Perception, 1997, volume 26, pages 977–994

Spatial scale interactions in stereo sensitivity and the neural representation of binocular disparity

Harvey S Smallman¶, Donald I A MacLeod
Department of Psychology, University of California at San Diego, La Jolla, CA 92093-0109, USA;
¶ also at Centre for Vision and Visual Cognition, Department of Psychology, University of Durham, Durham DH1 3LE, UK; e-mail: smallman@psy.ucsd.edu
Based on a paper presented at the Durham Workshop on Spatial Scale Interactions, Durham, UK, 16–18 September 1996
Received 15 November 1996, accepted 26 February 1997

Abstract. How are binocular disparities encoded and represented in the human visual system? An 'encoding cube' diagram is introduced to visualise differences between competing models. To distinguish the models experimentally, the depth-increment-detection function (discriminating disparity d from $d \pm \Delta d$) was measured as a function of standing disparity (d) with spatially filtered random-dot stereograms of different centre spatial frequencies. Stereothresholds degraded more quickly as standing disparity was increased with stimuli defined by high rather than low centre spatial frequency. This is consistent with a close correlation between the spatial scale of detection mechanisms and the disparities they process. It is shown that a simple model, where discrimination is limited by the noisy ratio of outputs of three disparity-selective mechanisms at each spatial scale, can account for the data. It is not necessary to invoke a population code for disparity to model the depth-increment-detection function.

This type of encoding scheme implies insensitivity to large interocular phase differences. Might the system have developed a strategy to disambiguate or shift the matches made at fine scales with those made at the coarse scales at large standing disparities? In agreement with Rohaly and Wilson, no evidence was found that this is so. Such a scheme would predict that stereothresholds determined with targets composed of compounds of high and low frequency should be superior to those of either component alone. Although a small stereoacuity benefit was found at small disparities, the more striking result was that stereothresholds for compound-frequency targets were actually degraded at large standing disparities. The results argue against neural shifting of the matching range of fine scales by coarse-scale matches posited by certain stereo models.

1 Introduction

The view that the senses are involved in the active construction of an internal representation of the external world has commanded wide belief since at least the time of Descartes. Vision may thus be considered the end result of a causal chain of processes that extract and abstract information from a hierarchy of neural representations of the two retinal images. Marr's (1982) work represents the most eloquent recent advocation of this position. In order to understand binocular stereopsis, on this view, an understanding of how binocular disparities are initially detected and represented in the human visual system is essential. But it is still unclear how this happens despite a century and a half of work since Wheatstone's (1838) initial demonstrations that binocular disparities form the basis of binocular stereopsis.

A number of subquestions can be identified. First, how are binocular disparities initially detected? Second, how many disparity detectors are required to 'code', or neurally represent, disparity at a given retinal locus? Third, what is the nature of the readout from this code to extract binocular disparity? Fourth, what role does spatial scale play and how does stereo processing at one spatial scale influence processing at another? In this paper we report psychophysical data and computational modelling that suggests possible answers to these questions. To take the second question first, there are several candidate schemes for the representation of binocular disparity [see Lehky et al

(1990) and Howard and Rogers (1995) for recent reviews]. To visualise these schemes we introduce an 'encoding-cube' diagram to distinguish competing models of neural representation of a stimulus attribute such as binocular disparity (see figure 1).

Models plot as points inside the encoding cube. The x-dimension of the cube represents the breadth of tuning of a mechanism to the stimulus attribute (from coarse, or broadly tuned, to fine, or narrowly tuned). The y-dimension represents the number of mechanisms in the code at a given retinal locus (from a discrete code with two or three mechanisms, to a continuous code with many mechanisms). The depth of the cube, or z-dimension, represents the number of mechanisms involved in the readout scheme from the code (from listening to one mechanism singing loudest, or winner takes all, to an ensemble code weighting all responses such as the centre of gravity of the responses).[1] Some well-known models are plotted in the cube. Richards (1970, 1971) suggested, by analogy with the Young–Helmholtz theory of colour vision, that there might be just three mechanisms, or pools of disparity detectors, each containing units broadly tuned to disparity, with one set tuned to crossed disparities, one to uncrossed, and one to disparities near the fixation plane. This is one example of a *coarse-code* model of disparity encoding. It plots at the bottom back left-hand corner of the encoding cube. Richards's code received some support from the first physiological recordings of disparity-selective neurons in awake behaving monkey visual area V1

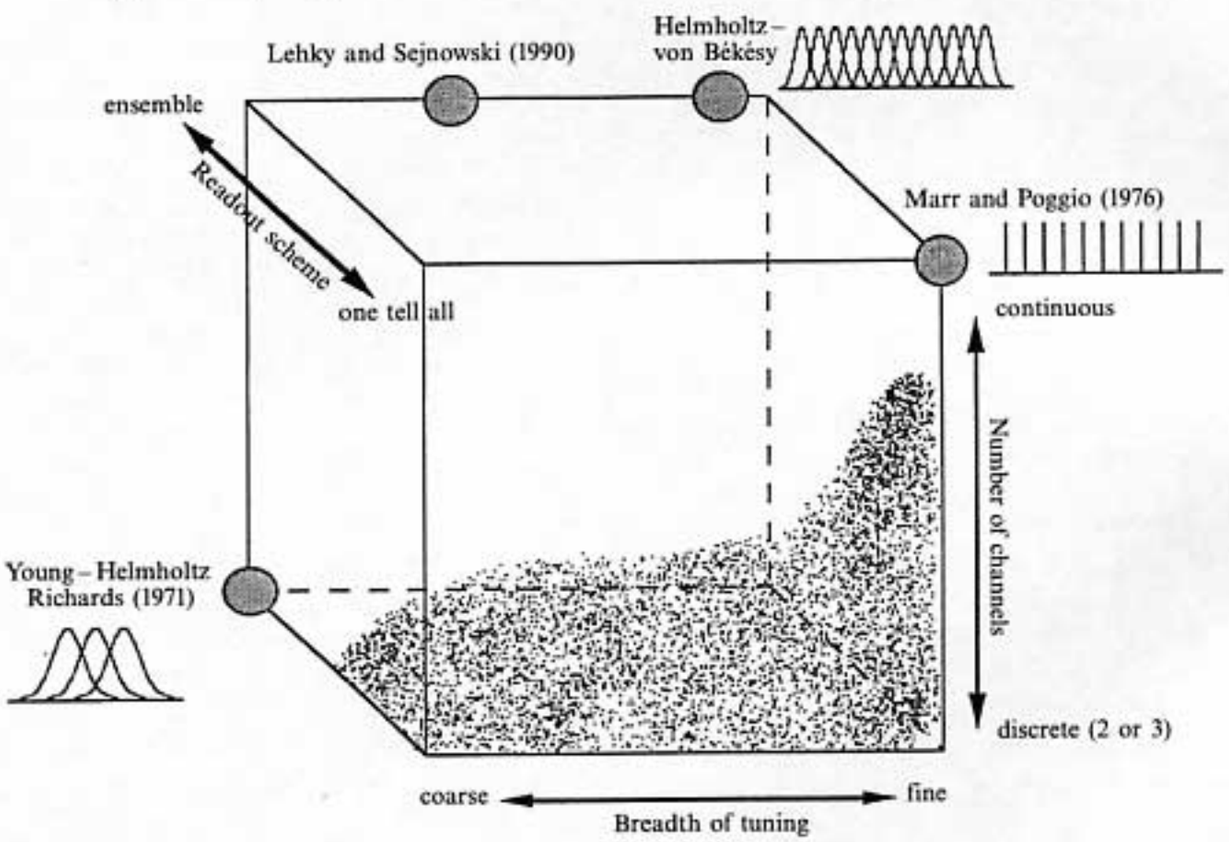

Figure 1. The encoding cube, a way of visualising competing models of the neural representation of a visual attribute, here binocular disparity. Models, shown by filled circles, plot as points in the cube. The x-dimension of the cube represents the breadth of tuning of the mechanisms making up the code for the stimulus attribute. The y-dimension represents the number of mechanisms in the representation. The z-dimension represents the number of mechanisms employed in the readout from the code. The dimensions are not necessarily linear. The dotted region represents the space of models that are not sensible to adhere to.

[1] An alternative way of conceiving of this dimension, suggested to us by Dr C W Tyler, is to think of it as the exponent (Q) to which the response of the mechanisms of the code is raised in a Quick-like readout scheme (Quick 1974). A low value of Q (eg 1) represents summation across all members of the code (an ensemble readout). A large value of Q (eg 4) weights most strongly the mechanisms responding most vigorously (ie essentially a winner-takes-all readout scheme).

(Poggio and Fischer 1977). But in more-recent physiology it has not been possible to find cells with disparity selectivity clustering into just three pools (LeVay and Voigt 1988; Poggio et al 1988).

The cube possesses a region (depicted as dotted in figure 1) where sensible models fear to tread. Because Richards's scheme employed just three mechanisms it *had* to employ an ensemble readout scheme. If Richards's scheme employed a 'one-tell-all' readout scheme, that is, if it plotted at the bottom front left corner of the cube in the dotted region, observers would never be able to distinguish more than three possible depths.

