Size–disparity correlation in stereopsis at contrast threshold

Harvey S. Smallman and Donald I. A. MacLeod

Department of Psychology, University of California, San Diego, La Jolla, California 92093-0109

Received April 1, 1993; revised manuscript received March 3, 1994; accepted March 8, 1994

Contrast thresholds for 75% correct depth identification in narrow-band filtered random dot stereograms were determined for different center spatial frequencies and binocular disparities. Rigorous control over vergence was maintained during testing, and a forced-choice procedure was used. The resulting contrast sensitivity function for stereopsis revealed sensitivity over a greater range of disparities at low than at high spatial frequencies. Sensitivity peaked for large disparities at low spatial frequencies and for small disparities at high spatial frequencies. When disparities were converted to effective binocular phase differences, the variation of contrast sensitivity with phase followed a consistent pattern across spatial frequencies, with peak sensitivity occurring mainly for binocular phases of between 90° and 180°. These results have implications for the extent of spatial integration at the input to the disparity sensing mechanism. A model postulating a spread of positional disparities independent of the spatial frequency selectivity of disparity-sensitive units cannot account for the results. But the size–disparity correlation strongly evident in our data is predicted by certain models of stereopsis, such as phase disparity encoding. An ideal observer analysis is developed that demonstrates that our results were not forced by the nature of the stimulus employed; rather, the quantum efficiency for stereopsis at contrast threshold follows the size–disparity correlation.

1. INTRODUCTION

A continuing issue concerning human stereopsis is the nature of the monocular preprocessing that precedes the extraction of binocular disparities. For the correspondence problem to be minimized, it is clearly desirable that the visual system restrict its matching to those elements in the two visual half-images that are as similar as possible. Over the past 25 years much evidence has accumulated that the early visual signal is processed by a set of channels, or arrays of neurons, each sensitive to a restricted range of spatial frequencies and orientations.1–3 The suggestion that a limited degree of selectivity for spatial configuration is operative in stereo matching comes from two sources. First, it has been shown that overlapping spatial frequency content between the two eyes’ views is necessary to support stereopsis.4,5 That this information may be represented in multiple channels was initially suggested by a second piece of evidence. The demonstrations by Julesz and Miller,6 who used random dot stereograms filtered to exclude all but a narrow band of spatial frequency components, showed that stereopsis was not impaired by the introduction of monocular band-limited masking noise if the noise energy was 2 octaves or more from that carrying the stereo disparities. A recent extension of the Julesz–Miller study systematically manipulating noise amplitude and frequency in a forced-choice masking paradigm study has confirmed and extended these observations.7 Finally, the fact that the disparity range yielding fusion for high spatial frequency components of an image can be influenced by lower spatial frequency components only if the latter are coarser by 2 octaves or less also implies the operation of spatial frequency tuned mechanisms in stereo matching.8

If spatially tuned mechanisms feed into stereopsis, then this is of interest because of the potential for solving the stereo correspondence problem—the problem of which features to match between the two eyes’ views—with them (see Ref. 9 for a review). Marr and Poggio10 suggested that, if the spatial range over which candidate matches could be chosen in a spatially filtered representation were limited, the otherwise severe correspondence problem could be greatly reduced, because matching elements are sparsely distributed in a coarsely filtered image representation. A physiologically inspired model of disparity extraction, sharing several features of the Marr–Poggio stereo theory, has been proposed by Ohzawa et al.11 This model is motivated by the observation that the fine structure of the receptive fields of cat simple cells can differ between the two eyes: in particular, the spatial phase can vary.12 Ohzawa et al. suggested that the disparity-selective responses of cat complex cells could be modeled if one assumed that these cells were innervated by simple cells possessing a receptive field phase mismatch of 90° between the two eyes. In their formulation the complex cells are disparity energy detectors; hence this model is analogous to the Adelson–Bergen13 energy model for motion. Another disparity energy model, incorporating many features of the formulation of Ohzawa et al.,14 has been proposed by Bovne et al.,15 this time motivated by psychophysical considerations.

A critical prediction of these models is that the range of disparities processed will covary with the spatial frequencies defining them: the so-called size–disparity correlation.15 This is because a filter passing only high spatial frequencies (with a correspondingly increased absolute bandwidth) will have a spatially restricted weighting function or receptive field and hence will register only small disparities, whereas a low frequency filter is...
spatially extended and can process a greater range of binocular disparities. Evidence for such a correlation was first suggested by the adaptation study of Felton et al.,\textsuperscript{16} who showed that the largest contrast threshold elevations following adaptation to a discrete grating occurred for disparities proportional to the widths of the bars constituting the grating. Kulikowski\textsuperscript{17} later demonstrated that the diplopic limit for a sine wave was larger than that of a square wave of the same fundamental spatial frequency; this suggested that the higher harmonic components had a lower diplopic limit than the fundamental; that is, the higher spatial frequencies processed a narrower range of disparities. Support for this contention was provided by Schor et al.,\textsuperscript{18} who showed that the diplopic limit for difference of Gaussian (DOG) stimuli occurred at a constant 90° phase disparity for spatial frequencies lower than 2.4 cycles/deg, and square bars having the same space constant as the DOG's shrank the diplopic range substantially. A different but perhaps related connection between the processing of luminance and disparity information was suggested in a neglected paper of Pulliam,\textsuperscript{19} who showed that there was a correlation between the tuning for spatial disparity modulations\textsuperscript{20} and the center spatial frequency of the luminance profile whose disparity was modulated. Taken together, these studies suggest that coarse-scale luminance information can be used to support stereopsis for larger disparities over greater spatial extents than can fine-scale information. It should be noted that the size–disparity correlation receives some physiological support. Two studies that measured the disparity selectivity of cat striate cells and the sizes of their monocular receptive fields have found that they correlate well (≈0.8).\textsuperscript{21,22}

However, other evidence has cast doubt on the notion of a size–disparity correlation, certainly as envisaged by Marr and Poggio.\textsuperscript{15,18,23–27} This evidence has been garnered by researchers presenting various spatial frequency–disparity pairings and looking at upper depth limits,\textsuperscript{15} thresholds for diplopa,\textsuperscript{19} contrast thresholds,\textsuperscript{23} stereo thresholds,\textsuperscript{20} and the limits of disparity discrimination\textsuperscript{28,29} to see whether certain pairings yielded impoverished performance compared with others. However, before one accepts the weight of evidence against the notion, one must carefully establish that the disparities that were claimed to be present in the stimuli were in fact so presented to subjects, because sustained fixation disparities or vergence eye movements may serve to convert the nominal stimulus disparities into unintended ones. For example, Mayhew and Frisby\textsuperscript{28} claimed that disparity discrimination was possible between two large convergent disparities, larger than the Marr–Poggio theory could permit, when the stimuli were random dot stereograms filtered to contain only high spatial frequencies.\textsuperscript{26} Although stimuli were presented with durations sufficiently short to prevent eye movements, no nonius lines were present to check the state of vergence; hence fixation disparities could have reduced effective binocular disparities to a size small enough for a mechanism tuned to high spatial frequencies and small disparities to discriminate reliably the small relative disparity between the two disparity pedestals.\textsuperscript{28} Other studies have used stimulus durations sufficiently long to permit vergence eye movements to modulate disparities in the same unwanted fashion as sustained fixation disparities.\textsuperscript{23,24} However, other research, particularly that of Schor and colleagues,\textsuperscript{15,18,25} certainly cannot be faulted in this regard, yet a size–disparity correlation is evident only below approximately 2.4 cycles/deg (cpd). Discontinuities of this kind have led others to conclude that the lowest mechanism processing stereo may be centered at this frequency.\textsuperscript{29} It is conceivable, however, that the use of 100% contrast, relatively broadband DOG's may have allowed stereo judgments to be mediated by contrast sensitive mechanisms possessing a lower preferred center spatial frequency and hence, possibly, a different disparity selectivity. To guard against off-frequency viewing, our strategy in this paper is to attempt to determine the disparity dependence of the spatial mechanisms processing stereopsis by presenting our stimuli at contrast threshold.

