Pergamon

0042-6989(95)00110-7

Vision Res. Vol. 36, No. 3, pp. 391–399, 1996

0042-6989/96 $15.00 + 0.00

Is the Site of Non-linear Filtering in Stereopsis Before or After Binocular Combination?

LAURIE M. WILCOX,*† ROBERT F. HESS*

Received 25 October 1994; in revised form 17 March 1995

There is recent evidence that both linear and non-linear filtering operations subserve stereoscopic localization. For example, for spatially band-pass stimuli, the overall Gaussian envelope, which is not explicitly represented by the output of linear filters, can provide coarse disparity information. Here we ask three questions about the nature of this non-linear processing in stereopsis. First, is the site of the non-linearity before or after binocular combination? Second, is the stimulus envelope extracted by orientation or non-orientation selective spatial filters? Finally, we ask whether the envelope-based 3-D localization performance is similar to that for monocular 2-D localization as would be the case if the localization of the monocular contrast envelope was common to both operations. Our results suggest that envelope extraction occurs before binocular combination and that the filters involved are orientation selective. Finally, we provide preliminary evidence that is compatible with the proposal that 3-D and 2-D localization use the same envelope extraction operations.

Stereopsis Interocular correlation Non-linear filtering

INTRODUCTION

Spatial localization can be based on positional information derived from either the local changes in luminance of an image through linear filtering, or from the contrast envelope of a stimulus through non-linear operations such as half-wave rectification. We have examined linear and non-linear operations in stereopsis by measuring stereoacuity using Gabor stimuli (Hess & Wilcox, 1994). Our stereoacuity results showed that under some conditions (spatially broad-band patches) stereopsis depends on the spatial frequency content of these stimuli, while under other conditions (spatially narrow-band patches) performance depends on the size of the Gaussian envelope. The former can be explained by the more traditional models of stereopsis in which the output of linear filters, such as simple cells, provide the disparity signal. In the latter case however, one would need to postulate non-linear operations to make the contrast envelope explicit. Complex cells exhibit such a type of non-linearity (Spitzer & Hochstein, 1985). In a subsequent publication (Wilcox & Hess, 1995) we showed that the upper disparity limit for stereopsis (D_{max}) is determined *solely* by the size of the envelope. The fact that the stimulus envelope can provide a disparity signal demonstrates that there is an early non-linear input to stereopsis (e.g. rectification). The contribution of both linear and non-linear operations to stereoacuity has also been confirmed by Sato and Nishida (1993, 1994).

We now ask three questions concerning the nature of this envelope extraction. First, does it occur before or after binocular combination? We model binocular combination as a cross-correlation since this is the most widely accepted model (Tyler & Julesz, 1978; Poggio, Gonzalez & Krause, 1988; Cormack, Stevenson & Schor, 1991; Stevenson, Cormack & Schor, 1994). We define envelope extraction as the outcome of non-linear operations which generate a d.c. component that can signal the position of the contrast envelope. If the envelope is extracted before the site of binocular combination, then it might be similar to the envelope representation that is used for 2-D localization (Toet & Koenderink, 1988; Hess & Holliday, 1992). This is depicted in Fig. 1(A) as the contrast energy model. The alternative, that the site of envelope extraction for stereopsis occurs after binocular combination, is depicted in Fig. 1(B) as the disparity envelope model.

To determine if the envelope used for stereopsis is extracted before or after binocular combination, we measured stereoscopic performance for stereo-pairs consisting of spatially band-pass 1-D noise. The stereo-pairs were either matched in the two eyes (correlated) or unmatched (uncorrelated). The logic of these experiments is as follows. In the correlated noise condition, stereoacuity should be quite accurate because stereopsis will depend on the operations performed by linear spatial frequency and disparity-tuned filters.

*McGill Vision Research, Dept of Ophthalmology, McGill University, 687 Pine Avenue West, H4-14, Montreal, Quebec, Canada H3A 1A1.

†To whom all correspondence should be addressed [*Fax*: +1-514-843-1691].