Alternatively, analogous to the Helmholtz–von Békésy code for audition (Helmholz 1863; Békésy 1968), which plots at the top right-hand back corner of the cube, there might be many disparity-selective mechanisms at a given retinal locus with narrow, but overlapping, tuning for disparity. This type of code is sometimes called a *fine code*, or in more-recent parlance, a *population code* (Lehky et al 1990). A population-code model for disparity has been advanced by Lehky and Sejnowski (1990) because it could be made to account for the falloff of stereoacuity away from the fixation plane. How it did so will be discussed in more detail later. There is recent psychophysical support for this model over a coarse-code formulation (Stevenson et al 1992; Cormack et al 1993). Lehky and Sejnowski's model plots at the centre back of the top of the cube because it differs from a true disparity analogue of a Helmholtz–von Békésy fine code by possessing some mechanisms with quite broad tuning for disparity.

The representation of binocular disparity implied by the Marr and Poggio (1976) theory of stereo vision, which plots at the front top right-hand corner of the cube, is even finer than that in the Helmholtz–von Békésy fine code inasmuch as their disparity-selective mechanisms are presumed to respond each to one unique disparity. Note that if one gradually removes mechanisms from the Marr and Poggio (1976) code, that is, one slips down the y-axis from its position, then one again moves into the dotted region. One cannot adhere to a model with a spiky representation of disparity with only a few mechanisms in the code. A visual system possessed of such a code would 'miss' certain planes in depth. We propose a hybrid model of disparity encoding in this paper which shares several features of Richards's code and that of Lehky and Sejnowski.

How are disparities initially detected? Meshing naturally with the idea of a population code for disparity is the default model of how neurons obtain their selectivity for binocular disparity. Since the initial reports from the 1960s of binocular-disparity-selective neurons in cat visual cortex (Barlow et al 1967; Pettigrew et al 1968) the default model of disparity detection has been that binocular neurons obtain that selectivity from having positional offsets of identical monocular receptive fields (RFs) in the two eyes (Maske et al 1984). If it is assumed, quite plausibly, that there are detectors fed by a spread of positional offsets at a given retinal locus, and hence many mechanisms tuned to disparity, then one has constructed a population code.

However, the most recent mappings of the substructure of monocular RFs of binocular neurons, again in cat, has suggested that the RFs may *not* be identical in the two eyes (DeAngelis et al 1991, 1995). Specifically, it has been suggested that a spatial phase difference between approximately sinusoidal RFs modulated by Gaussian windows that have no positional offset in the two eyes might be the basis of disparity encoding (Freeman and Ohzawa 1990; Ohzawa et al 1990, 1996). With this type of encoding scheme there could be either a coarse or a population code for disparity. To make a coarse code, a fixation-plane pool could be constructed from units possessing 0° interocular phase difference in their monocular RFs, with the crossed and uncrossed pools having $\pm 90°$ phase differences (Ohzawa et al 1990; DeAngelis et al 1991). Alternatively, to make a population code there would be a variety of more-finely graded interocular phase differences in the monocular RFs subserving binocular mechanisms at a given

RF locus (DeAngelis et al 1995; Ohzawa et al 1996). Recently, others have begun to envisage models of disparity encoding which entail a hybrid scheme of positional offsets and phase to encode binocular disparity (Jacobsen et al 1993; Fleet et al 1996).

Irrespective of whether there is a population or coarse code for disparity, the phase model, unlike the simplest positional model, makes certain key psychophysical predictions about the range of disparities that give rise to stereopsis. It is well established in cats and monkeys that RFs of simple cells come in a range of different sizes and that the linearity of spatial integration within their RFs endows them with selectivities for different ranges of image spatial frequencies (Movshon et al 1978; DeAngelis et al 1993). Large RFs select for low spatial frequencies (coarse scales) and small RFs select for high spatial frequencies (fine scales). Humans are also thought to possess similar mechanisms (Campbell and Robson 1968; Wilson et al 1990; Smallman et al 1996; Wilson and Wilkinson 1997). In the phase-disparity model, the disparity a cell is optimally responsive to is a fixed fraction of RF width. Thus, the range of disparities that support stereopsis should vary with the spatial-frequency content of a stimulus in such a way that large disparities should only be visible when the stimulus contains low spatial frequencies. This is the so-called 'size–disparity correlation' (Schor and Wood 1983; see Smallman and MacLeod 1994a for a recent review). There is consensus that the size–disparity correlation is supported for low spatial frequencies (<2.5 cycles deg^{-1}) with the range of disparities that supports stereopsis scaling with spatial frequency (Schor and Wood 1983; Smallman and MacLeod 1994a). At high spatial frequencies the size–disparity relationship breaks down in some studies, with the disparity range for stereopsis constant for frequencies above 2.5 cycles deg^{-1} (Schor and Wood 1983; Legge and Gu 1989). However, we previously (Smallman and MacLeod 1994a) used low-contrast-filtered random-dot stereograms (RDSs) to show that the range of disparities decreases with increasing spatial frequency across all spatial frequencies tested (1–15 cycles deg^{-1}).

Phase-disparity encoding and the size–disparity correlation make certain key predictions about stereo performance away from the plane of fixation. It is well established that stereoacuity falls off rapidly away from the plane of fixation (Ogle 1953; Blakemore 1970; McKee et al 1990). As the only spatial frequencies that should support stereopsis at large standing disparities are low ones then depth discrimination should fall off faster away from the fixation plane for high spatial frequencies than for low. But the available evidence so far has not supported this prediction (Mayhew and Frisby 1979; Badcock and Schor 1985; Siderov and Harwerth 1993a, 1995). These researchers have found instead an essentially spatial-frequency-invariant depth-increment-detection function.

How can these studies be reconciled with phase-disparity encoding and with the data that support the size–disparity correlation? In this paper we develop arguments to suggest that they can be reconciled and we provide evidence that the depth-increment-detection function does indeed follow the size–disparity correlation under certain conditions. We also show that the depth-increment function can be modelled with a coarse code for disparity—it is not necessary to invoke a population code to model the function as Lehky and Sejnowski (1990) suggested.

Phase-disparity encoding predicts insensitivity to large interocular phase differences (eg large disparities specified by high spatial frequencies). This raises an intriguing question. In fovea, stereoacuity in the fixation plane is best for high spatial frequencies (Siderov and Harwerth 1993a, 1995). However, matching ambiguity is potentially worst at these frequencies (Marr and Poggio 1979; but see Harris et al 1997). In certain computational models, to obviate this problem, matches made at coarse spatial scales shift the range of matching of fine spatial scales (Marr and Poggio 1979; Nishihara 1984; Quam 1987; see figure 2). This could be accomplished by several different mechanisms. In Marr and Poggio's well-known 1979 theory, vergence eye movements, driven by initial

disparity estimates from coarse scales, accomplished the shift by moving the fine scales into their restricted matching range. In other models (Nishihara 1984; Quam 1987) the shifting was done neurally in the absence of eye movements by something akin to the shifter circuits proposed by Anderson and van Essen (1987). Last, there is another possible mechanism which could achieve the same goals as a shift without the necessity for eye movements or elaborate circuitry. Cells tuned to high spatial frequencies often exhibit a tighter tuning to spatial frequency by virtue of having more cycles in their RF structure (Mullikin et al 1984; DeAngelis et al 1995). Responses from such cells might be expected to be more periodic with disparity and hence more ambiguous. Coarse-scale matches could resolve this ambiguity at large standing disparities and hence give rise to the same effect as a shift in matching range. There is evidence that matches made at one spatial scale can resolve matching ambiguity at another (Smallman 1995; Mallot et al 1996). If coarse scales could shift the matching range of fine scales in the absence of eye movements then it should be the case that stereothresholds for compound targets (of high plus low spatial frequency) should be better than low alone at large standing disparities. However, in agreement with the earlier study of Rohaly and Wilson (1993a), who reasoned similarly, we come to the same conclusion that they cannot.

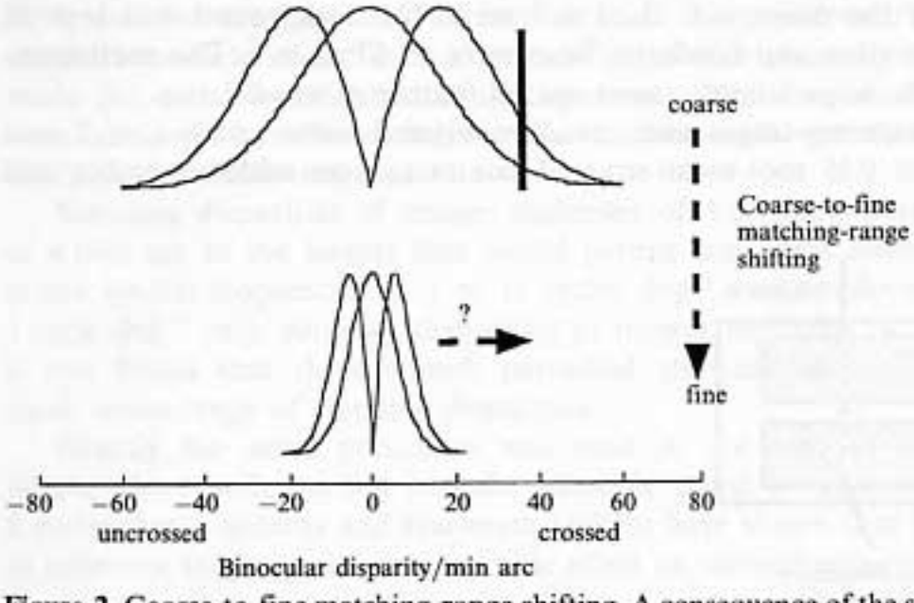

Figure 2. Coarse-to-fine matching-range shifting. A consequence of the size – disparity correlation in disparity encoding is an insensitivity to large disparities specified by high spatial frequencies. Stereothresholds for the thick black vertical line target at 40 min arc would be poor as they are determined only by coarse spatial scales which see it and not by fine scales which do not see it. Might the system overcome this deficiency by shifting the matching range of fine scales with a signal from the coarse? If so then stereothresholds for compound targets made of high plus low spatial frequency should be better than low alone at large standing binocular disparities. We find no evidence for such a scheme.