The effect of contrast on the processing of stereoscopic information, particularly stereocuity, has been investigated by Legge and Gu\textsuperscript{30} and Halpern and Blake.\textsuperscript{31} These studies revealed that stereocuity varied as a power function of contrast for low spatial frequencies. But we may also ask what stimulus contrast is necessary for support of a criterion level of stereo performance. If we measure stereo contrast thresholds as a function of the spatial frequency content of the stereograms, then we have the contrast sensitivity function (CSF) for stereopsis. An early attempt to measure the stereo CSF was made by Frisby and Mayhew,\textsuperscript{28} who used band-pass-filtered random dot stereograms. Their study essentially extended to stereo the methodology employed in the pattern detection domain by asking subjects to increase the contrast slowly in a stereo pair until depth was just perceived. With this method, Frisby and Mayhew obtained a disparity-invariant CSF that had its peak sensitivity shifted to slightly lower spatial frequencies than the equivalent functions for detection. At first glance this study also seems to argue against the size–disparity correlation, for that asserts that sensitivity should be poor or unattainable for high disparity–high spatial frequency combinations. However, because of the use of the method of adjustment, this study shares the problem of the contamination of vergence eye movements mentioned above. Unless controlled in a study of this kind, vergence eye movements can serve to modulate the disparities in a stereogram to ones that could be optimal for a mechanism that covaried in its spatial and disparity-processing capacities. What is needed is the determination of the CSF for stereopsis under conditions controlling vergence eye movements within a more rigorous two-alternative–forced-choice paradigm. A design that employs the presentation of both crossed and uncrossed disparities is also preferable, because it will make anticipatory eye movements essentially useless, unlike the Mayhew–Frisby\textsuperscript{28} design. Mansfield and Simmons\textsuperscript{34} have also measured stereo contrast thresholds for bandpass-filtered stereograms possessing different disparities with a view to explaining the contrast gap between detection and depth identification, but again with unrestricted eye movements. The present study seeks to remedy these deficiencies.
2. Method

Apparatus and Stimuli
All stimuli were presented on a Taxan Ultra Vision 1000 RGB monitor driven by a Macintosh IIcx microcomputer. We used a linearizing color lookup table to produce the desired intensity profiles and combined the 8-bit red, green, and blue outputs to yield 12-bit digital-to-analog converter precision on the green monitor input and hence permit very fine contrast modulations, using the video attenuator hardware and software of Pelli and Zhang.32 The monitor was viewed at 57 cm by the subjects through a simple mirror stereoscope that permitted each eye to view independently a different region of the screen.

We made the filtered textures that constituted the stimuli by isotropically digitally filtering a 256 × 64 patch of binary random texture with a radially symmetrical filter profile that passed unattenuated those frequencies that were ±0.5 cpd of center frequency and blocked higher and lower frequencies. We accomplished this filtering by setting the amplitude coefficients to zero in the two-dimensional (2D) Fourier transform for frequency components outside the annular passband and then taking the inverse Fourier transform. The amplitude density function of the resulting patterns was approximately Gaussian. This filtering scheme was chosen to permit a direct comparison with the contrast thresholds for stereopsis of Frisby and Mayhew,33 but it entailed employing a constant linear filter bandwidth rather than the more usual constant log bandwidth. The implications of this filter choice are discussed below.

Narrow-band-filtered 2D noise patterns possess certain advantages for stereo experiments: their luminance cross sections are locally approximately periodic. With one-dimensional (1D) patterns this approximate periodicity makes locally derived estimates of horizontal disparity somewhat ambiguous, as in the wallpaper illusion. The 2D patterns used here have the advantage that they practically eliminate this ambiguity, since the effective horizontal spatial frequency varies vertically as well as horizontally. But they also introduce a complication: the horizontal disparity associated with Fourier components of different orientations is always the same as the stimulus horizontal disparity, but the phase disparity measured in the direction of modulation varies as \( \cos(\theta) \), where \( \theta \) is the angle between the direction of modulation and the horizontal. However, it seems that the former of these disparity measures is more relevant for disparity-selective neurons.12

On each trial two random \( 1^\circ \times 1^\circ \) regions of this filtered texture were displayed to produce a stereogram of two patches in depth, one on each side of a fixation cross. Each had a crossed or uncrossed disparity relative to fixation of an amount always equal to and opposite to that of the patch on the other side of the fixation cross (Fig. 1). We introduced the disparities by displaying the shifted texture into the \( 1^\circ \times 1^\circ \) regions whose borders were still at zero disparity, to avoid possible monocular cues to the sign of the disparity of the panels. Nonius lines, of length 75 min, were continuously visible throughout the experiment along with a \( 1.5^\circ \times 3.1^\circ \) outline frame in the fixation plane (see Fig. 1). The mean luminance of the textured stimuli was 50 cd/m², and the background was kept at 5 cd/m². The fixation, nonius, and bordering lines were all 95 cd/m².

Procedure
A trial was commenced by the subject with a button press when she or he perceived the nonius lines to be collinear.33 Half a second later the stimuli were displayed for 150 ms, a duration insufficient to permit the initiation of eye movements.34 On successive trials, either the left or the right patch was randomly selected to have a crossed disparity, and the other had the same uncrossed disparity. When the stimuli had been removed, the subject indicated whether he or she had perceived the left or the right filtered texture patch as lying in front of fixation. The threshold root mean square (rms) contrast for this decision was determined with a two-alternative forced-choice staircase procedure. The filtered texture contrast was modulated on successive trials by a 3-up–1-down staircase35; in this way, trials were efficiently concentrated about the contrast, supporting 83% correct depth identification. Two such staircases were randomly interwoven for each spatial frequency–disparity pairing condition and constituted one block of trials. Forty-five data trials were run for each staircase, with an additional three practice trials before. No feedback was given about the correctness of the subject’s response. Data for the two
The CSF is bandpass in shape, both for detection and for sensitivity for the detection of a filtered patch, possessing different disparities (from 1 to 20 arcmin and for center frequencies ranging from 1 to 15 cpd). Contrast sensitivity was defined conventionally as the reciprocal of threshold rms contrast. The conventional binocular contrast sensitivity data for detection (line-dot curve) are also shown for comparison.