However in the uncorrelated noise condition there will be no reliable disparity information available to the linear filtering stage; the only reliable disparity information will be represented in the relative positions of the stimulus envelopes. Consider Fig. 1(B), where the envelope is extracted after combination. The output of a cross-correlation of two *uncorrelated* noise patches will be close to zero. Therefore, there will be no signal available to a subsequent stage of envelope extraction, and so, no envelope-based disparity signal. Thus, if the envelope is extracted after binocular correlation it should not be possible to make depth judgements using our uncorrelated stereo-pairs. If, as shown in Fig. 1(A), the envelope is extracted before combination, then the envelope disparity will be represented at the correlation stage and it should be possible to perceive depth with the uncorrelated noise patches. This prediction was tested in the first experiment.

The second question we ask here is whether the spatial filters that are responsible for the extraction of the stimulus envelope are orientationally tuned. To address this issue we measured stereoacuity using uncorrelated noise patches (as described above), and varied the relative orientation of the noise in the two eyes. We measured stereoacuity for three stimulus conditions: vertical uncorrelated noise in both eyes, horizontal uncorrelated noise in both eyes, and vertical uncorrelated noise in one eye, horizontal uncorrelated noise in the other. We predicted that if the envelope disparity signal is derived from the output of non-oriented filters in the two eyes then stereoacuity should be measurable when the uncorrelated noise is orthogonally oriented in both eyes (horizontal vs vertical).

Monocular localization experiments for non-abutting targets have shown that performance is limited by the contrast envelope of the stimulus and not by the carrier's spatial frequency or orientation content (Toet, von Eekhout, Simons & Koenderink, 1987; Hess & Holliday, 1992). We asked in our final experiment whether the same operations might be used to extract the stimulus envelope for 3-D and 2-D localization. To address this issue we compared monocular and binocular localization performance for an identical task.

METHODS

Subjects and apparatus

Extensive measurements were obtained from two experienced subjects. Both subjects had excellent stereopsis as assessed using the Randot Stereotest and by their performance in previous stereoacuity experiments. Both subjects wore their prescribed optical correction. Stimuli were presented on a Joyce Electronics display screen with a P3 phosphor. The display was refreshed at 200 Hz, and had a vertical 100 kHz raster. The dimensions of the display area were 29 × 22.5 cm. The mean luminance of the display, viewed through the shutter glasses, was approximately 45 cd/m^2. To verify the linearity of our display system we drifted a complex pattern past a narrow slit and measured the luminance at each pixel using a UDT photometer. We then examined the power spectrum of the stimulus. This procedure was repeated at a number of positions on the display and revealed no evidence of distortion.

Stereoscopic depth was achieved using "Display Tech" liquid crystal shutters mounted in trial frames. A ± 10 V signal, supplied via a digital-to-analogue port, controlled the state of the shutters and was synchronized with the onset of each frame of the Joyce display. The stimuli for each eye were presented on alternate frames at a rate of 100 Hz/eye.* The reference stimuli were presented with zero disparity on all trials, while the target patch half-images viewed by the two eyes were offset in equal and opposite directions, by the amount required for each test condition.

Stimuli

The stimuli were patches of one dimensional, band-limited, spatial noise multiplied by a 2-D Gaussian envelope (see Fig. 2). Except when orientation was varied, the noise was vertically oriented. The patches were generated using a commercially available image-processing package, HIPS (Landy, Cohen & Sperling, 1984). To create the noise pattern a random number generator was used to select one of 256 grey-levels for each line of the image. This noise pattern was then filtered (convolved) with a Gabor function of the form:

$$G(x, y) = A \sin(2\pi f x) \cdot \exp[-(x^2 + y^2)/(2\sigma^2)], \quad (1)$$

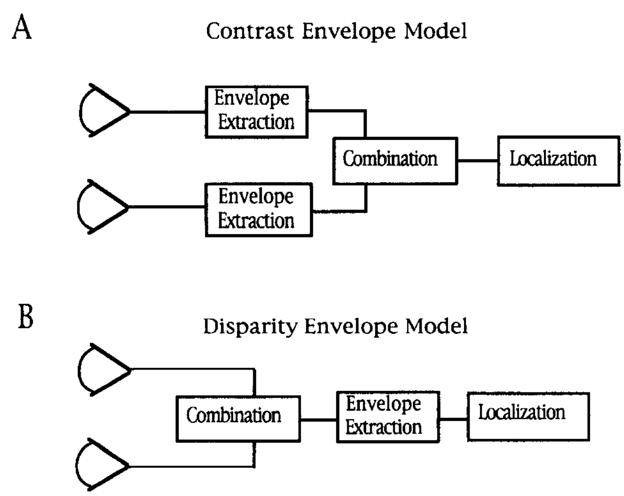

FIGURE 1. Shown here are two alternatives for the site of envelope extraction, relative to binocular combination. The envelope can be extracted monocularly, before binocular combination (contrast energy signal). The envelope could alternatively be obtained after binocular combination (disparity energy signal). The first set of experiments discriminates between these two alternatives.

*The liquid crystal shutters allow a very fast alternation rate, which can be faster than the decay time of a monitor's phosphor(s). In some situations (e.g. high contrasts) this results in cross-talk, or leakage, between the two eyes views. We have avoided this problem by using a display with a single, fast phosphor, and by ensuring that the stimulus contrasts used in our experiments were well below the threshold for detection of the cross-talk.

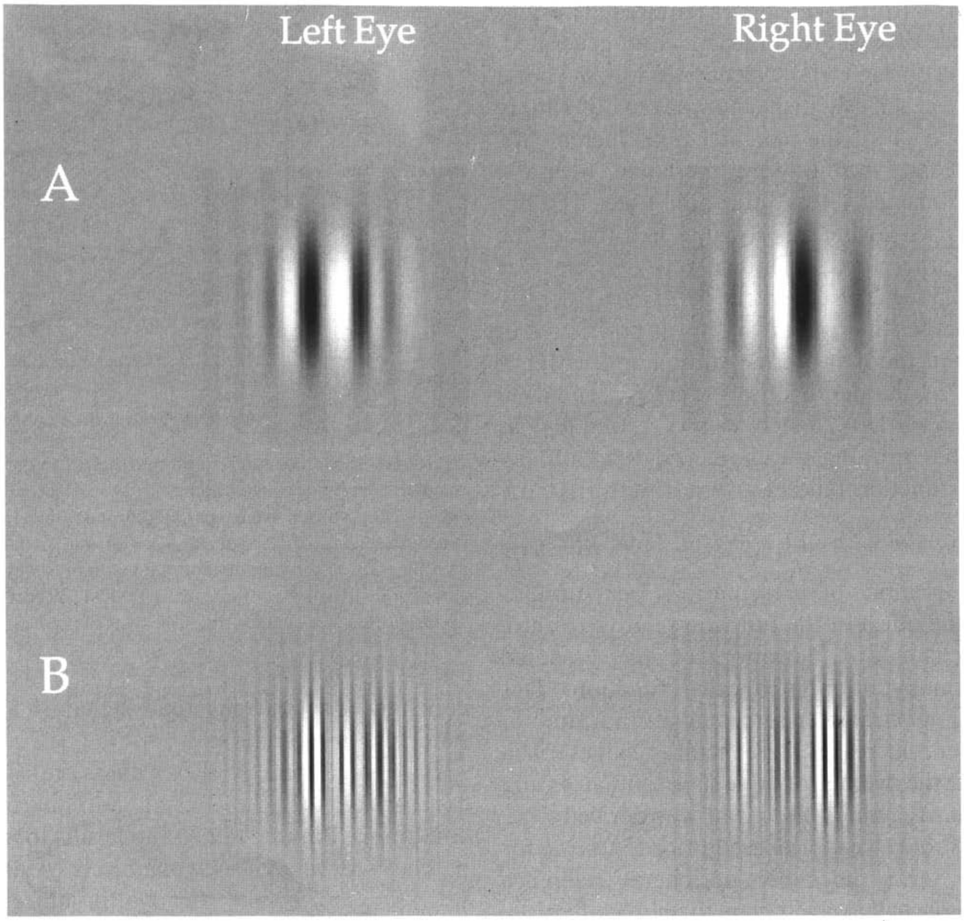

FIGURE 2. Shown here are examples of the band-limited, 1-D noise patches used in these experiments. The broad bandwidth patches (A) have an octave bandwidth of 1.89 with a centre frequency of 1.5 c/deg at 1 m, while the narrow bandwidth patches (B) have an octave bandwidth of 0.6 with a centre frequency of 5.76 c/deg at 1 m. Additional stimulus details are provided in the text.