2 Methods

2.1 *Apparatus and stimuli*

The stimuli were two-dimensional spatially filtered RDSs, henceforth called filtered-noise stereograms of the same bandwidth of ± 0.5 cycle deg^{-1} of centre frequency as were employed previously (Smallman and MacLeod 1994a). They were produced in exactly the same fashion and on the same equipment as in that study, hence only brief descriptions are presented here.

Stimuli were presented on a Taxan Ultra Vision 1000 RGB monitor driven by a Macintosh IIcx microcomputer. A linearising colour look-up table was employed to produce desired intensity profiles and the 8 bit red, green, and blue outputs were combined to yield 12 bit DAC (digital to analog conversion) precision on the green

monitor input, allowing very fine contrast modulations, by using the video attenuator hardware and software of Pelli and Zhang (1991). The monitor was viewed at 57 cm by the subjects through a simple mirror stereoscope which permitted each eye to view a different region of the screen independently.

On each trial two random patches of the filtered noise, 3.1 deg × 0.8 deg, were displayed to produce a stereogram. The centre spatial frequency of the noise was dependent on the particular condition being run. These patches were used to create two disparate panels, one above and one below a fixation cross (see figure 3a). The noise in both panels was given convergent (crossed) disparity by laterally shifting the noise presented in each without shifting the panel borders. Hence monocularly there were no cues available to subjects as to the binocular disparity of the noise. This technique contrasts with other methods of measuring the depth-increment-detection function with local targets where monocular cues were present (eg Badcock and Schor 1985). Subjects were forced to decide which panel lay closer to them in depth. The top panel always had the same disparity during a run. The disparity of the bottom panel was controlled, in an adaptive way, by the subject's responses in a staircase procedure (see figure 3b). Nonius lines, of length 75 min arc, were continuously visible throughout the experiment along with an outline frame 4.1 deg × 1.25 deg in the fixation plane. The mean luminance of the noise was 33 cd m^{-2} while the background was kept at 5 cd m^{-2}. The fixation, nonius, and bordering lines were all 67 cd m^{-2}. The root-mean-squared contrast of all the noise stimuli across spatial frequency was 0.3.

In the compound-frequency-target case, random filtered noise patches of 2 and 8 cycles deg^{-1}, each with 0.15 root-mean-squared contrast, were added together and presented as above.

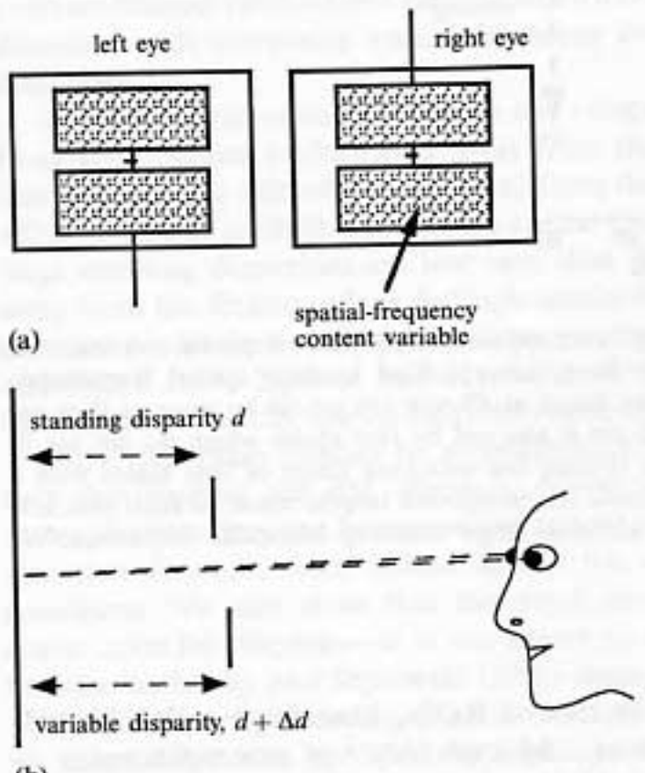

Figure 3. Schematic diagram depicting the stimulus and the depth relationships it gave rise to. (a) The left and right eyes' views which comprised the stereogram. The observer fixated the central cross, and filtered texture patterns with different convergent disparities were presented to above and below. (b) The depth relationships present during a typical trial. The observer's task was to determine whether the top or bottom texture patch lay closer. The disparity of the bottom patch was modulated on successive trials to determine a stereothreshold. To ensure fixation in the plane of the cross the observer was instructed to redo any trials on which the nonius lines were not collinear.

2.2 *Procedure*

The subject aligned the nonius lines and then hit a key to initiate stimulus presentation. The stimuli were then displayed for 250 ms, a duration insufficient to allow the completion of vergence eye movements (Rashbass and Westheimer 1961). On a given trial both panels had different crossed disparity; either could possess larger crossed disparity. When the stimuli had been removed the subjects indicated whether they had perceived the top or the bottom panel as lying closer to them. Stereothresholds for this task were determined with a modified staircase procedure (Levitt 1971). The disparity of the bottom filtered patch was modulated on successive trials by one 3-up-1-down and one 1-up-2-down staircase (Levitt 1971). These two staircases were randomly interwoven for each spatial-frequency/disparity pairing condition so that on any given trial subjects were unaware which staircase they were on. Forty-five data trials were run for each staircase, with an additional three practice trials beforehand. A condition was thus run as a block of ninety-six trials. Different conditions were run in a random order for both subjects. No feedback was given concerning whether responses were correct or incorrect. Data for the two staircases constituting a condition were combined and a cumulative normal was fit to the data (see Mulligan and MacLeod 1988). Stereothresholds were conventionally defined as the disparity difference between that giving 50% and that leading to 75% correct from the psychometric function. Data for a subject were determined over a series of days with generally three to four runs being made for each condition; although at some spatial frequencies more runs than this were made per condition. Reported stereothresholds are the weighted means (Klein 1992) from the different runs.

Standing disparities of integer multiples of 4 min arc were used, from the lowest of 4 min arc to the largest that would permit consistent stereothresholds. A range of centre spatial frequencies of 1 to 11 cycles deg^{-1} was employed. For the frequency of 1 cycle deg^{-1} only, standing disparities of integer multiples of 10 min arc were used as it was found that these stimuli permitted stereothresholds to be measured over a much wider range of standing disparities.

Exactly the same procedure was used in the case of the compound-frequency target, where both the test and the reference panel were composed of noise of 2 and 8 cycles deg^{-1}. Siderov and Harwerth (1993b) have shown that the spectral composition of reference targets has no measurable effect on stereothresholds.

2.3 *Subjects*

Two subjects participated in this study. One was the author, HSS. The other, DJP, was naive as to experimental intent. Both were experienced observers with good stereopsis and normal or corrected-to-normal visual acuity. Further data taken for another experienced observer confirmed the main trends in the data at two centre spatial frequencies.

3 Results and modelling

The stereothresholds for the single-frequency targets and the compound-frequency targets were collected over the same course of time. For clarity of exposition the results are presented separately.

3.1 *Depth-increment-detection function with single-frequency targets*

Mean stereothresholds are plotted against standing disparity for a range of noise centre spatial frequencies from 1 to 11 cycles deg^{-1} for the two observers in figure 4. The higher centre spatial frequencies are plotted in solid symbols and the lower ones in open symbols.