Subjects

Complete data were obtained with one experienced psychophysical observer, LW, who was naive of experimental intent. Further data were taken by the even more experienced observer SPM at three center spatial frequencies. Both have good stereo and normal or corrected visual acuity. Additional data taken from two more observers lead us to trust the generality of the findings.

3. RESULTS AND DISCUSSION

Contrast thresholds were determined for disparities ranging from 1 to 20 arcmin and for center frequencies ranging from 1 to 15 cpd. Contrast sensitivity was defined conventionally as the reciprocal of threshold contrast. The complete stereo CSF (composed of a family of curves for different disparities) for one observer is shown in Fig. 2. The detection curve represents the binocular contrast sensitivity for the detection of a filtered patch, possessing zero disparity, lying to the left or right of fixation in the same location as the patches composing the stereograms. The CSF is bandpass in shape, both for detection and for disparities lower than 11 arcmin. A bandpass shape was also found for one of Frisby and Mayhew’s observers and by Mansfield and Simmons and was suggested by the masking data of Yang and Blake. Peak sensitivity is at 3 cpd for stereo and 5 cpd for simple detection. The lowest threshold disparities of Legge and Gu’s observers was likewise at 3 cpd.

The difference between the stereo thresholds that could be measured and simple detection thresholds is on average 0.25 log unit and varied from 0.1 to 0.53. This is considerably lower than the average gap of 0.44 unit in the Frisby-Mayhew study but of the same order as that reported by Mansfield and Simmons. For one of our observers, SPM, stereo contrast thresholds were actually essentially equal to detection thresholds at 1 cpd for binocular disparities of 20 arcmin (threshold rms contrast of 0.0169 for detection and 0.0171 for stereo, a 0.1-dB increase); measurable stereoacuity, of the order of 1 min, has been reported for D10 stimuli of center frequency 1.2 cpd at only 3 dB above detection threshold. The average correlation between detection sensitivity and the stereo sensitivity was 0.99 for LW’s data; it was 0.96 for the data of Frisby and Mayhew. The correlation between threshold disparity and contrast sensitivity was 0.84 in the Legge-Gu study. A central feature of the data is that the range of disparities over which sensitivity could be measured varies with spatial frequency. At 15 cpd, the highest frequency tested, this range was the most severely restricted and ranged in disparity from 1 to 7 arcmin; with larger disparities than this were tested they resulted in chance performance. At the lowest tested frequency, sensitivity could be measured for all disparities, but it was lowest for 1 arcmin, the lowest disparity, and best for 20 arcmin, the highest disparity tested. Inspection of Fig. 2, particularly of the dashed curves depicting the highest disparities, confirms the trend for a narrowing range of disparities permitting sensitivity measurements as spatial frequency is increased. There is considerable consistency in the dropout of the larger disparities with increasing spatial frequency, with 20-arcmin disparity measurable only up to 3 cpd, 15-arcmin disparity measurable only up to 7 cpd, and so on.

The variation of disparity range with center spatial frequency can be seen more clearly with a three-dimensional (3D) representation of the data, and one is depicted in Fig. 3. This highlights the dramatic shrinkage in the range of disparities, permitting sensitivity measurements with increasing spatial frequency. Thus contrast sensitivity forms a ridge in the 3D space, with a region of pairings of high spatial frequency and disparity revealed on the floor of the cube, where there was no contrast sensitivity. The Frisby-Mayhew data, if replotted in the same fashion, would form a flat surface in the cube.

Another feature of the data, which is only somewhat apparent in the 3D representation of the data, is that the disparity giving peak sensitivity varied with spatial frequency. The stereo CSF of Frisby and Mayhew was disparity invariant, and Mansfield and Simmons reported that the peak sensitivity was for a 3-arcmin disparity across spatial frequencies. In Fig. 2, sensitivity for the lowest disparities of 1 and 3 arcmin is relatively better than for the higher disparities for only the highest spatial frequencies. The converse is also true: the best sensitivity at the lowest spatial frequency of 1 cpd is for
H. S. Smallman and D. I. A. MacLeod

Lent to a binocular phase of 90°. The data from Fig. 2 binocular phases. For example, a binocular disparity of view the data would be to transform the disparities into regularity suggested to us that a more instructive way to the highest disparities of 20 and then 15 arcmin. This pairings of large disparity-high center spatial frequency. sensitivity forms a ridge in the 3D space, with no sensitivity for pairings of large disparity-high center spatial frequency.

the highest disparities of 20 and then 15 arcmin. This regularity suggested to us that a more instructive way to view the data would be to transform the disparities into binocular phases. For example, a binocular disparity of 5 arcmin at a center spatial frequency of 3 cpd is equivalent to a binocular phase of 90°. The data from Fig. 2 are replotted this way in Fig. 4 for each of the spatial frequencies tested.

Figure 4 reveals systematic variations of contrast sensitivity for stereopsis with phase across a range of spatial frequencies. Each curve exhibits a phase that gave optimum performance with sensitivity compromised for lower and higher phases. This figure permits a key prediction of the quadrature model of Ohzawa et al. to be tested: that the peak contrast sensitivity will be found for binocular phases of ~90°; this follows simply from the fact that a 90° disparate stimulus should match up to the receptive fields closely in the two eyes' views and hence require the minimum contrast to trigger the cells' response. Figure 4 offers some support to this prediction, although there is a tendency for sensitivity to peak at phase disparities slightly higher than 90°. This discrepancy has a plausible explanation: if not only vertically tuned disparity-selective cells but also cells that are sensitive to orientations deviating substantially (say, by θ deg) from the vertical are optimally stimulated by a quadrature phase relationship between the relevant Fourier components of the monocular images, the horizontal phase disparity that best stimulates these will be 90°/cos(θ) degrees, increasing from 90° for vertically tuned units to 180° for units with preferred orientations that deviate by 60° from vertical. This explanation remains speculative, although it is known that binocularly driven neurons actually do exhibit a variation of receptive field phase differences between the two eyes with preferred orientation. Finally, there is a suggestion that sensitivity extends to higher phases for the stimuli composed of higher spatial frequencies. Sensitivity could be measured up to one spatial period of the center frequency at 3 cpd but up to nearly two periods at 11 and 13 cpd.

The variations of sensitivity with phase are all the more impressive when they are compared with the data of Frisby and Mayhew,23 which have been replotted as phases in Fig. 5. Note that, in those data, sensitivity is essentially invariant with phase and could be measured up to nearly 2000°, or 5.5 spatial periods of center spatial frequency, which corresponds to a disparity of 22 arcmin at 15 cpd, the highest disparity–spatial frequency combination that those authors tested. The mean variation of sensitivity with phase across frequencies for their two observers was 0.2 log unit and ranged from 0.1 to 0.4. In LW's data, on the other hand, there was a mean variation of 0.8 log unit, and this ranged from 0.2 to 1.4 log units across frequency. This mean value is derived from only those disparity–frequency combinations that would permit sensitivity to be tested and is therefore an underestimate of the real variation in sensitivity. We argue that the phase invariance in Frisby and Mayhew's data may have resulted from the ability of vergence eye movements to reduce or equate all the disparities tested and therefore result in a disparity-invariant CSF for stereopsis.

