THE ROYAL
SOCIETY

Size-disparity correlation in human binocular depth perception

Simon J. D. Prince[1*] **and Richard A. Eagle**[2]

[1]*Laboratory of Physiology, University of Oxford, Parks Road, Oxford OX1 3PT, UK*
[2]*Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK*

To use the small horizontal disparities between images projected to the eyes for the recovery of three-dimensional information, our visual system must first identify which feature in one eye's image corresponds with which in the other. The earliest level of disparity processing in primates (V1) contains cells that are spatial-frequency tuned. If such cells have a disparity range that covers only a single period of their mean tuning frequency, there will always be exactly one potential match within this range. Here, this 'size-disparity' hypothesis was tested by measuring the contrast sensitivity of stereopsis as a function of disparity for single bandpass-filtered items. It was found that thresholds were low and relatively constant up to disparities an order of magnitude larger than is predicted by this constraint. Furthermore, peak sensitivity was relatively independent of spatial frequency. A control experiment showed that binocular correlation of the carrier is necessary for this task. In a third experiment, the maximum disparity that supports threshold performance was compared for an isolated bandpass item and bandpass-filtered noise. This limit was found to be five times larger for the isolated stimuli. In summary, these findings show that the initial stage of disparity detection is not limited by the size-disparity constraint. For stimuli with multiple false targets, however, processes subsequent to this stage reduce the disparity range over which the correspondence problem can be solved.

Keywords: human vision; stereopsis; depth; correspondence problem

1. INTRODUCTION

There is now a compelling body of psychophysical evidence to suggest that the early stages of stereo processing occur within independent channels each tuned to a narrow range of spatial frequencies (Julesz & Miller 1975; Yang & Blake 1991; Prince *et al.* 1998). Computational studies have demonstrated that this feature of the visual system can be exploited to help solve the correspondence problem (e.g. Marr & Poggio 1979). Within one of these bandpass channels, false matches are on average separated by one period of the filtered stimulus. Hence, schemes which limit the search for correspondence to half the tuning period in either direction ensure a unique solution to the correspondence problem within that channel. The prevailing model of the binocular cells in the primary visual cortex is that they are spatial-frequency tuned and have receptive fields covering the same position in both eyes, but with differing monocular phases in each input (Ohzawa *et al.* 1996). This model also predicts that disparities in vertically orientated, narrow band stimuli may only be detected at up to half a cycle of the peak spatial frequency and that optimal detection will occur at one-quarter of a cycle (see figure 1).

Surprisingly, there is still no consensus on whether the size-disparity constraint is actually employed in human stereopsis (Mayhew & Frisby 1979; Schor & Wood 1983; Smallman & MacLeod 1994). Smallman & MacLeod (1994) measured the minimum contrast required for disparity discrimination in bandpass-filtered noise stereograms as a function of disparity. They found that both the optimal disparity and the maximum disparity for performing the judgement were linked to the centre frequency of the filter and argued for the presence of a size-disparity constraint. However, for filtered noise the number of false matches in a given stimulus region is proportional to the mean frequency. Hence, their experiment may have confounded the spatial frequency with the complexity of the correspondence problem.

Here, we re-examine the issue of size-disparity correlation by measuring disparity discrimination in bandpass-filtered stereograms comprising either isolated items, where the correspondence problem is minimal and independent of spatial frequency, or noise, where the correspondence problem is significant and dependent on spatial frequency.

2. EXPERIMENT 1: CONTRAST SENSITIVITY FOR DEPTH DISCRIMINATION IN SINGLE GABOR TARGETS

In this experiment, luminance-contrast thresholds were measured for a depth discrimination task as a function of disparity and spatial frequency for an isolated bandpass element (a Gabor patch). Stimuli consisted of Gabor patches of odd phase viewed foveally. The luminance profile of each patch was defined by the following equation:

$$L(x,y) = L_{\text{mean}} + k \times e^{-x^2/2\sigma_x^2} \times e^{-y^2/2\sigma_y^2} \times (\sin[2\pi f x + \theta]).$$

---

[*] Author for correspondence (simon.prince@physiol.ox.ac.uk).