Stereothresholds for all spatial frequencies rose away from the fixation plane. This effect has been robust in the literature and is repeated here (Ogle 1953; Blakemore 1970; Regan and Beverly 1973; Westheimer and McKee 1978; Badcock and Schor 1985;

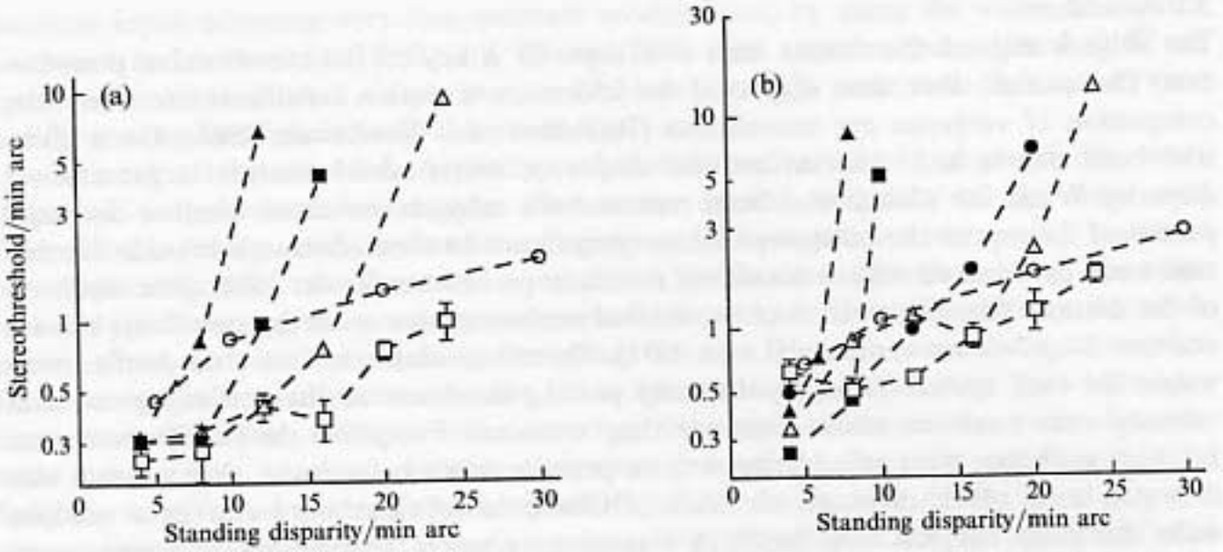

Figure 4. Stereothresholds away from the plane of fixation for a range of centre spatial frequencies of the filtered texture patterns. The *x*-axis shows standing disparity, the *y*-axis depicts stereothresholds. Data are shown for two observers, DJP in (a) and HSS in (b). The different symbols depict different centre spatial frequencies of the filtered texture (○ 1, □ 2, △ 3, ● 5, ■ 8, and ▲ 11 cycles deg^{-1}). Error bars, depicting 95% confidence intervals, are only shown on the 2 cycles deg^{-1} data, for clarity. Stereothresholds rose faster away from the fixation plane for stimuli defined by high spatial frequencies (solid symbols) rather than low spatial frequencies (open symbols).

McKee et al 1990; Siderov and Harwerth 1993a, 1995). The shape of the stereothreshold rise has been reported to be an exponential in threshold as a function of disparity (Blakemore 1970; Siderov and Harwerth 1993a). In fact, here the curves for most centre spatial frequencies accelerate on the log–linear axes of figure 4, hence an exponential fit to the data is poor. There was no hint of a discontinuity or flattening in the functions that could be measured for disparities greater than 20 min arc as is present in other data (Regan and Beverly 1973; Badcock and Schor 1985). The shape of the functions will be discussed in further detail later; of more-immediate interest is the variation in threshold with spatial frequency.

This variation is dramatic, and consistent between the two observers. The increase in thresholds for the higher-spatial-frequency targets is extremely rapid as the stimulus moves away from the fixation plane. At 11 cycles deg^{-1}, an increase in standing disparity from 4 min arc to 12 min arc leads to an approximate 2000% increase in stereothreshold (averaged across both observers). But at the lower spatial frequency of 2 cycles deg^{-1}, the same increase in standing disparity leads to a modest 30% increase in thresholds.

The largest disparity that permitted stereoacuity measurement at 11 cycles deg^{-1} was 8 min arc for HSS and 12 min arc for DJP. As spatial frequency decreased from 11 cycles deg^{-1} the largest disparity permitting stereothreshold measurements systematically increased; the same quantitative trend was observed previously in contrast-threshold measurements (Smallman and MacLeod 1994a) which involved the same class of stimuli.

The systematic drop-off in maximum disparity permitting stereoacuity measurement with spatial frequency suggested to us that a more instructive way to view the data would be to transform the disparities into interocular phases, as we also did with the contrast-threshold data (Smallman and MacLeod 1994a). For example, a disparity of 15 min arc at a centre frequency of 1 cycle deg^{-1} corresponded to a 90° binocular phase. Data from figure 4 are replotted this way in figure 5.

Figure 5 reveals a systematic variation in (now) threshold interocular phase as a function of standing binocular phase. All of the data for the different spatial frequencies which were distinct in figure 4 now collapse approximately onto a single function in

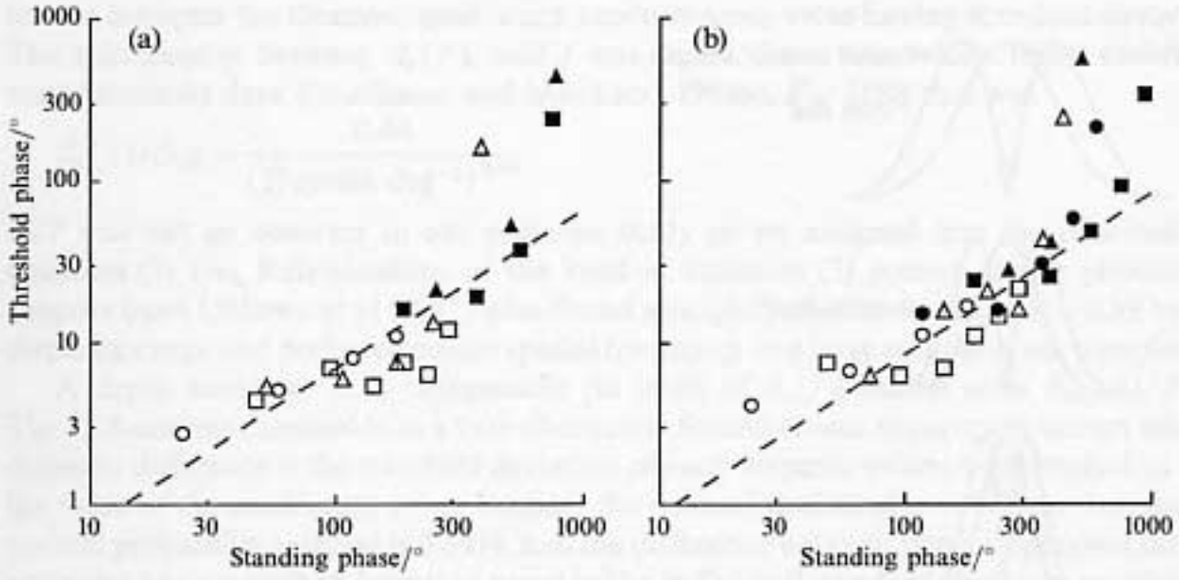

Figure 5. Stereothresholds from figure 4 replotted as threshold interocular phase differences (y-axis) as a function of standing interocular phase (x-axis). All of the data now collapse onto a single function. Data are shown for the two observers, DJP in (a) and HSS in (b). The key for the different spatial frequencies is the same as that in figure 4. To replot the data disparities were converted to interocular phases with the formula

$$\text{phase}/^\circ = 360\,\frac{\text{disparity/min arc}}{60}\,\text{spatial frequency/deg}^{-1}.$$

The dashed line depicts Weber's-law behaviour. This holds over a range of interocular phases but at larger phases thresholds are worse than Weber's law.

figure 5. This is good evidence for the notion that phase disparities may be encoded by the visual system (Freeman and Ohzawa 1990; Ohzawa et al 1990, 1996; DeAngelis et al 1991, 1995). Or, if one adhered to a position model, it is good evidence that the spread of positional offsets inversely correlates with spatial frequency.

Others have found that the depth-increment-detection function exhibits Weber's-law behaviour (McKee et al 1990). Our data obey Weber's law over a restricted range of standing phases (Φ) from approximately 90° to 360° ($\Delta\Phi/\Phi = 0.08$, the dashed line in figure 5). This means that, for those standing phases, an 8% change in standing phase was detectably different for our observers across all experimental conditions. At larger binocular phases the observers' performance is worse than that predicted by Weber's law (the data lie above the dashed line).

How can we account for this? If we assume that the data for crossed and uncrossed standing disparities are symmetrical reflections of the data of figure 5 about the vertical axis then it is simple to account for the bowl-shaped threshold surface with a very simple ratio model of depth discrimination. We assume that at a given retinal locus at each spatial scale there is a Richards-like coarse code for disparity [see figure 6a; see DeAngelis et al (1991) for a similar scheme]. There are three pools of disparity detectors; a near pool for crossed disparities, a far pool for uncrossed disparities, and a horopter pool for disparities around the fixation plane. Observers base their estimate of crossed disparity, say for the purposes of depth discrimination, on the ratio of neural activity in the near to the horopter pool.