We discovered empirically that there were consistent variations in contrast sensitivity with disparity at each spatial frequency in our data. This enabled us to characterize the underlying processes further by fitting the same mathematical function, a Gaussian derivative of sensitivity with disparity, to all the data:

$$S(f) = S_0(f) d/f \exp\left\{-\frac{[d/f]^2}{2}\right\}.$$  \hspace{1cm} (1)
Spatial Periods of Center Frequency

<table>
<thead>
<tr>
<th>Binocular Phase (degrees)</th>
<th>Spatial Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 cpd</td>
</tr>
<tr>
<td>180</td>
<td>3.75 cpd</td>
</tr>
<tr>
<td>360</td>
<td>5 cpd</td>
</tr>
<tr>
<td>540</td>
<td>7.5 cpd</td>
</tr>
<tr>
<td>720</td>
<td>10 cpd</td>
</tr>
<tr>
<td>900</td>
<td>12.5 cpd</td>
</tr>
<tr>
<td>1080</td>
<td>15 cpd</td>
</tr>
</tbody>
</table>

Fig. 5. Contrast sensitivity plotted as in Fig. 4, for the data of the two observers of Frisby and Mayhew\textsuperscript{23} for each of the spatial frequencies that they tested. Note the difference in scale on the abscissa between this figure and Fig. 4. Sensitivity is essentially independent of phase. The two panels represent the data from the two observers in the Frisby–Mayhew study.

Spatial Periods of Center Frequency

<table>
<thead>
<tr>
<th>Binocular Phase (degrees)</th>
<th>Spatial Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2 cpd</td>
</tr>
<tr>
<td>180</td>
<td>3.75 cpd</td>
</tr>
<tr>
<td>360</td>
<td>5 cpd</td>
</tr>
<tr>
<td>540</td>
<td>7.5 cpd</td>
</tr>
<tr>
<td>720</td>
<td>10 cpd</td>
</tr>
<tr>
<td>900</td>
<td>12.5 cpd</td>
</tr>
<tr>
<td>1080</td>
<td>15 cpd</td>
</tr>
</tbody>
</table>

Fig. 6. Log contrast sensitivity as a function of disparity for all spatial frequencies superimposed. The fitted curve is a Gaussian derivative of contrast sensitivity as a function of disparity, thus making sensitivity proportional to disparity for small disparities. We accomplished the superposition by scaling each curve by the disparity range parameter \(d_0(f)\) factor to establish horizontal congruence and by adding the sensitivity scaling factor \(S_0(f)\) to line the curves up vertically (see text). The key for spatial frequency is the same as for Fig. 4.

We used two free parameters: a sensitivity scaling factor to make the curves vertically pass through the data points, \(S_0(f)\), and a disparity range factor \(d_0(f)\) (the space constant of the Gaussian) to constrain the curve to capture the data horizontally. Figure 6 shows all these curve fits superimposed and reveals that the chosen function performs well at capturing the main trends in the data. \(S_0(f)\) was largest at 5 cpd and decreased monotonically for spatial frequencies higher than that.

The estimated parameters from the curve fits enabled us to ask further questions of the data. Figure 7 shows a scatter plot of the disparity range parameter \(d_0(f)\) against spatial frequency. The absolute value of the disparity range parameter itself has no clear significance, but its variation with spatial frequency is illustrative of the main trends in the data: it is proportional to the disparity yielding optimal sensitivity for each spatial frequency (by a scaling factor of \(1/\sqrt{2}\)), and this is plotted on the right-hand axis of the figure. This plot quantifies and confirms the trend discussed above for the existence of a systematic variation in the range of disparities processed at each spatial frequency. There is a more than fivefold variation in the range of disparities processed from the lowest to the highest spatial frequency tested. That is, the data strongly support the notion of a size–disparity correlation. The same type of Gaussian derivative curve also fitted SPM's data well for the three spatial frequencies on which she was tested, and these are also plotted in Fig. 7. Both observers show the same tendency for a decreasing range of disparities processed with increasing spatial frequency, although SPM's falloff is steeper. In the same plot are shown linear regressions of disparity range on spatial frequency. The slope of the lines was \(-0.97\) for observer SPM and \(-0.56\) for observer LW. The regressions accounted for 98% and 99% of LW's and SPM's data, respectively. There is no suggestion of a discontinuity with frequency and for disparity range to flatten above 3 cpd as there was in the data of Schor and Wood.\textsuperscript{15} On the contrary, disparity range continues to decline for center spatial frequencies above 3 cpd.

It is known that the processing of stereo can be significantly influenced by the temporal aspects of the stimulus,\textsuperscript{25,28} and Panum's fusional area varies significantly with spatiotemporal disparity modulation\textsuperscript{39} and stimulus duration.\textsuperscript{40} In short, stereopsis seems particularly sluggish, and it is conceivable that the narrow range of disparities found for high frequencies could have been due to the tardiness of high relative to lower frequencies. To ensure that our results generalized beyond the brief...
150-ms stimulus duration employed above, a control experiment was run in which the contrast sensitivity function for stereopsis was redetermined as before but with a 2-s stimulus duration. The results for one observer are shown in Fig. 8(a).

Comparison of Figs. 8(a) and 2 shows that the main conclusions that were made on the basis of a short exposure duration can be generalized to longer durations as well. The trend of a greater range of disparities permitting sensitivity determination at the lower frequencies is present. Also, sensitivity is best at the largest disparities tested at low frequencies and best for small disparities at high frequencies. There are some minor inconsistencies between observer LW's short- and observer HSS's long-duration data. At 5 cpd, sensitivity could be measured up to 20 arcmin for HSS, higher than the 15 arcmin of observer LW. In addition, the disparity yielding optimal contrast threshold was consistently nearer a 90° phase shift in HSS's long-duration data. This is revealed more clearly in Fig. 8(b), which replots the long-duration data of HSS as phases, in exactly the same manner as was done in Fig. 4. Up to 5 cpd the binocular phase giving optimal sensitivity was 90°, with a tendency for the higher frequencies to shift to slightly higher peak phases. Again there was an envelope of sensitivity to phase extending up to nearly two spatial periods of center frequency.

Another necessary control was to ensure that the filtering scheme employed was not responsible for the size–disparity correlation evident in the data. Above, we showed how stimuli were filtered with a constant linear bandwidth in order to make the contrast thresholds directly comparable with those of the study of Frisby and Mayhew rather than with the conventional constant log bandwidth. To ensure that this procedure was not responsible for the narrow range of disparities permitting contrast thresholds to be measured, we reran the experiment at one high center spatial frequency (13 cpd) over the entire range of disparities that would permit sensitivity measurements. HSS was the subject in the experiment, and three bandwidths were employed: 0.1, 0.5, and 1.0 octaves, with a stimulus duration of 1 s.