Proc. R. Soc. Lond. B (1999) **266**, 1361–1365
Received 3 March 1999 *Accepted* 18 March 1999

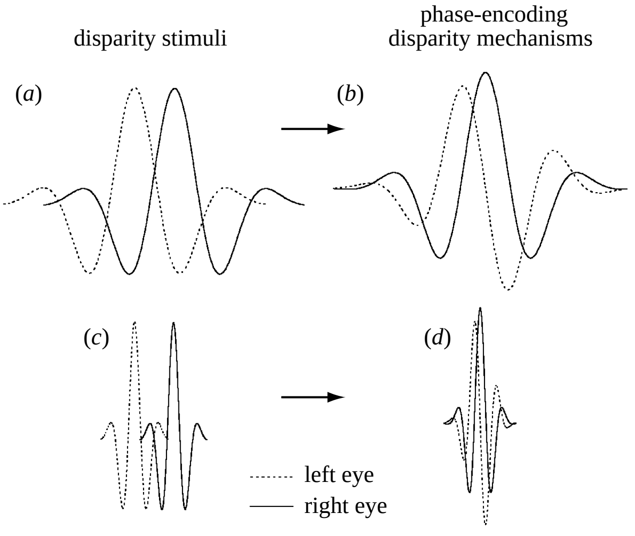

Figure 1. Light from the Gabor stimuli will strike different parts of the left and right retina, relative to the fovea. For a certain depth from the fixation spot, the same disparity in position will be introduced for both low- and high-frequency Gabors, as indicated by the offset in the luminance profiles in (*a*) and (*c*), respectively. (*b*,*d*) One scheme for encoding disparities that uses a size-disparity correlation. In this phase model, the left and right eyes' receptive fields are designed to optimally encode disparities of one-quarter of a cycle of the peak spatial-frequency sensitivity and no larger than half a cycle. The low-frequency stimulus in (*a*) can be detected by the phase-disparity mechanism in (*b*), but the disparity is too large for the high-frequency stimulus in (*c*) to be detected by the mechanism (*d*). This model and others using the size-disparity constraint predict that large disparities can only be encoded by low-frequency detectors.

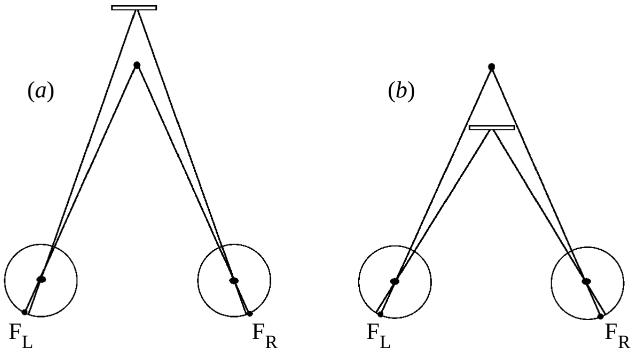

Figure 2. In each trial subjects fixated a central spot binocularly and viewed two Gabor pattern presented in random order, (*a*) one behind and (*b*) one in front of fixation. Subjects were asked to discriminate in which interval the Gabor pattern appeared behind the fixation spot. Light from these stimuli will strike different parts of the left (L) and right (R) retina, relative to the fovea (F).

The size of the horizontal Gaussian envelope was adjusted to provide a constant band width of 1.5 octaves full width at half height. The size of the vertical Gaussian envelope was set to twice the size of the horizontal envelope. The mean luminance of the stimuli was always 20.5 cd m^{-2}. These patterns were presented using a Wheatstone stereoscope configuration driven by a Macintosh 7500 Power PC.

Subjects were asked to fixate the central spot which was present throughout the whole experiment. Between trials, nonius lines were provided on either side of the fixation spot, so that the subjects could check their fixation was correct. Observers were successively presented with two Gabor patches of equal but opposite disparity in each trial (see figure 2). It is important to note that both the carrier and envelope components of the patches were always given the same disparity. The subjects' task was to indicate which interval contained the uncrossed disparity stimulus (i.e. the interval in which the stimulus appeared behind the fixation spot). Each stimulus was displayed for only 150 ms. This ensured that the subjects could not make vergence movements that moderated the magnitude of the stimulus disparity. The interstimulus interval was 500 ms. Two subjects were used, both of whom were experienced psychophysical observers.

In a given condition spatial frequency and disparity were fixed, but contrast was varied. An adaptive Bayesian psychophysical technique was used to make an estimate of the contrast threshold (defined here as the lowest contrast at which 75% correct performance was attained). Each run consisted of 60 trials. There were four runs per condition. The range of stimuli spanned four octaves in spatial frequency and the entire range of disparities across which performance was possible.