We needed to choose disparity-tuning functions for the mechanisms constituting the neural pools. Several possible functions have been discussed in the literature. Ohzawa et al (1990, 1997) found that Gabor functions fit their complex-cell disparity-response profiles well. Other modellers have used a difference-of-Gaussians (DOG) function (Lehky and Sejnowski 1990; Stevenson et al 1992). We chose instead to use a simple Gaussian of disparity for the horopter pool, and Gaussian derivatives of disparity

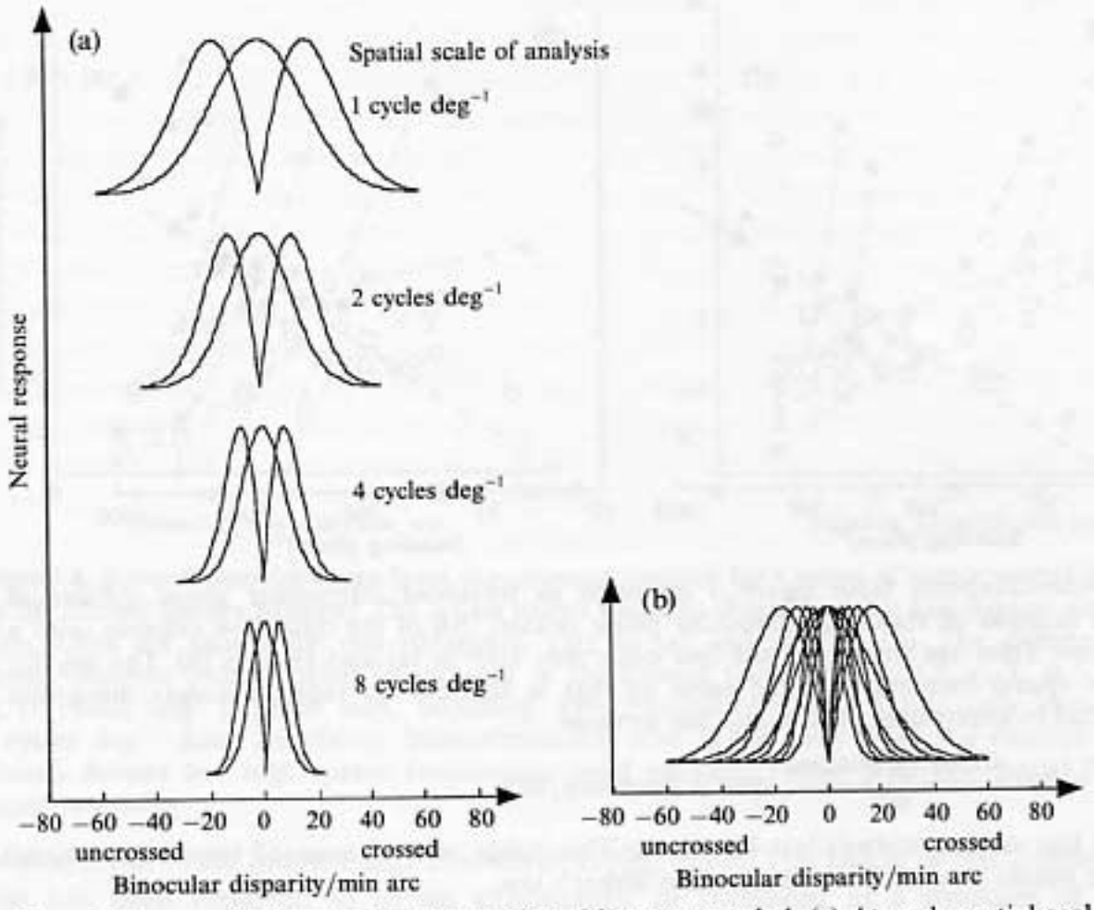

Figure 6. A quantitative model of how disparities are encoded. (a) At each spatial scale of analysis there is a coarse code for binocular disparity, with three disparity-selective mechanisms. The range of disparities coded varies with spatial scale such that at low spatial frequencies there is a wider range covered than at high spatial frequencies. This embodies the size–disparity correlation. Further, disparity discrimination is modelled by assuming that observers base their estimates of disparity on the ratio of adjacent mechanisms in the code. This works well. The predictions of the code for the task of figure 3b is shown in figure 7 in solid lines. (b) The code from (a) appears 'population like' (many mechanisms tuned to disparity) when tested with a broadband pattern that stimulates all spatial scales of analysis.

for the near and far pools. We had three reasons for so doing. First, the Gaussian derivative does not differ greatly from the other modellers' chosen functions over a range of disparities. Second, we had found empirically that Gaussian derivatives fit our depth-contrast-threshold data well (see Smallman and MacLeod 1994a, figure 6). Third, our choice of functions offers an attractive computational rationale for the system to base its code, for crossed binocular disparities say, on the ratio of activity of the near pool to the horopter pool. The ratio is proportional to the dimension that the code is designed to recover, binocular disparity, denoted by d.

The range of disparities covered at each spatial scale decreases with increasing spatial frequency in quantitative accordance with our previous contrast-threshold data (Smallman and MacLeod 1994a). Specifically, the decreasing range of disparities, $d_0(f)$, covered with increasing spatial frequency scaled the responses of, for example, the near pool at a given spatial frequency, f, such that the response of the pool as a function of disparity was given by

$$R_{Nf}(d) = \frac{d}{d_0(f)} \exp\{-[d/d_0(f)]^2\} + \varepsilon, \tag{1}$$

for the near pool, and by

$$R_{Hf}(d) = \exp\{-[d/d_0(f)]^2\} + \varepsilon, \tag{2}$$

for the horopter (or fixation) pool. ε is a random-noise value having standard deviation σ. The relationship between $d_0(f)$, and f was taken from that which fitted earlier contrast-threshold data (Smallman and MacLeod 1994a). For HSS this was

$$d_0(f)/\text{deg} = \frac{0.64}{(f/\text{cycles deg}^{-1})^{0.67}}. \tag{3}$$

DJP was not an observer in our previous study so we assigned him the relationship of equation (3) too. Relationships of the kind in equation (3) receive recent physiological support from Ohzawa et al (1997) who found a (slightly shallower) slope of -0.39 between disparity range and preferred centre spatial frequency in a large sample of cat complex cells.

A depth estimate, $\hat{d}$, is computable [in units of $d_0(f)$] as the ratio $R_{Nf}(d)/R_{Hf}(d)$. The 75%-correct thresholds in a two-alternative forced-choice experiment occurs when the disparity difference is the standard deviation of each disparity estimate multiplied by 0.907; the value of the coefficient arises because the normal deviate at $p = 0.75$ on a cumulative normal probability integral is 0.6414, and the difference between two independent disparity estimates has a standard deviation equal to the individual standard deviation multiplied by $2^{1/2}$. As will be made clear shortly, the theoretical elaboration of the present model is simplified by discussing the logarithm of the disparity estimates, and by assuming that this ratio, rather than the linear-disparity estimate, is normally distributed. The Weber fraction for disparity is proportional (if not too large) to the threshold difference in the natural logarithm of the disparity, which is the standard deviation multiplied by 0.907 of the logarithm of the two independent disparity estimates for the two fields compared. (For small Weber fractions, these two standard deviations will be approximately equal.) The natural logarithm of the disparity estimate, $\hat{d}$, is equal to $\ln R_N - \ln R_H$, and hence has a standard deviation equal to the square root of the summed variances of $\ln R_N$ and $\ln R_H$. The standard deviation of $\ln(x+\varepsilon)$ is approximately σ/x, where σ is the standard deviation of ε. If this approximation is adopted, the variance in the natural logarithm of a disparity estimate has one component equal to $\sigma^2\exp\{2[d/d_0(f)]^2\}$, the variance in $\ln R_H$, and a second component $\sigma^2[d_0(f)/d]^2\exp\{2[d/d_0(f)]^2\}$, the variance in $\ln R_N$. The Weber fraction for disparity thus becomes

$$\frac{\Delta d}{d} = 0.907\left\{\sigma^2\left[1+\frac{d_0(f)}{d}\right]^2\exp\{2[d/d_0(f)]^2\}\right\}^{1/2}. \tag{4}$$

Simplifying, we get

$$\frac{\Delta d}{d} = 0.907\sigma\exp\{[d/d_0(f)]^2\}\left\{1+\left[\frac{d_0(f)}{d}\right]^2\right\}^{1/2}. \tag{5}$$

The threshold-disparity difference is obtained by multiplying by the base disparity, d, to get

$$\Delta d = 0.907\sigma\exp\{[d/d_0(f)]^2\}\{d^2+[d_0(f)]^2\}^{1/2}. \tag{6}$$

An identical equation, of course, results for far disparities if these are computed from $R_{Ff}(d)/R_{Hf}(d)$.