The results from the control experiment are shown in Fig. 9. As above, sensitivity could be measured only over a narrow range of disparities (in this case up to 11 arcmin), and a tenfold increase in stimulus bandwidth affected contrast sensitivity only marginally. There was an apparent tendency for the disparity range to expand slightly for the broader bandwidths, but this would be expected from the extra low frequency content present in

---

**Fig. 7.** Scatter plot of range of disparities processed as a function of center spatial frequency of the filtered texture for the two observers LW and SPM, with best-fitting straight lines shown. The left-hand axis plots the disparity range factor \( d_\text{r} / f \), defined as the space constant of the Gaussian derivatives used to fit the curves of Fig. 6. The right-hand axis, for the same data, indicates the disparities giving optimal contrast sensitivity at each spatial frequency. There is at least a fivefold reduction in the range of disparities processed from the lowest to the highest center frequency; that is, the data exhibit a strong size–disparity correlation.

---

**Fig. 8.** (a) CSF for stereopsis for observer HSS plotted as in Fig. 2, determined for a display duration of 2 s. Sensitivity (reciprocal of threshold rms contrast) is plotted against center spatial frequency. The variation of sensitivity with disparity is in close accordance with that found for observer LW, shown in Fig. 2. This establishes that the results found above are not specific for a short display duration (used in that case to guard against eye movements). (b) Data from (a) replotted to show sensitivity as a function of binocular phase for each spatial frequency tested; the curves were constructed in exactly the same way as for Fig. 4. The 90° phase prediction for optimal sensitivity from the model of Ohzawa et al. is shown by the arrow. The number of spatial periods of center spatial frequency represented by a certain binocular phase is shown on the top axis.
there are significant variations in the contrast required for depth identification with the disparity and spatial frequency content of the stimuli. A significant variation in the range of processing of disparities at each spatial frequency is clearly revealed (see Fig. 7). Our data suggest that the labeled lines\(^{43}\) that process stereopsis differ in their disparity selectivity with spatial frequency, because at contrast threshold at 1 cpd observers discriminate best between disparities of 20 arcmin but at 15 cpd this number is reduced to 3 arcmin. Some of the previous work that has not revealed a size-disparity correlation has been performed under conditions that may have permitted the modulation of disparities in the stimulus by vergence eye movements\(^{23}\) or by sustained fixation disparities.\(^{26}\) Our data do not show the independence of stereo processing with spatial frequencies higher than 2.4 cpd that is evident in the data of Schor and colleagues,\(^{15,18,55}\) which was obtained with high contrast DOG targets (note that in our data the decline in disparity range across all frequencies was well fitted by a straight line; see Fig. 7). There are many methodological differences between our study and those of Schor, and the reasons for the apparent discrepancies between the sets of data are the subject of ongoing research. Some possible reasons are outlined briefly here. First, one potential reason for finding a large disparity range with 100% contrast high frequency DOG stimuli is that these stimuli may give rise to a contrast signal in sensitive low frequency channels—channels, if we are correct, that may process a wider range of disparities, through off-frequency viewing. Our study was conducted at contrast threshold so that we could be more certain of which mechanisms were determining sensitivity. Second, our Julesz-style stimuli contained no local sign information in the conventional sense, and in this regard it is interesting to note that bisection thresholds show a frequency dependence similar to that of the Schor data.\(^{44}\)

The size—disparity correlation evident in our data places constraints on formulations of the early encoding of binocular disparities. Freeman and Ohzawa\(^{45}\) have drawn a dichotomy between two disparity-encoding schemes, one employing receptive field phase and one employing a receptive field positional shift between the two eyes to give a putative cortical neuron disparity selectivity. Figure 10 illustrates this phase-encoding scheme in the right and center columns, and it shows that the phase-encoding scheme must predict a size—disparity correlation. The simplest position-encoding scheme might have a receptive field positional jitter which is independent of the preferred spatial frequency of the units innervating the disparity detectors; in fact, several stereopsis models employ such schemes.\(^{46,46}\) This scheme would predict a disparity-invariant stereo CSF, as was reported by Frisby and Mayhew.\(^{23}\) However, our results show that we must revise this scheme by making the positional shift (or jitter) frequency dependent, as shown in the left and center columns of Fig. 10. Mean values of the shift can be obtained from the right-hand axis of Fig. 7. We cannot at present distinguish between the two encoding schemes that can handle our data shown in Fig. 10; one reason for this, as a close inspection of Fig. 10 reveals, is that the receptive field profiles in these two schemes are actually very similar. Other researchers have recently tried to discriminate between Freeman's two notions of disparity encoding by using a different paradigm.\(^{47}\) It is interesting to note in this regard that a fairly good characterization of the variation of the disparity that yielded optimal sensitivity and spatial frequency for LW's data (from Fig. 7) is a quadrature phase shift between the eyes plus 5 arcmin, across all frequencies. Thus one rapprochement between the phase- and position-encoding alternatives may be a hybrid model of the two, and other researchers also are entertaining this possibility.\(^{48}\)

Another prediction that the phase-encoding model of Ohzawa \textit{et al}.\(^{11}\) makes is that peak contrast sensitivity should be found for a quadrature, or 90°, interocular phase shift of the center spatial frequency defining a stimulus. Inspection of the sensitivity data replotted against phase in Figs. 4 and 8(b) show that this prediction is fairly well upheld (although we would not wish to claim that the data
discriminate against other models that predict, say, a 120° phase shift or even a 180° shift that the model of McKee et al. makes; the latter model, however, would seem to receive little support if it employed any obliquely oriented mechanisms, as that would shift its predicted peak sensitivity to phases higher than 180°; see below). One might argue, though, that the reason that optimal sensitivity was found near quadrature phase is that this was where the stimulus was least ambiguous, and hence an ideal observer making optimal use of the information provided by the stimulus might perhaps have generated similar data. To ensure that the size-disparity correlation that we obtained was not due to the nature of the stimulus employed, we have formulated an ideal observer for disparity discrimination at contrast threshold (which is developed in Appendix A). This reveals that optimal use of the information in the stimulus would not favor a phase disparity of 90° over much larger ones (over 1000°). In addition, the quantum efficiency for disparity discrimination at threshold was calculated from our ideal observer and is plotted for several center spatial frequencies in Fig. 11. Quantum efficiency reflects the ratio of the human performance in a task to that of the ideal observer.