Because these stimuli pose a minimal correspondence problem while containing only a narrow frequency range, they are ideal for probing the disparity range of initial stereo processing. Models employing a size-disparity constraint predict that performance will be fixed when expressed in terms of cycles of the stimulus centre frequency. In particular, the phase-disparity model predicts that peak performance will occur at a disparity equal to one-quarter of a cycle of the stimulus frequency and that the upper limit of performance will occur at around half a cycle, independent of spatial frequency.

The results are presented in figure 3. The graphs show that the threshold functions do not peak at a quarter cycle of disparity. Indeed, sensitivity is approximately constant over two cycles for all frequencies. Moreover, for the highest frequency, performance is not extinguished at disparities of 16 cycles. This strongly suggests that a size-disparity limit at the encoding stage is not used to constrain the solution to the correspondence problem.

Simmons & Kingdom (1995) measured contrast sensitivity for chromatic and achromatic stimuli and also found contrast thresholds for disparity discrimination in 0.5 c deg^{-1} Gabor patches were low at a disparity of 1.3 cycles (the largest disparity they measured). One apparent discrepency is that thresholds were increased at around 0.75 cycles. This can probably be attributed to the fact that Simmons & Kingdom (1995) used a relatively narrow band (1.1 octave full width at half height) stimulus in which more sinusoidal cycles were visible. Hence, their stimulus contained more potential false matches which would be expected to impair stereo performance at these disparities.

3. EXPERIMENT 2: CONTRAST SENSITIVITY FOR DEPTH DISCRIMINATION OF SINGLE GABOR TARGETS WITH CONFLICTING CARRIER INFORMATION

One possible objection to this conclusion is that it is the disparity of the contrast envelope and not the enclosed

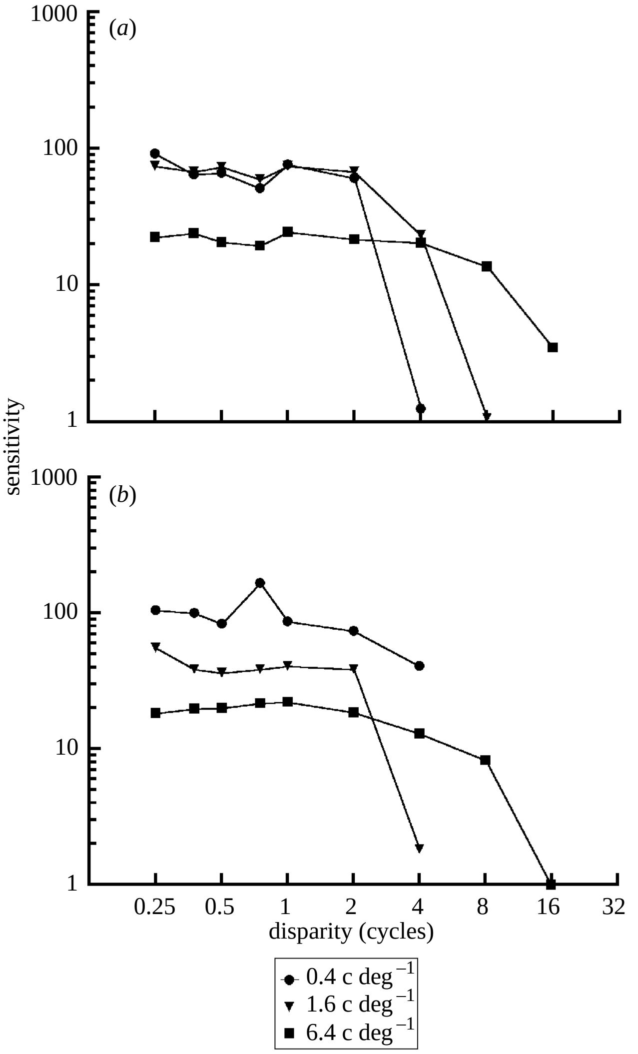

Figure 3. The results for (*a*) subject S.J.P. and (*b*) subject S. A. S. are plotted as sensitivity (the reciprocal of contrast threshold) against stimulus disparity for three spatial frequencies. The sensitivity for depth discrimination in Gabor patches remains approximately constant for a large range of disparities. There is no evidence to suggest that peak performance occurs at a disparity of a quarter cycle. Even the lowest spatial frequencies have a range of almost four cycles and at high spatial frequencies the range extends to 16 cycles.