The predicted performance of the model is shown in figure 7. Here, thresholds are plotted against base disparity normalised by $d_0(f)$ so that predictions for all data across spatial frequency can be visualised. Only one parameter controlled the model fit of equation (6) to the data, the noise parameter σ, which controlled how rapidly stereothresholds rose relative to the strength of putative neural signals in the pools. σ translates the prediction of the model vertically on the log–log axes of figure 7. The depicted values of σ are 0.018 for observer DJP in figure 7a and 0.03 for HSS in figure 7b. The model does a good job of capturing the acceleration of thresholds away from the fixation plane on log–log axes. The model is interesting because it suggests

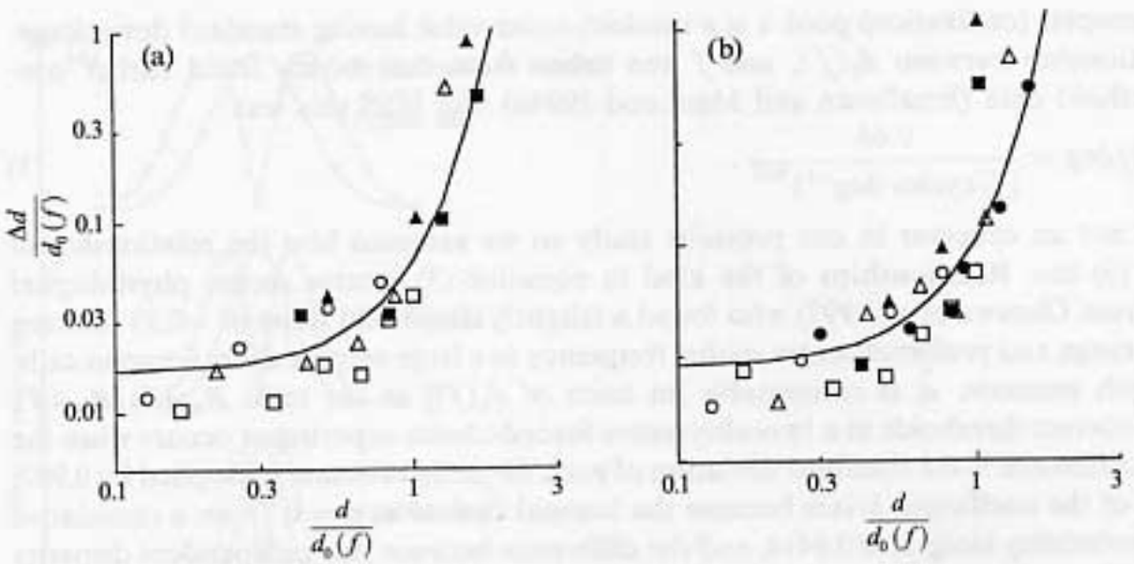

Figure 7. Stereothresholds (y-axis) from figure 4 replotted against standing disparity (x-axis) in units of $d_0(f)$, the range of disparities encoded at each spatial frequency, f, in the model of figure 6. Data are shown for the two observers, DJP in (a) and HSS in (b). The key for the different spatial frequencies is the same as that in figure 4. The solid curve depicts the predicted discrimination of the simple coarse-code-ratio model of equation (6) described in the text. It captures well the acceleration of thresholds away from the plane of fixation.

that a population code for disparity is unnecessary to capture the depth-increment-detection function and it sheds some light on the representation of binocular disparity. It is analogous to one we recently put forward for spatial-frequency discrimination near the visual-resolution limit (Smallman et al 1996).

3.2 *Depth-increment-detection function with compound-frequency targets*

In figure 8 are shown the results of the experiment in which stereothresholds were measured for compound-frequency targets made of noise at 2 and 8 cycles deg^{-1}, along with the stereothresholds for the two component frequencies when determined alone replotted from figure 4. Error bars are 95% confidence intervals.

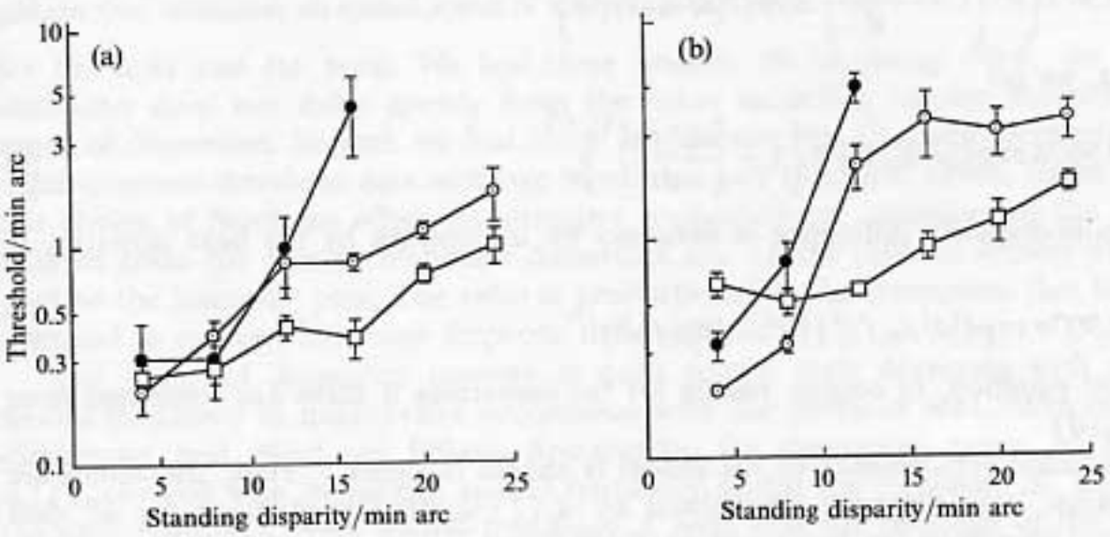

Figure 8. Results of testing the coarse-to-fine matching-range-shifting hypothesis. Stereothresholds away from the plane of fixation are shown for targets defined by low spatial frequency (□ 2 cycles deg^{-1}), high spatial frequency (■ 8 cycles deg^{-1}), and a compound of the two (⊙ 2 + 8 cycles deg^{-1}). Data are shown in the same format as figure 4, with the x-axis representing standing disparity and the y-axis stereothreshold. Data are shown for two observers, DJP in (a) and HSS in (b). Error bars depict 95% confidence intervals. Stereothresholds are worse at large disparities for the compound-frequency targets than for the low-frequency-only targets suggesting that coarse scales cannot shift the matching range of fine scales in the absence of vergence eye movements.

Recall that on the hypothesis of coarse-to-fine range shifting of figure 2 the predicted stereothresholds for the compound-spatial-frequency target should be independent of standing disparity, and at least as good as stereothresholds for the fine frequency component measured alone near the fixation point. This hypothesis is clearly not supported in our data. For small standing disparities (4 and 8 min arc) there is a hint of an improvement in stereothresholds over that of the 8 or 2 cycles deg^{-1} component alone. This improvement is statistically significant for HSS (at 4 min arc, $t_8 = 4.867$, $p < 0.005$; and at 8 min arc, $t_4 = 4.673$, $p < 0.005$) but not for DJP (at 4 min arc, $t_4 = 1.578$, $p < 0.10$; at 8 min arc, $t_3 = 0.979$, $p < 0.20$). At larger standing disparities (>10 arc min) a very surprising result is obtained. Completely counter to the hypothesised improvement, stereothresholds for the compound target are actually *worse* than that of the low-frequency component determined alone. In further control experiments, we halved the contrast of the component targets to see if the better stereothresholds at small and poorer thresholds at large standing disparities could be due to a difference in this contrast between the conditions depicted in figure 8 (Legge and Gu 1989). This led to no measurable change in stereothresholds. They were neither worse, as some have found (Rohaly and Wilson 1993b), nor better, as reported in an earlier study (Richards and Foley 1974). We thus conclude that the different contrast of the components versus that in the compound target in figure 8 cannot account for the effects reported here. We found the same pattern of results of figure 8 when we used targets of 3, 11, and 3 + 11 cycles deg^{-1}.

4 Discussion

The main results of this study can be summarised as follows. First, stereothresholds for spatially filtered RDSs of high centre spatial frequency degrade faster with distance in depth from the fixation plane than do those of low-spatial-frequency RDSs. This demonstrates a size–disparity correlation in the underlying disparity-detection mechanisms, with detectors fed by fine spatial scales covering a narrow range of disparities and those fed by coarse spatial scales covering a broader range of disparities (Smallman and MacLeod 1994a). Second, although Lehky and Sejnowski (1990) have argued that a population code of disparity detectors is necessary to model the depth-increment-detection function, we show that the results could be accounted for more simply by a coarse code of disparity detectors at each spatial scale, whose tuning functions fit previously presented contrast-threshold data (Smallman and MacLeod 1994a). Third, we find no evidence that the system employs coarse-to-fine matching-range shifting to bring to bear fine-scale stereo processing away from the fixation plane (see also Rohaly and Wilson 1993a). This suggests that certain computer models (Nishihara 1984; Quam 1987) need further development if they are to account for human vision, where stereothresholds are actually worse for compound targets (high + low frequency than those of the low component alone.