where A is the amplitude of the function, σ is the standard deviation of the Gaussian envelope defining the patch and f is the frequency of the carrier grating. Gabors with different centre frequencies but similar standard deviations were used to generate noise patches of two octave bandwidths: 1.89 ($f = 1.5$ c/deg, $\sigma = 0.22$ deg at 1 m) and 0.6 ($f = 5.76$ c/deg, $\sigma = 0.17$ deg at 1 m). Examples of these broad and narrow bandwidth stimuli are shown in Fig. 2(A) and (B), respectively. In the final stage of processing, the patches were multiplied spatially by a broad 2-D Gaussian window with a standard deviation of 0.57 deg.

It is essential to the validity of our experiments that the independently generated noise patches be uncorrelated. While the use of a random number generator to assign luminance values to each pixel should ensure that each patch is uncorrelated, we confirmed this assumption by cross-correlating each of the individual images and averaging across the output images. We compared the *averaged* cross-correlation output with the results of a cross-correlation of two identical patches (autocorrelation). The peak of the autocorrelation function is 1.0 and that of the averaged cross-correlation distribution is 0.00397 (both functions were normalized). Thus, the independently-generated noise patches are, for all practical purposes, uncorrelated.

Procedure

The accuracy with which a single noise patch could be localized in depth was measured relative to two identical peripheral patches which formed the fixation plane. The two reference stimuli were located directly above and below the stereo-target and provided a stable fusion stimulus. The distance between the target and the reference patches was held constant at approximately 4 times the standard deviation of the Gaussian. Stereoacuity was measured using the method of constant stimuli, with a set of 11 stimuli which covered a range of crossed and uncrossed disparities. This range was chosen individually for each stimulus condition to bracket the subject's stereo-threshold, or the point at which the perceived location of the central stimulus changed from being "in front" to "behind" the peripheral patches. When required, sub-pixel spatial accuracy was achieved by introducing a lateral shift in the position of the noise patch as it was created. Sub-pixel accuracy was only required for correlated noise patches. When measuring stereoacuity with uncorrelated noise patches lateral displacements of < 1 pixel were never required. Thus, we were able to store several (11) independently generated, noise patches and on each trial select at random the noise to be presented to each eye; disparity was created simply by repositioning each of the pair in 1-pixel increments.

In all conditions, stimuli were presented within a temporal raised cosine of total duration 1 sec. The flat portion of the raised cosine was equal to half that of the rise and fall; therefore, stimuli were visible for approximately 0.5 sec. The observers' task was to identify on each trial whether the central target was positioned in front of or behind the two outside stimuli and within a single run each of the depth offsets were presented 20 times in random order. A stereoacuity estimate was derived from the resulting psychometric function, by fitting the error function (cumulative normal), ERF(x):

$$P(x) = A[0.5 + 0.5\text{ERF}((x - B)/(\sqrt{2.0}C))], \quad (2)$$

where A is the number of presentations per stimulus condition, B is the offset of the function relative to zero, and C is the standard deviation of the assumed underlying, normally distributed error function. This standard deviation parameter serves as an indicator of stereoacuity for as it increases, stereoacuity declines. Each datum represents the average of three such estimates from which the standard error of the mean was derived.

Contrast thresholds were measured prior to testing for all conditions. Since the more peripheral reference patches were likely to have a higher threshold than the foveated target (Pointer & Hess, 1989), we measured the contrast threshold for the reference patches (while fixating centrally) and set the contrast of all patches equal to 8 dB above this value. This ensured that all three patches were *at least* 8 dB above their contrast threshold. For all contrast measurements, we used the method of adjustment with a randomized starting point to obtain 7 binocular threshold estimates which were then averaged. Contrast was controlled by varying a (14 bit) voltage from the digital signal generator and multiplying it by the Gabor stimuli output from graphics memory, the contrast of which could also be scaled (8 bit resolution). Because the Joyce display has a linear Z-amplifier, no further adjustments were required to ensure accurate contrast manipulation.