Several important points can be made about Fig. 11. Disparity-specific quantum efficiency (DSQE) is better for small disparities (mean across spatial frequencies of 15% at 1 arcmin) than for larger disparities (e.g., mean at 15 arcmin when measurable of 4.1%). Of direct relevance here, though, is the relationship between center spatial frequency and disparity. It can be seen that the quantum efficiency at the larger disparities is much better for the low than for the high center spatial frequencies. For example, at a disparity of 20 arcmin the efficiency at 1 cpd is still 12.7%, whereas it is 2 log units worse for the next-higher spatial frequency of 3 cpd and could not even be measured for frequencies higher than that. This reinforces the idea of the size-disparity correlation: there do not appear to be mechanisms with a preference for large disparities and high spatial frequencies when one tests at contrast threshold. However, with regard to the model of Ohzawa et al., it is noteworthy that the
The systematic variations in the sensitivity-versus-disparity functions of Fig. 6 further constrain notions of disparity encoding. Our Gaussian derivative sensitivity functions are similar to the DOG's chosen by Lehky and Sejnowski\(^5\)\(^4\) to make up their population code for disparity and similar to those that have been successfully employed in the modeling of recent psychophysical depth adaptation data.\(^6\)\(^5\)\(^6\) These models do not at present consider the spatial integration preceding disparity extraction or the response of these mechanisms with contrast. As suggested above, our data suggest that there may be different populations of disparity detectors at different spatial scales, with preferred disparities extending to larger disparities at coarser spatial scales. Our fitted curves were Gaussian derivatives, but these do not differ greatly from sine Gabor functions (especially for small disparities), and hence our curves provide additional support for the model of Ohzawa et al., which models the discharge of disparity-selective complex cells with a sine Gabor as a function of disparity (see Fig. 3C of Ref. 11).

Finally, how are large binocular phases processed by the stereopsis mechanism? A system exhibiting a size-disparity correlation is particularly insensitive to large binocular phases, say, large disparities presented with fine-scale information. This insensitivity is undesirable, because fine-scale stimuli yield the best stereocuity.\(^2\)\(^7\)\(^5\) One answer to this question is suggested by an inspection of the data of Frisby and Mayhew plotted in Fig. 5. If our interpretation of their results is correct, then perhaps eye movements (given time) can defend the system against large phases by converting them into small phases again. This is very similar to the role posited for vergence in the theory of Marr and Poggio,\(^1\)\(^0\) which was later discredited.\(^5\) A more exciting idea is that coarse-scale matching (which can represent large disparities) is able to cooperate with and perhaps drive fine-scale matching to derive depth estimates that are both unambiguous and precise, in the absence of eye movements; there is good evidence for other suprathreshold interactions across spatial scales in stereopsis.\(^8\)\(^5\)\(^9\) We are currently addressing this question.

**APPENDIX A**

The variation of contrast sensitivity with disparity depends substantially on the nature of the stimulus that is used to measure it. For example, if the stimulus is horizontally periodic, such as a vertical sine wave, then contrast sensitivity must have the same periodicity, whatever the nature of the disparity-sensing mechanism, because of the wallpaper illusion.\(^6\) It is therefore important to consider how our results may be constrained by the characteristics of the stimulus used. This may be done by a comparison of the results with the ones that would be expected for an ideal observer that makes optimal use of the information provided by the stimulus. We will sketch an analysis of disparity discrimination and pattern detection for such an observer, whose performance is limited only by the presence of Gaussian noise of uniform variance in the stimulus, generated independently from trial to trial, from pixel to pixel, and in left and right eyes; the inevitably present photon noise conforms approximately to this description if the pattern contrast is not too high.
On any trial each pixel in each of the two binocularly viewed panels is assumed to generate for each eye a monocular excitation equal to the sum of (1) the mean stimulus intensity, over trials, for that pixel and (2) a Gaussian error term of mean zero and standard deviation $\sigma$. The ideal observer decides which panel has the crossed disparity and which the uncrossed by comparing the likelihood of the right-eye excitation pattern, given the left-eye excitation pattern, on the two alternative hypotheses (panel 1 uncrossed, panel 2 crossed; and vice versa: stimulus choices A and B in Fig. 12 below).

Random noise will cause the right-eye excitation to differ from the corresponding left-eye values. The binocular excitation difference between corresponding points that is due to noise, denoted $n_i$ for each pixel $i$ in the right-eye image, will have independent Gaussian distributions with zero mean and with variance $2\sigma^2$ (which is simply the sum of the noise variances for the two monocular excitations). The probability-density function of $n_i$,

$$p(n_i) = \frac{1}{(4\pi\sigma^2)^{1/2}} \exp \left[ -\frac{(n_i)^2}{4\sigma^2} \right], \quad (A1)$$

gives the likelihood of observing a difference $n_i$ between supposedly (and actually) corresponding pixels, on the correct hypothesis about the sign of the stimulus disparity. Consider next the incorrect hypothesis. Right- and left-eye pixels that actually represent corresponding points are now incorrectly assumed to represent points horizontally separated by a distance $2d$, where $d$ is the crossed or the uncrossed disparity present in each panel. The left- and right-eye excitations from actually corresponding pixels are accordingly expected to differ by the sum of (1) the random component $n_i$, which varies over trials with variance $2\sigma^2$, and (2) a systematic component $\Delta I_i$, which is due to the intensity difference between the two image locations involved and is constant over trials. The theoretical likelihood, on this incorrect hypothesis, of one's observing a difference $n_i$ between two actually corresponding left- and right-eye pixels becomes a Gaussian function of the deviation of $n_i$ from its theoretically expected value $\Delta I_i$:

$$p'(n_i) = \frac{1}{(4\pi\sigma^2)^{1/2}} \exp \left[ -\frac{(n_i - \Delta I_i)^2}{4\sigma^2} \right]. \quad (A2)$$

The natural log of $p'(n_i)/p(n_i)$, the relative likelihood of an observed binocular excitation difference $n_i$ (on a particular trial for pixel $i$), under the incorrect hypothesis as compared with the correct one, is the difference between the two exponents above:

$$\text{LLR}_i = -[(n_i - \Delta I_i)^2 - n_i^2]/4\sigma^2$$
$$= -[(\Delta I_i)^2 - 2\Delta I_in_i]/4\sigma^2. \quad (A3)$$

Over trials, the log likelihood ratio $\text{LLR}_i$ has a mean of $-(\Delta I_i)^2/4\sigma^2$ (since the term in $n_i$ averages to zero) and a variance equal to $(2\Delta I_i/4\sigma^2)^2\text{Var}(n_i) = (\Delta I_i)^2/2\sigma^2$. Since the differences observed for different pixels are statistically independent, the likelihood ratios for different pixels can be multiplied, or their logs added, to yield the overall likelihood ratio:

$$\text{LLR} = \sum_{2N} \frac{[(\Delta I_i)^2 - 2\Delta I_in_i]}{4\sigma^2}, \quad (A4)$$

where the sum is taken over all $2N$ right-eye pixels in the display. The mean and the variance of LLR are likewise obtainable through summing over pixels: the mean or expected value of LLR over trials is thus

$$E(\text{LLR}) = \sum_{2N} \frac{-(\Delta I_i)^2}{4\sigma^2}, \quad \text{(A5)}$$

and the variance is

$$\text{Var}(\text{LLR}) = \sum_{2N} \frac{(\Delta I_i)^2}{2\sigma^2}. \quad \text{(A6)}$$

