 1094-111.
Devalois, R. L., D. G. Albrecht and L. G. Thorell (1982). "Spatial-Frequency Selectivity of Cells in Macaque Visual-Cortex." Vision Research **22**(5): 545-559.
Dodd, J. V., K. Krug, B. G. Cumming and A. J. Parker (2001). "Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT." J Neurosci **21**(13): 4809-21.
Filippini, H. R. and M. S. Banks (2009). "Limits of stereopsis explained by local cross-correlation." J Vis **9**(1): 8 1-18.
Fleet, D. J., A. D. Jepson and M. R. M. Jenkin (1991). "Phase-Based Disparity Measurement." Cvgip-Image Understanding **53**(2): 198-210.
Fleet, D. J., H. Wagner and D. J. Heeger (1996). "Neural encoding of binocular disparity: energy models, position shifts and phase shifts." Vision Res **36**(12): 1839-57.
Gattass, R. and C. G. Gross (1981). "Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque." J Neurophysiol **46**(3): 621-38.
Geisler, W. S. and K. D. Davila (1985). "Ideal discriminators in spatial vision: two-point stimuli." J Opt Soc Am A **2**(9): 1483-97.
Georgeson, M. A. and N. E. Scott-Samuel (2000). "Spatial resolution and receptive field height of motion sensors in human vision." Vision Res **40**(7): 745-58.
Gepshtein, S. and A. Cooperman (1998). "Stereoscopic transparency: a test for binocular vision's disambiguating power." Vision Research **38**(19): 2913-2932.
Gillam, B., S. Blackburn and K. Brooks (2007). "Hinge versus twist: the effects of 'reference surfaces' and discontinuities on stereoscopic slant perception." Perception **36**(4): 596-616.
Glennerster, A. (1996). "The time course of 2-D shape discrimination in random dot stereograms." Vision Res **36**(13): 1955-68.

Glennerster, A. (1998). "dmax for stereopsis and motion in random dot displays." Vision Res **38**(6): 925-35.
Glennerster, A. and S. P. McKee (1999). "Bias and sensitivity of stereo judgements in the presence of a slanted reference plane." Vision Res **39**(18): 3057-69.
Graham, N. and J. Nachmias (1971). "Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channels models." Vision Res **11**(3): 251-9.
Grove, P. M. and D. Regan (2002). "Spatial frequency discrimination in cyclopean vision." Vision Res **42**(15): 1837-46.
Haefner, R. M. and B. G. Cumming (2008). "Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model." Neuron **57**(1): 147-58.
Hannah, M. (1974). Computer matching of areas in stereo imagery, Stanford University.
Harris, J. M., S. P. McKee and H. S. Smallman (1997). "Fine-scale processing in human binocular stereopsis." Journal of the Optical Society of America a-Optics Image Science and Vision **14**(8): 1673-1683.
Hess, R. F., F. A. Kingdom and L. R. Ziegler (1999). "On the relationship between the spatial channels for luminance and disparity processing." Vision Res **39**(3): 559-68.
Howard, I. P. and B. J. Rogers (1995). Binocular Vision and Stereopsis, Oxford University Press.
Janssen, P., R. Vogels and G. A. Orban (1999). "Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes." Proc Natl Acad Sci U S A **96**(14): 8217-22.
Julesz, B. (1971). Foundations of cyclopean perception. Chicago, University of Chicago.
Kanade, T. and M. Okutomi (1994). "A Stereo Matching Algorithm with an Adaptive Window - Theory and Experiment." Ieee Transactions on Pattern Analysis and Machine Intelligence **16**(9): 920-932.
Kleiner, M., D. H. Brainard and D. G. Pelli (2007). "What's new in Psychtoolbox-3?" Perception **36**: ECVP Abstract Supplement.
Kulikowski, J. J. (1978). "Limit of single vision in stereopsis depends on contour sharpness." Nature **275**(5676): 126-7.
Lankheet, M. J. M. and P. Lennie (1996). "Spatio-temporal requirements for binocular correlation in stereopsis." Vision Research **36**(4): 527-538.
Majaj, N. J., M. Carandini and J. A. Movshon (2007). "Motion integration by neurons in macaque MT is local, not global." J Neurosci **27**(2): 366-70.
Marcar, V. L., D. K. Xiao, S. E. Raiguel, H. Maes and G. A. Orban (1995). "Processing of kinetically defined boundaries in the cortical motion area MT of the macaque monkey." J Neurophysiol **74**(3): 1258-70.
McKee, S. P. and P. Verghese (2002). "Stereo transparency and the disparity gradient limit." Vision Research **42**(16): 1963-1977.
Mitchison, G. J. and G. Westheimer (1984). "The perception of depth in simple figures." Vision Res **24**(9): 1063-73.
Morgan, M. J. and P. Thompson (1975). "Apparent motion and the Pulfrich effect." Perception **4**(1): 3-18.