RESULTS AND DISCUSSION

Is the site of envelope extraction before or after binocular combination?

Figures 3 and 4 show the stereoacuity performance of two subjects, for spatially broad (1.89 octaves) and narrow (0.6 octaves) bandwidth noise patches, respectively. First consider the uncorrelated noise condition (open circles). Although the stereo-thresholds are quite high, at all scales both subjects were able to make reliable depth judgements using these uncorrelated stimuli. Recall that if the envelope is *not* extracted before combination we would expect that subjects would be unable to perform this task. Given the results shown in Figs 3 and 4 we conclude that the envelope signal is extracted *before* binocular combination.

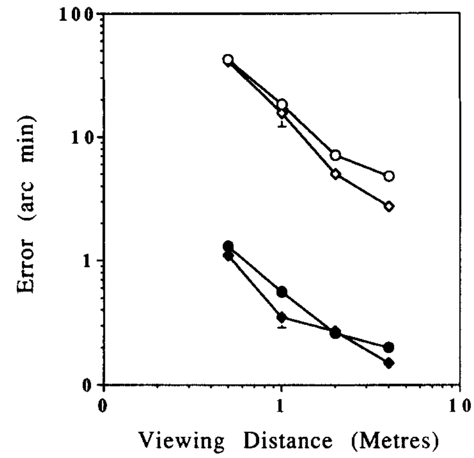

FIGURE 3. Stereoacuity is plotted here for correlated (filled) and uncorrelated (open) noise patches for two subjects LMW (◆◇) and JH (●○). The stimuli were spatially broadband (1.89 octaves) with a centre frequency of 1.5 c/deg and $\sigma = 34$ min, at 1 m. Data are plotted as a function of viewing distance and so represent the effects of scaling the stimuli. Error bars indicate ±1 SEM. Where error bars are not apparent they are smaller than the symbol used to represent the point.

In Figs 5 and 6 these data are replotted to aid comparison of the broad and narrow bandwidth conditions. Figure 5 shows the results for correlated noise patches separately for two subjects (A and B). In Fig. 6 we plot the results for both subjects (circles and diamonds) for uncorrelated noise patches (open symbols represent the broadband condition, filled symbols the narrowband condition). Although there is some variability, stereopsis is very similar for quite different bandwidths, with the exception of the longest viewing distance in the correlated noise condition.

At the longest viewing distance, stereoacuity for the narrow bandwidth patches deteriorates rapidly (Fig. 5). Most likely this occurs because the narrow bandwidth patches have a higher centre frequency than the broad bandwidth patches (see Methods/Stimuli). A similar

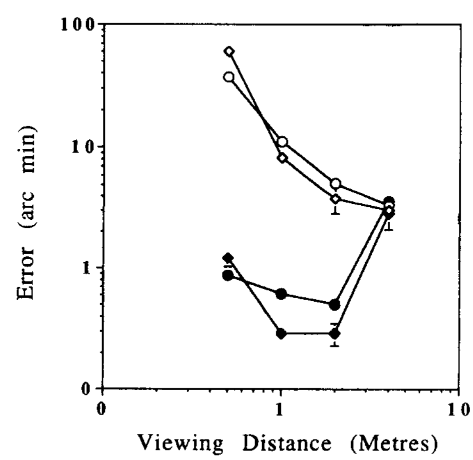

FIGURE 4. Stereoacuity is plotted for correlated (filled) and uncorrelated (open) noise patches for two subjects (●○ and ◆◇). In this graph the spatial bandwidth is narrow (0.6 octaves), the centre frequency is 5.76 c/deg and $\sigma = 34$ min at 1 m. Error bars indicate ±1 SEM.