grating that mediates stereopsis at large disparities. In one sense, this must be true because it is only the envelope information which distinguishes a Gabor patch from a sinusoid, in which depth discrimination performance is necessarily cyclical. However, some authors (e.g. Hess & Wilcox 1994) have proposed the existence of a specialized stereoscopic mechanism for processing contrast envelopes, termed nonlinear or second-order stereopsis. One simple way in which such a specialized mechanism might work is by applying a nonlinear transformation of the stimulus in order to extract the envelope prior to disparity processing.

To examine whether this type of nonlinear mechanism could account for our findings, a second experiment was carried out in which either the spatial frequency or the orientation of the carrier component was different in the two eyes' images, but the contrast envelope was identical.

condition		sensitivity	sensitivity
L	R	S. A. S.	S.J.P.
		67.392	38.46
		12.525	6.0569
		< 1	< 1
		< 1	< 1

Figure 4. The data show contrast sensitivities for the four conditions used in experiment 2, each performed by two subjects (see text for details). In the two conditions, where either carrier spatial frequency or orientation were not matched in the two eyes, performance was abolished. This suggests that the contrast envelope is not being used prior to spatial frequency and orientation filtering.

In two baseline conditions, the carrier was vertically orientated and the spatial frequency was the same in both eyes (either 1.6 c deg^{-1} or 6.4 c deg^{-1}). In a third condition, the carrier was vertically orientated but the frequency presented to each eye differed by two octaves (6.4 c deg^{-1} and 1.6 c deg^{-1}). In the final condition, one eye was presented with a horizontally orientated carrier of 1.6 c deg^{-1} and the other was presented with a vertically orientated carrier of the same frequency. In all of these conditions, the contrast envelope was the same for both eyes' images and always appropriate for the 1.6 c deg^{-1} stimulus. For each of these four conditions the contrast threshold was measured for a large disparity (250 arc min). The task and all other procedural details were identical to those in experiment 1.

If performance at large disparities is mediated entirely by a nonlinear extraction of the contrast envelope then similar thresholds in same- and different-frequency or orientation conditions should be found. However, figure 4 shows that the task was in fact now impossible for our subjects to perform at a 75% correct level at any contrast in the different frequency or orientation conditions. The results here are consistent with the findings of both Wilcox & Hess (1996) and Schor *et al.* (1998) who showed that decorrelating the carrier information significantly reduces the percentage of trials in which the sign of a disparate Gabor can be discriminated. Our results suggest that, if specialized envelope-disparity processing exists, then it occurs subsequent to a common filtering stage. This implies that spatial frequency and orientation filtering precedes disparity extraction and that the extended range of good performance found in experiment 1 requires a binocularly correlated carrier. This suggests that, if a separate second-order system exists, then it is neither independent of the first-order system, nor is it due to a simple early nonlinear transformation of the stimulus intensity information.

This finding is also important in ruling out two other potential accounts of the good performance achieved in experiment 1. First, it shows that performance is not based on monocular positional cues provided by the Gabors, as this cue is equally present in the control stimuli here in which performance is abolished. For the same reason, this result also rules out a strategy based on the dichoptic width cues which would result from systematic misconvergence.

4. EXPERIMENT 3: D_{max} IN LOCALIZED PATCHES AND NOISE

These results appear to contradict the findings of Smallman & MacLeod (1994) who did find a size-disparity correlation in their experiment measuring contrast thresholds for disparity discrimination in bandpass-filtered noise. In the present experiment, D_{max} (the largest disparity at which depth discrimination can be performed) was measured explicitly for both Gabor stimuli and bandpass-filtered noise. The Gabor stimuli had identical properties to those used in experiment 1. Each patch was presented at five times the measured contrast threshold for 75% crossed versus uncrossed discrimination at a 90° phase disparity. Noise stimuli consisted of white noise which had been passed through a one-octave rectangular filter in spatial frequency. The filter was isotropic in orientation to replicate the conditions of Smallman & MacLeod (1994). These were also presented at five times the measured contrast threshold for these stimuli in a fixed rectangular contrast envelope of size $6° \times 6°$. A large field size was presented so that stimuli did not become significantly decorrelated at large disparities.