To evaluate $\Sigma_{2N} (\Delta I_i)^2$, we denote the deviations of the monocular intensities for pixel $i$ from the mean intensity $I$ by $\Delta I_L$ and $\Delta I_R$, so that the binocular difference $\Delta I_i$ is $\Delta I_L - \Delta I_R$, and its square is $\Delta I_L^2 - 2\Delta I_L\Delta I_R + \Delta I_R^2$. Each of the two squared terms in this expression has a sum, over the $2N$ pixels of each eye's image, of $2N(\text{IC})^2$, where $C$ is the rms stimulus contrast and $I$ is the mean stimulus intensity. The mean of the cross product $2\Delta I_L\Delta I_R$ is equal to the mean-squared intensity excursion $(\text{IC})^2$ multiplied by $r(2d)$, the correlation between the stimulus intensities between the eyes at a horizontal separation (or disparity) of $2d$ (Fig. 12). The sum, over $2N$ pixels, of the cross-product term $2\Delta I_L\Delta I_R$ is thus $-4N(\text{IC})^2 r(2d)$. This yields

$$\Sigma (\Delta I_i)^2 = 4N(\text{IC})^2 [1 - r(2d)]. \quad \text{(A7)}$$

We can now specify the proportion of trials on which the ideal observer makes the correct decision. These are the trials on which LLR is negative, making the observations more likely under the correct hypothesis. The distribution of LLR is approximately Gaussian; when $\Sigma(\Delta I_i)^2$ is substituted for Eq. (A7) in Eqs. (A6) and (A5) above, the mean of LLR is seen to be $-N(\text{IC})^2(1 - r(2d))/\sigma^2$, and its variance $2N(\text{IC})^2 (1 - r(2d))/\sigma^2$. The desired probability is thus the integral of a normal probability-density function from $-\infty$ up to a standard deviate, or $Z$ value, of

$$d'_\text{discrim} = \sqrt{N/\sigma^2} \text{IC} \sqrt{1 - r(2d)}. \quad \text{(A8)}$$

In Eq. (A8), $d'_\text{discrim}$ remains constant if the stimulus contrast $C$ varies with disparity to satisfy

$$1/C = \sqrt{1 - r(2d)}. \quad \text{(A9)}$$

This relationship specifies the relative contrast sensitivity of the ideal observer in our experiment for different binocular disparities, in terms of the value of the autocorrelation function of the stimulus pattern for a displacement of twice the disparity. We have computed the autocorrelation functions of our stimuli for different disparities $r(d)$, and part of the autocorrelation function for the 3-cpd noise is presented in Fig. 12, along with an illustration of the task faced by the ideal observer. As is common, the autocorrelation functions are roughly quadratic for small disparities, so that the ideal observer's contrast sensitivity is directly proportional to test disparity, as it was for our observers. The proportionality between contrast sensitivity and disparity for small disparities
to the human tendency to be most sensitive to disparities that were approximately a quarter cycle (90°) of the spatial period of the center spatial frequency and for sensitivity to fall with higher phases. Our ideal observer of Eq. (A8) below may have some further generality; in particular, it could easily be modified to predict the data of Nakayama and Silverman for the detection of motion displacements at contrast threshold (see their Fig. 5a), without recourse to their less mechanistic vector model.

In the detection case a pattern appeared in one or other of the two binocularly observed panels, with zero binocular disparity, and the observer's task was to say in which panel the pattern appeared. Since the stimulus to be detected is the same in the two eyes, the appropriate strategy is to operate on the summed monocular excitations rather than on their difference. The summed excitation has a variance of $2\sigma^2$ for each of the $N$ pixels in each panel. Since the observer did not have prior knowledge of the particular pattern to be presented, the optimal detector is the "energy detector" of Green and Swets, which chooses the panel exhibiting the greatest mean-square deviation from uniformity. The observer does, however, in principle, have prior knowledge (after the first few trials in a block) of the stimulus center spatial frequency and hence knows which Fourier components are included in the stimulus pattern and which have been filtered out. This knowledge can be exploited for efficiency in detection by restriction of attention in the noisy image only to those frequency components that are transmitted by the filter. Note that prior knowledge of stimulus center frequency is in principle unnecessary for disparity discrimination; it need only be known that the pattern, whatever it is, is the same in the left and the right eyes.

In the case of the panel that has no pattern but only spatially random noise, the variance of $2\sigma^2$ associated with each of the $N$ pixels' summed binocular excitation is resolvable into independent contributions from $N$ sine implied by Eq. (1) and the curve in Fig. 6 is therefore expected on fairly general grounds for the task and stimuli involved. This is truer at lower than at high center spatial frequencies. At larger disparities the autocorrelation functions exhibit some oscillation, as is inevitable for a narrow-band filtered stimulus. But this oscillation only weakly affects the predicted contrast sensitivity and cannot account for the precipitous loss of sensitivity that we find when the disparity becomes large relative to the center period of the stimulus: the ideal observer continues to perform well with our stimuli at arbitrarily large disparities. In fact, the ideal observer possesses peak sensitivity at disparities much higher than the 90° phase shift found for the human observers, and therefore the loss of sensitivity was not due to the task and stimuli involved. These observations can be seen more clearly by inspection of Fig. 13. This figure shows the predictions of the contrast sensitivity for stereopsis of the ideal observer of Eq. (A8) against the human data of observer LW for center spatial frequencies of 1, 3, and 9 cpd [Figs. 13(a), 13(b), and 13(c), respectively]. We calculated the ideal observer's sensitivities by assuming that the same noise that limits disparity discrimination also limits simple contrast detection (see below), and this noise was set at a value that equated the contrast detection sensitivity of the ideal observer and of LW. Figure 13 shows that the ideal observer captures the trends in the human data for small disparities [note the proportionality evident for small disparities in Fig. 13(a) for the human and ideal observer's performance], but it continues to possess good sensitivity as disparity increases, whereas the human's sensitivity falls off. Although the ideal observer's sensitivity is fairly constant for large disparities, the peak predicted sensitivity is for more than 1000° at the frequencies of 3 and 9 cpd. This bears no relation to the human tendency to be most sensitive to disparities that are in principle unnecessary for disparity discrimination; it need only be known that the pattern, whatever it is, is the same in the left and the right eyes.