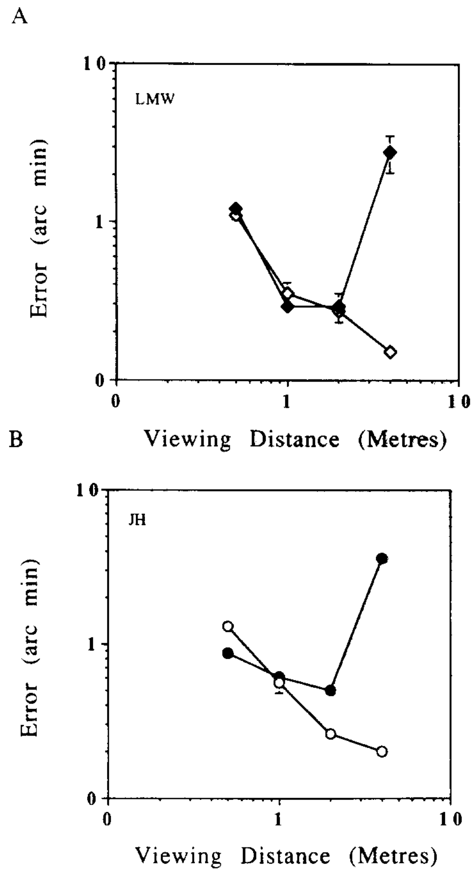

FIGURE 5. (A, B) The data from Figs 3 and 4 are replotted here to illustrate the effects of spatial bandwidth on stereoacuity for correlated noise for both subjects (A and B). Stereoacuity as a function of viewing distance is plotted for spatially narrowband (filled) and broadband (open) noise patches, their octave bandwidths were 0.6 and 1.89, respectively. Error bars indicate ± 1 SEM.

deterioration of stereoacuity occurred in a previous experiment, in which we measured stereoacuity as a function of the centre frequency of Gabor patches (Hess & Wilcox, 1994). At each of three patch sizes we found that over a large range of frequencies there was improved performance with increasing frequency. However, as the spatial frequency was increased, and the octave bandwidth was narrowed, there was a point on each curve (at approximately 0.6 octave bandwidth) where performance suddenly deteriorated.*

There is evidence for an upper limit on the spatial frequency information that can be utilized by the linear stereoscopic system (Westheimer & McKee, 1980). Since the narrowband noise patch has a higher centre frequency this upper limit will be attained at a shorter viewing distance than for the broad bandwidth patch. Interestingly, performance is *identical* in the correlated and uncorrelated narrowband conditions for the 4 m viewing distance (see Fig. 4). Our explanation for this correspondence is that in both cases the disparity signal provided by the stimulus envelope limits performance. For uncorrelated noise stereo-pairs the envelope is used at all viewing distances because the phase structure is randomized. However, when the narrow bandwidth, correlated noise is viewed from 4 m, the stimulus envelope is used because the amplitude spectra of the stimuli are outside the range for which linear filtering operations can be used to extract a disparity signal.

Are envelope filters orientation tuned?

If the envelope disparity signal is extracted from the outputs of non-oriented linear filters, then it should be possible to extract an envelope-based disparity signal when orthogonally oriented noise patches are presented to the two eyes. For example, the simplest model of envelope extraction would involve the summing of geniculate afferents which are known to be half-wave rectified. However, if the envelope can only be extracted from the outputs of oriented filters, then it will be impossible to see depth when viewing orthogonally oriented stereo-pairs.

We tested this prediction using the broad bandwidth, uncorrelated, noise patches described above with the same methodology and subjects. Prior to measuring stereopsis we obtained subject's detection threshold for these noise patches, using the method of adjustment (see Methods). In all subsequent testing the contrast was set to 8 dB above the contrast threshold. We measured stereoacuity for three types of noise stereo-pairs: horizontal–horizontal, vertical–vertical and horizontal–vertical. The results for two subjects are shown in Fig. 7. It is clear that while there is little difference in stereoacuity for uncorrelated horizontal and vertical pairs, stereoacuity for the horizontal–vertical condition is severely impaired. In fact, the psychometric functions for this