In each trial, subjects had to indicate which of two presentation intervals contained crossed disparity and which contained uncrossed disparity. The magnitude of the disparity of the patches was varied and the point at which disparity discrimination performance was at 75% correct was estimated. Disparity was varied in steps of 0.05 log units of disparity.

The results are plotted in figure 5. Straight lines in log disparity and log frequency have been fitted. The diagonal line represents a constant 360° phase disparity. These data show that the range of stereoscopic performance is very different for filtered noise patches than for isolated Gabor elements. For the Gabor patch condition, the D_{max} is 0.5–1.5 log units larger than one cycle. This confirms the findings from experiment 1 that disparity encoding extends well beyond that predicted by the phase-disparity model. In the noise condition, D_{max} is around five times smaller, approaching a 360° phase disparity as found by Smallman & MacLeod (1994) at low frequencies. It should be noted that all measures of D_{max} were substantially larger than the 180° limit predicted by a phase-disparity encoding model. The data for Gabor patches show that, for both subjects, D_{max} decreases as spatial frequency increases, with a slope of -0.3 for subject S.J.P. and -0.43 for subject S. A. S. This is considerably smaller in magnitude than the predicted -1 slope of size-disparity models. For filtered noise, D_{max} was measured to be greater than 360° phase and had exponents of -0.47 for subject S.J.P. and -0.5 for subject S. A. S. This is similar to Smallman & MacLeod's (1994) reported value of -0.67.

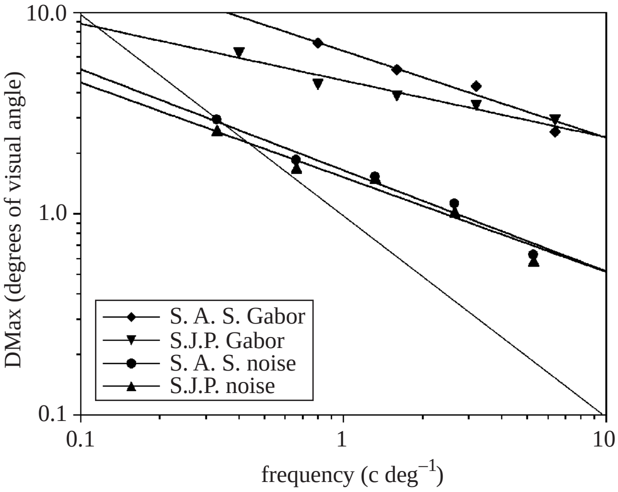

Figure 5. The maximum disparity at which disparity discrimination can be performed (D_{max}) as a function of spatial frequency expressed in degrees of visual angle for two subjects. The diagonal line represents a constant 360° phase disparity as found by Smallman & MacLeod (1994). D_{max} is considerably larger for Gabor patches than for noise stimuli. Moreover, the slope relating D_{max} to spatial frequency is considerably shallower than that predicted by a strict size-disparity correlation.

5. DISCUSSION

Two earlier psychophysical studies have also addressed the question of size-disparity correlation. Schor & Wood (1983) employed 100% contrast difference-of-Gaussian stimuli to measure the maximum disparity at which depth is perceived as a function of frequency and also found a larger disparity range than predicted by size-disparity correlation. However, these results are inconclusive as the use of such high-contrast stimuli will activate a wide range of spatial-frequency tuned detectors. Smallman & Macleod (1994) also noted that this experiment was performed with a long presentation time which potentially allows observers to use vergence movements to bring stimuli into range.

Smallman & MacLeod (1994) measured contrast thresholds for depth discrimination using bandpass-filtered noise patterns in a fixed window. Peak performance was observed at slightly more than one-quarter cycle of the stimulus frequency and cessation of performance at one cycle—consistent with the size-disparity constraint. Smallman & MacLeod (1994) presented an ideal observer model of disparity discrimination which does not predict such a size-disparity correlation. Hence, they concluded that their data must reflect a property of the visual system, rather than of the stimulus itself. One possible explanation for their data is that this limit reflects the small range of disparities encoded at the initial stage of detection. However, another account based on the complexity of the correspondence problem cannot be ruled out as this covaried with the spatial-frequency content of their stimuli.