Others having examined the depth-increment-threshold function with spatially filtered patterns (Mayhew and Frisby 1979; Badcock and Schor 1985; Siderov and Harwerth 1993a, 1995) and not found evidence for the size–disparity correlation. Why not? Mayhew and Frisby (1979) conducted a study similar to ours. Using filtered FDS stimuli, they measured duration thresholds for convergent-disparity discrimination away from the fixation plane. They claimed that a disparity difference of 2.6 min arc was discriminable 10.4 min arc away from the fixation plane with noise at 10.8 cycles deg^{-1} in 140–390 ms, which is in direct opposition to what our model predicts (see figure 6a). The most comparable thresholds in our conditions were twice these (see figure 4). It seems likely to us that the reason for Mayhew and Frisby's good performance away from the fixation plane was due to subjects fixating not on the fixation plane offered by experimental conditions but rather much nearer the stimuli (ie by adopting sustained

convergent fixation disparities). This strategy would have turned the large standing disparities in the stimuli into little ones; ones small enough for the disparity code at this spatial frequency to handle. To prevent fixation disparities, instead of using a fixation spot, as did Mayhew and Frisby, we only presented stimuli when nonius lines were aligned. We did this because a fixation spot will have appeared fused with a fixation disparity. Mayhew and Frisby note in their paper (page 66) that "subjects required some practice to settle to their task" (time that could have been used to learn to adopt a fixation disparity). Ours did not.

Badcock and Schor (1985) also found a different pattern of results from ours using local DOG targets in their determination of the depth-increment-threshold function. They found that thresholds for fine scale DOGs (of 9.6 cycles deg^{-1}) flattened for standing disparities >20 min arc and that these thresholds were better than for low-frequency DOGs. However, both McKee et al (1990) and Siderov and Harwerth (1993a) have pointed out that there were cues other than binocular disparity available to Badcock and Schor's subjects and these could have affected their judgments. Both monocular vernier cues and dichoptic width cues were available. When these cues were randomised, no flattening of stereothresholds above 20 min arc was found (Siderov and Harwerth 1993a). In our experiments which involved RDSs, none of these cues were available to subjects.

The most problematic data to explain are those of Siderov and Harwerth (1993a, 1995), carefully measured by using DOG stimuli. These data show little evidence for the size–disparity correlation. One possible reconciliation of our study and theirs is suggested by Harris and McKee's (1995) recent study. They showed that subjects can extract a rough estimate of depth from monocular targets far from fixation (perhaps from the vergence system). It is possible, therefore, that Siderov and Harwerth's subjects could have used this signal at large standing disparities to mediate a poor, but measurable, stereothreshold for fine-scale targets.

We were able to fit a coarse code for disparity to our discrimination data. Of course, this in no way *proves* that there is a coarse code for disparity. But we find it interesting that we do not need to invoke any more than a coarse code to capture the data. In our model, observers base their discrimination of convergent disparity on an estimate they recover from the ratio of activity of the near pool to the horopter pool. This is very different from Lehky and Sejnowski (1990) who, in their line element model, argue that a population code is necessary to capture the depth-increment-detection function. Lehky and Sejnowski rejected all coarse-code models based on simulations done with one particular such model that had a very narrow tuning in the middle pool (see their figure 6). Their population-code model was set up to account for the Badcock and Schor (1985) data discussed above, and so was not likely to reflect entirely stereoscopic mechanisms.

It was noted in section 1 that some physiological data do not seem to support a coarse code for disparity (eg LeVay and Voigt 1988; Poggio et al 1988). But these data do not rule out our code. This is because cells from this data set came from a range of different retinal loci and probably had a range of different RF sizes. When lumped together such data will appear to support a population code. Only recently has detailed RF mapping been done in conjunction with measurements of disparity selectivity (eg Ohzawa et al 1990), although even this method presently cannot return exact RF positional offsets. Our model makes the physiological prediction that all cells at a given RF locus of a given RF size should cluster into three pools. Moreover, our results leave open the question whether the implied diversity in RF size is found at each retinal location or is due to the variation in cortical magnification with retinal eccentricity. Other psychophysical evidence also seems initially at odds with our code. For example, both Stevenson et al (1992) and Cormack et al (1993) have found evidence for population-code-like behaviour. However, in both of these studies the stimuli were

spectrally broadband and so would stimulate all cells. When coarse codes across all spatial scales are stimulated at one time then the visual system would then appear population-like (see figure 6b). That is, full stimulation masks the specific three-pool action at each given frequency (see figure 6b). Population-like behaviour is necessary to account for perceptual depth shifts following disparity adaptation of Blakemore and Julesz (1971). But with test and adapting stimuli designed to selectively stimulate particular spatial scales of analysis coarse-like depth-shift behaviour would be expected (Smallman and MacLeod 1994b).

The idea of using the stereoacuity measure to look for spatial-scale interactions in stereopsis is not a new one. To test for independent spatial-frequency channels for stereo, Heckmann and Schor (1989) compared stereoacuity in the fixation plane for compound sinusoidal gratings with that for the component frequencies alone. They found no difference in these two conditions, supporting the theory of independent channels. However, as is clear from figure 2, it is away from the fixation plane that independence might be expected to break down and scale interactions play a beneficial role. But in agreement with the earlier study of Rohaly and Wilson (1993a), who used spatially localised sixth-derivative-of-Gaussian targets, we find no evidence for this beneficial interaction. Stereothresholds for compound targets were not independent of standing disparity. This suggests that the system actually does not employ neural coarse-to-fine matching-range shifting (Nishihara 1984; Quam 1987). It may instead be that the visual system circumvents the limitations of the size–disparity correlation through the use of vergence eye movements as Marr and Poggio (1979) originally suggested. All of our data were taken as stimulus-presentation times insufficient to allow completion of such eye movements, which may be why we observed no shift-like behaviour.

We did observe a worsening of stereothresholds for compound targets (see figure 8). This regressive fine-to-coarse interaction in stereoacuity stands in contrast to the beneficial matching disambiguation that fine scales can impose on coarse (Smallman 1995). Perhaps the inflexible channel combination of the system sometimes sacrifices sensitivity. Though unexpected in direction, this result adds to a growing body of evidence for spatial-scale interactions in binocular stereopsis (Prazdny 1987; Rohaly and Wilson 1993a, 1994; Smallman 1995; Mallot et al 1996) and fusion (Wilson et al 1991; Rohaly and Wilson 1993a).

Unlike the data that suggest independent channel access in the fixation plane (Heckmann and Schor 1989), our data suggest that away from the fixation plane when lacking a signal the fine-scale signals actively degrade the stereoacuity signal of the coarse scales. This degradation could form the basis for the reported decrease in the diplopic limit with the addition of high-spatial-frequency components (Kulikowski 1978; Schor et al 1984). Our argument is reminiscent of that of Cleary and Braddick (1990) for the shrinking of D_{max} in broadband kinematograms. Others have recently begun to challenge this account by putting forward models that suggest that the high frequencies serve only to mask the detection of primitives extracted from low-pass spatially filtered image representations (Morgan 1992; Eagle 1996). It remains to be seen whether the stereo data here can be similarly challenged.

Acknowledgements. This research was supported by NIH grant EY-01711, and by grants from the Royal Society, the Nuffield Foundation, and the University of Durham to Harvey Smallman. Some of the work was first presented at the 1993 Annual Meeting of the Optical Society of America and the 1994 Annual Meeting of the Association for Research in Vision and Ophthalmology (Smallman and MacLeod 1993, 1994b). Thanks to Elisa Sobo, Julie Harris, Ariella Popple, and John Findlay for helpful comments on an earlier draft of this paper.

References

Anderson C H, Essen D C van, 1987 "Shifter circuits: A computational strategy for dynamic aspects of visual processing" *Proceedings of the National Academy of Sciences of the USA* **84** 6297–6301

Badcock D R, Schor C M, 1985 "Depth-increment detection function for individual spatial channels" *Journal of the Optical Society of America A* **2** 1211–1216

Barlow H B, Blakemore C, Pettigrew J D, 1967 "The neural mechanisms of binocular depth discrimination" *Journal of Physiology (London)* **193** 327–342

Békésy G von, 1968 *Sensory Inhibition* (Princeton, NJ: Princeton University Press)

Blakemore C, 1970 "The range and scope of binocular depth discrimination in man" *Journal of Physiology (London)* **211** 599–622

Blakemore C, Julesz B, 1971 "Stereoscopic depth aftereffect produced with monocular cues" *Science* **171** 286–288

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

Cleary R, Braddick O J, 1990 "Masking of low frequency information in short-range apparent motion" *Vision Research* **30** 317–328

Cormack L K, Stevenson S B, Schor C M, 1993 "Disparity-tuned channels of the human visual system" *Visual Neuroscience* **10** 585–596

DeAngelis G C, Ohzawa I, Freeman R D, 1991 "Depth is encoded in the visual cortex by a specialized receptive field structure" *Nature (London)* **352** 156–159

DeAngelis G C, Ohzawa I, Freeman R D, 1993 "Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. 2. Linearity of temporal and spatial summation" *Journal of Neurophysiology* **69** 1118–1135

DeAngelis G C, Ohzawa I, Freeman R D, 1995 "Neural mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity?" *Perception* **24** 3–31

Eagle R A, 1996 "What determines the maximum displacement limit for spatially broadband kinematograms" *Journal of the Optical Society of America A* **13** 408–418

Fleet D J, Wagner H, Heeger D J, 1996 "Neural encoding of binocular disparity: energy models, position shifts and phase shifts" *Vision Research* **36** 1839–1857

Freeman R D, Ohzawa I, 1990 "On the neurophysiological organization of binocular vision" *Vision Research* **30** 1161–1675