Neri, P. and D. M. Levi (2008). "Evidence for joint encoding of motion and disparity in human visual perception." J Neurophysiol **100**(6): 3117-33.

Nguyenkim, J. D. and G. C. DeAngelis (2003). "Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons." J Neurosci **23**(18): 7117-28.
Nienborg, H., H. Bridge, A. J. Parker and B. G. Cumming (2004). "Receptive field size in V1 neurons limits acuity for perceiving disparity modulation." J Neurosci **24**(9): 2065-76.
Ohzawa, I., G. C. DeAngelis and R. D. Freeman (1990). "Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors." Science **249**(4972): 1037-41.
Ohzawa, I., G. C. DeAngelis and R. D. Freeman (1997). "Encoding of binocular disparity by complex cells in the cat's visual cortex." J Neurophysiol **77**(6): 2879-909.
Orban, G. A., P. Dupont, B. De Bruyn, R. Vogels, R. Vandenberghe and L. Mortelmans (1995). "A motion area in human visual cortex." Proc Natl Acad Sci U S A **92**(4): 993-7.
Panton, D. J. (1978). "Flexible Approach to Digital Stereo Mapping." Photogrammetric Engineering and Remote Sensing **44**(12): 1499-1512.
Parker, A. J. (2007). "Binocular depth perception and the cerebral cortex." Nature Reviews Neuroscience **8**(5): 379-391.
Pelli, D. G. (1997). "The VideoToolbox software for visual psychophysics: transforming numbers into movies." Spat Vis **10**(4): 437-42.
Prince, S. J., B. G. Cumming and A. J. Parker (2002). "Range and mechanism of encoding of horizontal disparity in macaque V1." J Neurophysiol **87**(1): 209-21.
Prince, S. J. and R. A. Eagle (1999). "Size-disparity correlation in human binocular depth perception." Proc Biol Sci **266**(1426): 1361-5.
Qian, N. (1994). "Computing Stereo Disparity and Motion with Known Binocular Cell Properties." Neural Computation **6**(3): 390-404.
Qian, N. and R. A. Andersen (1994). "Transparent motion perception as detection of unbalanced motion signals. II. Physiology." J Neurosci **14**(12): 7367-80.
Qian, N. and R. A. Andersen (1997). "A physiological model for motion-stereo integration and a unified explanation of Pulfrich-like phenomena." Vision Res **37**(12): 1683-98.
Qian, N., R. A. Andersen and E. H. Adelson (1994). "Transparent motion perception as detection of unbalanced motion signals. III. Modeling." J Neurosci **14**(12): 7381-92.
Qian, N. and R. D. Freeman (2009). "Pulfrich phenomena are coded effectively by a joint motion-disparity process." J Vis **9**(5): 24 1-16.
Qian, N. and S. Mikaelian (2000). "Relationship between phase and energy methods for disparity computation." Neural Computation **12**(2): 279-292.
Qian, N. and Y. D. Zhu (1997). "Physiological computation of binocular disparity." Vision Research **37**(13): 1811-1827.
Raiguel, S., M. M. Van Hulle, D. K. Xiao, V. L. Marcar and G. A. Orban (1995). "Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque." Eur J Neurosci **7**(10): 2064-82.
Read, J. C. A. (2002). "A Bayesian model of stereopsis depth and motion direction discrimination." Biol Cybern **86**(2): 117-36.
Read, J. C. A. (2005). "Early computational processing in binocular vision and depth perception." Progress in Biophysics & Molecular Biology **87**(1): 77-108.

Read, J. C. A. (2010). "Vertical Binocular Disparity is Encoded Implicitly within a Model Neuronal Population Tuned to Horizontal Disparity and Orientation." Plos Computational Biology **6**(4): 1-15.
Read, J. C. A. and B. G. Cumming (2004). "Understanding the cortical specialization for horizontal disparity." Neural Comput **16**(10): 1983-2020.
Read, J. C. A. and B. G. Cumming (2005a). "The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth." J Vis **5**(5): 417-34.
Read, J. C. A. and B. G. Cumming (2005b). "Effect of interocular delay on disparity-selective v1 neurons: relationship to stereoacuity and the pulfrich effect." J Neurophysiol **94**(2): 1541-53.
Read, J. C. A. and B. G. Cumming (2005c). "All Pulfrich-like illusions can be explained without joint encoding of motion and disparity." J Vis **5**(11): 901-27.
Read, J. C. A. and B. G. Cumming (2006). "Does depth perception require vertical-disparity detectors?" J Vis **6**(12): 1323-55.
Read, J. C. A. and B. G. Cumming (2007). "Sensors for impossible stimuli may solve the stereo correspondence problem." Nat Neurosci **10**(10): 1322-8.
Read, J. C. A. and R. A. Eagle (2000). "Reversed stereo depth and motion direction with anti-correlated stimuli." Vision Res **40**(24): 3345-58.
Read, J. C. A., A. J. Parker and B. G. Cumming (2002). "A simple model accounts for the response of disparity-tuned V1 neurons to anticorrelated images." Vis Neurosci **19**(6): 735-53.