In fact, our results suggest that the mechanisms at the initial encoding stage span a much larger disparity range than has previously been thought. However, such a scheme necessitates that a second stage of disparity processing is then required to sort through the multiple matches made at this first stage and, thus, solve the correspondence problem. One possible explanation for the reduced $D_{\max}$ for noise stimuli, where the correspondence problem is great, is that this later stage of processing involves a preference for small disparities over large. In the domain of motion detection, $D_{\max}$ has also been shown to vary with the spacing of false targets for a band-pass-filtered stimulus and a matching model favouring small displacements has been found to provide a good quantitative account of the data (Eagle & Rogers 1996).

Our results place heavy constraints on models of disparity encoding. Recent computational models incorporating the size-disparity relationship have fallen into one of two types, often referred to as phase- or position-based models (see Fleet *et al.* 1996). In the former, disparity is encoded by a pair of cells at the same location in each eye and tuned to the same spatial frequency but with a receptive field structure that is shifted by one-quarter-cycle phase. Alternatively, a pair of detectors tuned to the same frequency and phase, but spatially offset by one-quarter cycle, can encode the same disparity information. Physiological evidence in area 17 of the cat has provided evidence for the former model (Ohzawa *et al.* 1996). However, such a mechanism cannot account for our results as the disparities would fall outside the receptive range of these cells. While our results also rule out a position-based model where the disparity tuning is one-quarter cycle, versions of that model that allow larger shifts could account for the data. Any candidate model would have to account for the wide range of disparity sensitivity.

Recent physiological findings using anti-correlated random-dot stereograms (Cumming & Parker, 1997) have shown that V1 cells give a strong response even when perceptually the direction of disparity is undetectable. Our results complement these, showing that there are stimuli for which the sign of disparity is perceptually discernible even though the disparity is too large to be coded by known V1 cells. Both studies demonstrate that the measured properties of V1 cells cannot account for psychophysical observations of stereopsis. Furthermore, because these results preclude the use of the size-disparity relationship as recent models envisage, the correspondence problem would seem to require at least two stages. First, a detection stage in which a wide range of disparities at each spatial frequency are encoded. Second, a stage in which the many false matches made at this early stage are eliminated and the correct ones enhanced.

This work was supported by the Christopher Welch Fund and the Royal Society.

REFERENCES

Cumming, B. G. & Parker, A. 1997 Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* **389**, 280–282.

Eagle, R. A. & Rogers, B. J. 1996 Motion detection is limited by element density not spatial frequency. *Vision Res.* **36**, 545–558.

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

Hess, R. F. & Wilcox, L. M. 1994 Linear and non-linear filtering in stereopsis. *Vision Res.* **34**, 2431–2438.

Julesz, B. & Miller, J. E. 1975 Independent spatial frequency tuned channels in binocular fusion and rivalry. *Perception* **4**, 125–143.

Mayhew, J. E. W. & Frisby, J. P. 1979 Convergent disparity discriminations in narrow-band-filtered random-dot stereograms. *Vision Res.* **19**, 63–71.

Marr, D. & Poggio, T. 1979 A computational theory of human stereo vision. *Proc. R. Soc. Lond.* B **204**, 301–328.

Ohzawa, I., Sclar, G. & Freeman, R. D. 1996 Encoding of binocular disparity by simple cells in the cat's visual cortex. *J. Neurophysiol.* **77**, 1779–1805.

Prince, S. J. D., Eagle, R. A. & Rogers, B. J. 1998 Contrast masking reveals spatial frequency channels in stereopsis. *Perception* **27**, 1345–1357.

Schor, C. M. & Wood, I. 1983 Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Res.* **23**, 1649–1654.

Schor, C. M., Edwards, M. & Pope, D. R. 1998 Spatial-frequency and contrast tuning of the transient stereopsis system. *Vision Res.* **38**, 3057–3068.

Simmons, D. R. & Kingdom, F. A. A. 1995 Differences between stereopsis with isoluminant and isochromatic stimuli. *J. Opt. Soc. Am.* A **12**, 2094–2104.

Smallman, H. S. & MacLeod, D. I. A. 1994 Size-disparity correlation in stereopsis at contrast threshold. *J. Opt. Soc. Am.* A **11**, 2169–2183.

Wilcox, L. M. & Hess, R. F. 1996 Is the site of non-linear filtering in stereopsis before or after binocular combination. *Vision Res.* **36**, 391–399.

Yang, Y. & Blake, R. 1991 Spatial frequency tuning of human stereopsis. *Vision Res.* **31**, 1177–1189.