Harris J M, McKee S P, 1995 "Monocular location provides a signed depth signal for discriminating the sign of large disparities" *Investigative Ophthalmology and Visual Science, Supplement* **37** 1300

Harris J M, McKee S P, Smallman H S, 1997 "Fine-scale processing of human binocular stereopsis" *Journal of the Optical Society of America A* **14** 1673–1683

Heckmann T, Schor C M, 1989 "Is edge information for stereoacuity spatially channeled?" *Vision Research* **29** 593–607

Helmholtz H L F von, 1863 *Die Lehre von den Tonempfindugen als physiologische Grundlage für die Theorie der Musik* (Braunschweig: Vieweg)

Howard I P, Rogers B J, 1995 *Binocular Vision and Stereopsis* (New York: Oxford University Press)

Jacobsen L D, Gaska J P, Pollen D A, 1993 "Phase, displacement, and hybrid models for disparity coding" *Investigative Ophthalmology and Visual Science, Supplement* **34** 908

Klein S, 1992 "An Excel macro for transformed and weighted averaging" *Behaviour Research Methods, Instruments and Computers* **24** 90–96

Kulikowski J J, 1978 "Limits of single vision in stereopsis depends on contour sharpness" *Nature (London)* **275** 126–127

Legge G E, Gu Y, 1989 "Stereopsis and contrast" *Vision Research* **29** 989–1004

Lehky S R, Pouget A, Sejnowski T J, 1990 "Neural models of binocular depth perception" *Cold Spring Harbor Symposia in Quantitative Biology* **55** 765–777

Lehky S R, Sejnowski T J, 1990 "Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity" *Journal of Neuroscience* **10** 2281–2299

LeVay S, Voigt T, 1988 "Ocular dominance and disparity coding in cat visual cortex" *Visual Neuroscience* **1** 395–414

Levitt H, 1971 "Transformed up–down methods in psychoacoustics" *Journal of the Acoustical Society of America* **4** 467–477

McKee S P, Levi D M, Bowne S F, 1990 "The imprecision of stereopsis" *Vision Research* **30** 1763–1779

Mallot H A, Gillner S, Arndt P A, 1996 "Is correspondence search in human stereo vision a coarse-to-fine process?" *Biological Cybernetics* **74** 95–106

Marr D, 1982 *Vision* (San Francisco, CA: W H Freeman)
Marr D, Poggio T, 1976 "The cooperative computation of stereo disparity" *Science* **194** 283 – 287
Marr D, Poggio T, 1979 "A computational theory of human stereo vision" *Proceedings of the Royal Society of London, Series B* **204** 301 – 328
Maske R, Yamane S, Bishop P O, 1984 "Binocular simple cells for local stereopsis: comparison of receptive field organizations for the two eyes" *Vision Research* **12** 1921 – 1929
Mayhew J E W, Frisby J P, 1979 "Convergent disparity discrimination in narrow-band-filtered random dot stereograms" *Vision Research* **19** 63 – 71
Morgan M J, 1992 "Spatial filtering precedes motion detection" *Nature (London)* **355** 344 – 346
Movshon J A, Thompson I D, Tolhurst D J, 1978 "Spatial summation in the receptive fields of simple cells in the cat's striate cortex" *Journal of Physiology (London)* **283** 53 – 77
Mulligan J B M, MacLeod D I A, 1988 "Reciprocity between luminance and dot density in the perception of brightness" *Vision Research* **28** 503 – 519
Mullikin W H J, Jones J J, Palmer L A, 1984 "Periodic simple cells in cat area 17" *Journal of Neurophysiology* **52** 372 – 387
Nishihara H K, 1984 "Practical real time imaging stereo matcher" *Optical Engineering* **23** 536 – 545
Ogle K N, 1953 "Precision and validity of stereoscopic depth perception from double images" *Journal of the Optical Society of America* **43** 906 – 913
Ohzawa I, DeAngelis G C, Freeman R D, 1990 "Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors" *Science* **249** 1037 – 1041
Ohzawa I, DeAngelis G C, Freeman R D, 1997 "Encoding of binocular disparity by complex cells in the cat's visual cortex" *Journal of Neurophysiology* **77** 2879 – 2909
Ohzawa I, Sclar G, Freeman R D, 1996 "Encoding of binocular disparity by simple cells in the cat's visual cortex" *Journal of Neurophysiology* **75** 1779 – 1805
Pelli D G, Zhang L, 1991 "Accurate control of contrast in microcomputer displays" *Vision Research* **31** 1337 – 1350
Pettigrew J D, Nikara T, Bishop P O, 1968 "Binocular interaction on single units in cat striate cortex: Simultaneous stimulation by single moving slits with receptive fields in correspondence" *Experimental Brain Research* **6** 391 – 410
Poggio G F, Fischer B, 1977 "Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey" *Journal of Neurophysiology* **40** 1392 – 1405
Poggio G F, Gonzales F, Krause F, 1988 "Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity sensitivity" *Journal of Neuroscience* **8** 4531 – 4550
Prazdny K, 1987 "On the coarse-to-fine strategy in stereomatching" *Bulletin of the Psychonomic Society* **25** 92 – 94
Quam L H, 1987 "Hierarchical warp stereo", in *Readings in Computer Vision* Eds M A Fischler, O Firschein (Los Altos, CA: Kauffman) pp 80 – 86
Quick R F, 1974 "A vector-magnitude model of contrast detection" *Kybernetik* **16** 65 – 67
Rashbass C, Westheimer G, 1961 "Disjunctive eye movements" *Journal of Physiology (London)* **159** 149 – 170
Regan D, Beverly K I, 1973 "Some dynamic features of depth perception" *Vision Research* **13** 2369 – 2379
Richards W, 1970 "Stereopsis and stereoblindness" *Experimental Brain Research* **10** 380 – 388
Richards W, 1971 "Anomalous stereoscopic depth perception" *Journal of the Optical Society of America* **61** 410 – 414
Richards W, Foley J M, 1974 "Effect of luminance and contrast on the processing of large disparities" *Journal of the Optical Society of America* **64** 1703 – 1705
Rohaly A M, Wilson H R, 1993a "The nature of coarse-to-fine constraints on binocular fusion" *Journal of the Optical Society of America* **10** 2433 – 2441
Rohaly A M, Wilson H R, 1993b "The role of contrast in depth perception" *Investigative Ophthalmology and Visual Science, Supplement* **34** 1437
Rohaly A M, Wilson H R, 1994 "Disparity averaging across spatial scales" *Vision Research* **34** 1315 – 1325
Schor C M, Wood I, 1983 "Disparity range for local stereopsis as a function of luminance spatial frequency" *Vision Research* **23** 1649 – 1654
Schor C M, Wood I C, Ogawa J, 1984 "Binocular sensory fusion is limited by spatial resolution" *Vision Research* **24** 661 – 665
Siderov J, Harwerth R S, 1993a "Precision of stereoscopic depth perception from double images" *Vision Research* **33** 1553 – 1560
Siderov J, Harwerth R S, 1993b "Effects of the spatial frequency of test and reference stimuli on stereo-thresholds" *Vision Research* **33** 1545 – 1551

Siderov J, Harwerth R S, 1995 "Stereopsis, spatial frequency and retinal eccentricity" *Vision Research* **35** 2329–2337

Smallman H S, 1995 "Fine-to-coarse scale disambiguation in stereopsis" *Vision Research* **35** 1047–1060

Smallman H S, MacLeod D I A, 1993 "Interactions across spatial scales and the size–disparity correlation in stereopsis", in *OSA Annual Meeting Technical Digest, 1993* (Washington, DC: OSA) pp 186ff

Smallman H S, MacLeod D I A, 1994a "Size disparity correlation in stereopsis at contrast threshold" *Journal of the Optical Society of America A* **11** 2169–2183

Smallman H S, MacLeod D I A, 1994b "Paradoxical effects of adapting to large disparities: constraining population code models of disparity" *Investigative Ophthalmology and Visual Science, Supplement* **35** 1917

Smallman H S, MacLeod D I A, He S, Kentridge R W, 1996 "Fine grain of the neural representation of human spatial vision" *Journal of Neuroscience* **16** 1852–1859

Stevenson S B, Cormack L K, Schor C M, Tyler C W, 1992 "Disparity tuning in mechanisms of human stereopsis" *Vision Research* **32** 1685–1694

Westheimer G, McKee S P, 1978 "Stereoscopic acuity for moving retinal images" *Journal of the Optical Society of America* **68** 450–455

Wheatstone C, 1838 "Contributions to the physiology of vision. Part the first, on some remarkable and hitherto unobserved phenomena of binocular vision" *Philosophical Transactions of the Royal Society of London* **148** 371–394

Wilson H R, Blake R, Halpern D L, 1991 "Coarse spatial scales constrain the range of binocular fusion on fine scales" *Journal of the Optical Society of America A* **8** 229–236

Wilson H R, Levi D, Maffei L, Rovamo J, DeValois R L, 1990 "The perception of form: retina to striate cortex" (San Diego, CA: Academic Press) pp 317–347

Wilson H R, Wilkinson F, 1997 "Evolving concepts of spatial channels in vision: from independence to nonlinear interactions" *Perception* **26** 939–960