Reppas, J. B., S. Niyogi, A. M. Dale, M. I. Sereno and R. B. Tootell (1997). "Representation of motion boundaries in retinotopic human visual cortical areas." Nature **388**(6638): 175-9.
Roe, A. W., A. J. Parker, R. T. Born and G. C. DeAngelis (2007). "Disparity channels in early vision." J Neurosci **27**(44): 11820-31.
Rossi, E. A. and A. Roorda (2009). "The relationship between visual resolution and cone spacing in the human fovea." Nat Neurosci **13**(2): 156-157.
Sakata, H., M. Taira, M. Kusunoki, A. Murata, K. Tsutsui, Y. Tanaka, W. N. Shein and Y. Miyashita (1999). "Neural representation of three-dimensional features of manipulation objects with stereopsis." Exp Brain Res **128**(1-2): 160-9.
Schor, C. M. and I. Wood (1983). "Disparity range for local stereopsis as a function of luminance spatial frequency." Vision Res **23**(12): 1649-54.
Schumer, R. and L. Ganz (1979). "Independent stereoscopic channels for different extents of spatial pooling." Vision Res **19**(12): 1303-14.
Serrano-Pedraza, I., G. P. Phillipson and J. C. Read (in press). "A specialization for vertical disparity discontinuities." Journal of Vision.
Serrano-Pedraza, I. and J. C. Read (2009). "Horizontal/vertical anisotropy in sensitivity to relative disparity depends on stimulus depth structure." Perception **38**: 155-155.
Serrano-Pedraza, I. and J. C. A. Read (2010). "Multiple channels for horizontal, but only one for vertical corrugations? A new look at the stereo anisotropy." Journal of Vision **10**(12): 1-11.
Smallman, H. S. and D. I. MacLeod (1994). "Size-disparity correlation in stereopsis at contrast threshold." J Opt Soc Am A Opt Image Sci Vis **11**(8): 2169-83.
Snowden, R. J., S. Treue, R. G. Erickson and R. A. Andersen (1991). "The response of area MT and V1 neurons to transparent motion." J Neurosci **11**(9): 2768-85.

Snyder, L. H., A. P. Batista and R. A. Andersen (2000). "Intention-related activity in the posterior parietal cortex: a review." Vision Res **40**(10-12): 1433-41.
Steingrube, P., S. K. Gehrig and U. Franke (2009). "Performance Evaluation of Stereo Algorithms for Automotive Applications." Computer Vision Systems, Proceedings **5815**: 285-294.
Sugihara, H., I. Murakami, K. V. Shenoy, R. A. Andersen and H. Komatsu (2002). "Response of MSTd neurons to simulated 3D orientation of rotating planes." J Neurophysiol **87**(1): 273-85.
Thomas, O. M., B. G. Cumming and A. J. Parker (2002). "A specialization for relative disparity in V2." Nat Neurosci **5**(5): 472-8.
Tsai, J. J. and J. D. Victor (2003). "Reading a population code: a multi-scale neural model for representing binocular disparity." Vision Research **43**(4): 445-466.
Tsirlin, I., R. S. Allison and L. M. Wilcox (2008). "Stereoscopic transparency: Constraints on the perception of multiple surfaces." Journal of Vision **8**(5): 1-10.
Tyler, C. W. (1973). "Steroscopic vision: cortical limitations and a disparity scaling effect." Science **181**(96): 276-8.
Tyler, C. W. (1974). "Depth perception in disparity gratings." Nature **251**(5471): 140-2.
Tyler, C. W. (1975). "Spatial organization of binocular disparity sensitivity." Vision Res **15**(5): 583-90.
Tyler, C. W. (1975). "Stereoscopic Tilt and Size Aftereffects." Perception **4**(2): 187-192.
van der Willigen, R. F., W. M. Harmening, S. Vossen and H. Wagner (2010). "Disparity sensitivity in man and owl: Psychophysical evidence for equivalent perception of shape-from-stereo." Journal of Vision **10**(1): 1-11.
von der Heydt, R., H. Zhou and H. S. Friedman (2000). "Representation of stereoscopic edges in monkey visual cortex." Vision Res **40**(15): 1955-67.
Wallace, J. M. and P. Mamassian (2004). "The efficiency of depth discrimination for non-transparent and transparent stereoscopic surfaces." Vision Research **44**(19): 2253-2267.
Watson, A. B. and D. G. Pelli (1983). "QUEST: a Bayesian adaptive psychometric method." Percept Psychophys **33**(2): 113-20.
Zhu, Y. D. and N. Qian (1996). "Binocular receptive field models, disparity tuning, and characteristic disparity." Neural Comput **8**(8): 1611-41.
Ziegler, L. R., R. F. Hess and F. A. Kingdom (2000). "Global factors that determine the maximum disparity for seeing cyclopean surface shape." Vision Res **40**(5): 493-502.