[Fig. 12. At the right, a robotic ideal observer faced with the disparity-discrimination task; the observer must decide between stimulus presentation alternatives A and B. To do so it makes use of information in the autocorrelation function of the stimulus as a function of disparity (an example of which is shown at the left). The ideal observer that we develop in Eq. (A8) has a predicted contrast sensitivity proportional to the square root of the length of the thick black bar depicted at the extreme left of the figure, which is the difference in the autocorrelation function of a 2d shift. Ripples in the autocorrelation function affect predicted sensitivity only weakly. See text for details.]
H. S. Smallman and D. I. A. MacLeod

or cosine phase Fourier components, varying randomly and independently in amplitude, and each with a Gaussian zero-mean distribution of amplitude over trials. The sum of the squares of the deviations from the space average, summed for all $N$ pixels, has an expected value over trials of $2N\sigma^2$; this implies a contribution of $2\sigma^2$ from each of the $N$ Fourier components. The ideal observer, however, considers only the smaller number of Fourier components, $N_t$, that could be present in the narrow-band-filtered stimulus and that hence could be useful for inferring the presence or absence of a signal. The sum of squares of the deviations from uniformity arising from the combined effects of these $N_t$ components alone has an expected value

$$E(SS_n) = 2N_t\sigma^2.$$  \hspace{1cm} (A10)

The distribution of this sum of squares over trials is the chi-square distribution for $N_t$ degrees of freedom and is approximately Gaussian when $N_t$ is not very small (see Ref. 65: for our stimuli $N_t > 28$); its variance is the sum of contributions from all the $N_t$ Fourier components, with each component contributing a variance equal to twice the square of its expected value:

$$\text{Var}(SS_n) = 8N_t\sigma^4.$$  \hspace{1cm} (A11)

In the case of the patterned panel, the effect of the pattern is to add to the noise amplitude for each of the sine or cosine Fourier components an additional pattern-derived component having a Gaussian distribution with zero mean over trials. The pattern variance for each transmitted component exceeds the noise variance by the factor $2N(IC)^2/N_t\sigma^2$, since the excursion in the mean-square binocular excitation that is due to pattern, $4(IC)^2$, is generated by only $N_t$ components, whereas the mean-square excursion that is due to noise, $2\sigma^2$, originates from $N$ components. When pattern and noise contributions are combined, the expected value of the squared excursion in binocular excitation summed over all $N$ pixels is thus increased from $2\sigma^2$ to $2\{2N(IC)^2/N_t + \sigma^2\}$ for each Fourier component. Once again, the variance from trial to trial is twice the square of the expected value. The mean and the variance of the summed squared deviation from the space average in this signal-plus-noise case are produced by summing over all $N_t$ Fourier components:

$$E(SS_m) = 4N(IC)^2 + 2N_t\sigma^2,$$  \hspace{1cm} (A12)

$$\text{Var}(SS_m) = 8N_t\{2N(IC)^2/N_t + \sigma^2\}.$$  \hspace{1cm} (A13)

Let $\Delta SS$ represent the amount by which the sum of squares for the patterned panel exceeds that for the panel without a pattern. This has a Gaussian distribution over trials with a mean equal to $E(SS_m) - E(SS_n)$ and a variance equal to $\text{Var}(SS_m) + \text{Var}(SS_n)$:

$$E(\Delta SS) = 4N(IC)^2,$$  \hspace{1cm} (A14)

$$\text{Var}(\Delta SS) = 8N_t\{2N(IC)^2/N_t + \sigma^2\}^2 + 8N_t\sigma^4.$$  \hspace{1cm} (A15)

Thus the signal detectability ($d'_{\text{detect}}$) for contrast detection in the conditions of our experiment was the expected value of $\Delta SS$ divided by its standard deviation:

$$d'_{\text{detect}} = \frac{\sqrt{NI}C}{\sqrt{2\sigma^2 + 2N(IC)^2/N_t + N(IC)^2}}$$  \hspace{1cm} (A16)

This analysis is useful because it now enables us to compare the disparity discrimination achieved by the human observer with the performance expected of the ideal observer. The ideal observer's performance in Eqs. (A8) and (A16) depends on the noise value $\sigma$. We assign...
to $\sigma$ the value that equates the ideal observer's detection performance [Eq. (A16)] and that of the human observer LW. The ideal observer's predicted disparity discrimination performance shown in Fig. 13 is actually better than detection performance for all but small disparities. Inspection of LW's data in Fig. 2 shows that the predicted superiority of discrimination over detection sensitivity was never realized by the human observer. However, the differences between detection and stereopsis contrast thresholds were small in some cases (0.1 log unit), and, as mentioned above, observer SPM's contrast and discrimination thresholds were essentially identical for 1-cpd noise presented with a 20-arcmin disparity. We can express the deviations of the human from the ideal level of performance, revealed in Fig. 13, in several ways. It seems clear that, for our ideal observer to be made more human, it would have to be limited by more noise than actually limits contrast detection (e.g., vergence jitter). One way to express the human deviations is in terms of the ratio of the observed to the expected contrast sensitivity; but more useful is the square of the ratio, which is then a DSQE measure, which is the factor by which the discrimination stimulus intensity for the ideal observer must be reduced to make the ideal observer's performance match the human's:

$$\text{DSQE} = \left(\frac{C_{\text{discrim}}}{C_{\text{ideal,discrim}}}\right)^2,$$

(A17)

where $C_{\text{discrim}}$ is the human contrast sensitivity for disparity discrimination and $C_{\text{ideal,discrim}}$ is the observed contrast sensitivity for disparity discrimination of the ideal observer. The latter quantity was calculated by determination of the noise value $\sigma$ that equated the ideal and the human contrast sensitivity. The DSQE is plotted against disparity in Fig. 11. This plot reinforces the notion of a size-disparity correlation operative in stereopsis; further discussion of the relevance of this plot can be found in Section 4.

ACKNOWLEDGMENTS

Portions of this research have been reported previously.\textsuperscript{9,10} This research was supported by National Institutes of Health grant EY-01711 to D. I. A. MacLeod and by a McDonnell—Pew graduate fellowship to H. S. Smallman. Thanks to S. P. McKee and C. W. Tyler for useful discussion, L. Welch for her splendid efforts as an observer, R. O. Brown and S. M. Anstis for useful comments on a previous draft of the paper, and S. E. Stevenson for so closely reviewing the manuscript.

Address any correspondence to Harvey S. Smallman, The Smith-Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco, California 94115.

REFERENCES AND NOTES


28. Using only a fixation spot to constrain vergence state without verification by nonius lines is unsatisfactory because of the ability of the eye to fuse with fixation disparities up.
to Panum's limit. Additionally, the sole use of convergent disparities in this study could have permitted observers to develop an anticipatory fixation disparity. It is interesting, further, to consider that such a fixation disparity would have reduced the 2.6° disparity between the panels of highest spatial frequency in the Mayhew–Frisby study to a 168-deg phase disparity, which is close to the point at which we find peak contrast sensitivity in the present study. Our results make the counterintuitive prediction that the use of a bigger disparity difference between the two Mayhew–Frisby panels would have made the discrimination harder.


33. After extensive use of this procedure it was decided that it was unnecessary to inquire of the subject the collinearity of the nonius lines before every trial. Instead, subjects were asked to redo trials in which they judged that the nonius lines were incorrectly aligned.


37. This prediction applies to the model in its simplest form employing solely vertically oriented mechanisms finely covering the range of spatial frequencies tested here and with no absolute positional disparity between the locations of the left and right eye’s receptive fields.


62. Included in this plot, but not shown in Fig. 2, are data taken at 1 cpd with 25-arcmin disparity.