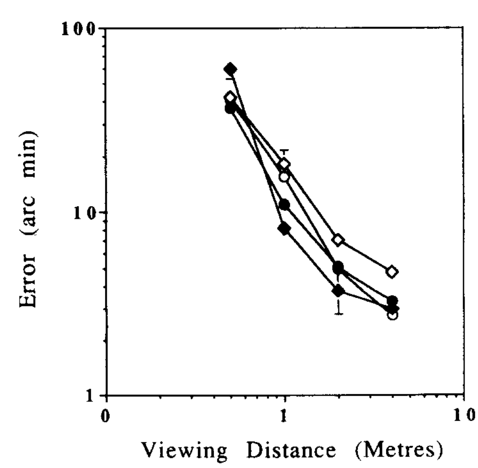

FIGURE 6. The data from Figs 3 and 4 are replotted here to illustrate the effects of spatial bandwidth on stereoacuity for uncorrelated noise for both subjects (◆◇ and ●○), for narrow (filled) and broad (open) bandwidth stimuli. Error bars indicate ± 1 SEM.

*In their investigation of the spatial frequency dependence of stereoacuity Schor and Wood (1983) did not observe this decline in stereoacuity at high spatial frequencies. The key to this apparent discrepancy lies in the stimuli used in the two experiments. We used Gabor stimuli which allowed us to independently vary size and frequency content. Therefore, we were able to test a wide range of octave bandwidths. Schor and Wood (1983) used broad bandwidth, difference-of-Gaussian (doG) patches for which size and spatial frequency covary. The octave bandwidth of their patches was fixed at 1.75. Schor and Wood (1983) did not observe this deterioration in stereoacuity at high frequencies because it only occurs for narrow (less than 0.6 octaves) bandwidth stimuli.

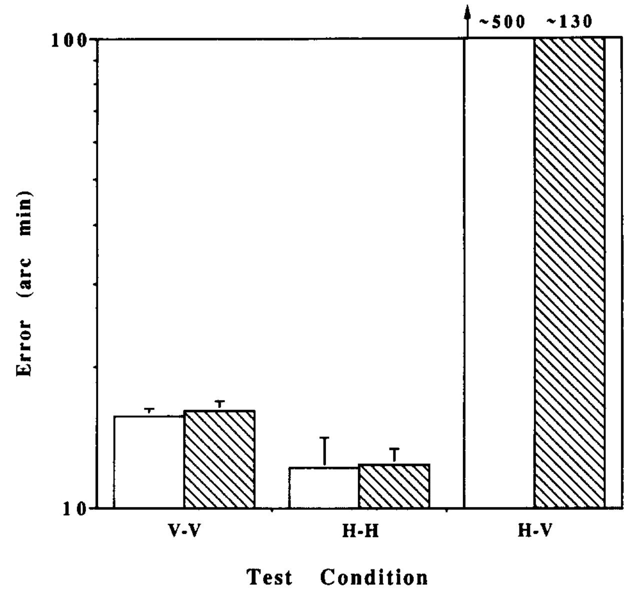

FIGURE 7. Stereoacuity is plotted as a function of the interocular orientation of *uncorrelated* noise patches. From left to right are data for the vertical–vertical, horizontal–horizontal and vertical–horizontal conditions, for two subjects LMW (□) and JH (▧). Although there appears to be an index of stereoacuity for the vertical–horizontal condition, these numbers are misleading. In all cases the psychometric functions oscillated about 50% indicating that stereopsis was impossible in this condition. Error bars indicate +1 SEM.

condition were essentially flat, corroborating subjects' reports that depth could not be perceived for these stimuli.

If envelopes extracted from the outputs of filters tuned to different orientations can be compared to derive a depth signal then performance in the horizontal–vertical condition should be comparable to that in the two matched-orientation conditions. However, subjects were unable to make reliable relative depth judgements for these stereo-pairs, but did make consistent depth judgements in the matched orientation conditions. We conclude that the envelope disparity signal is extracted by orientation-tuned filters and that the disparity signal is derived from a comparison of the outputs of like-oriented filters.

An alternative explanation of these results is that depth was not perceived with orthogonally-oriented noise stereo-pairs because of an interference effect of rivalry caused by placing orthogonally oriented stimuli in similar locations in the two eyes. At the low contrasts used in this experiment rivalry was imperceptible. However, to be sure that some aspect of rivalry was not responsible for the absence of depth perception for orthogonal stereo-pairs we conducted a simple test. Both subjects viewed horizontal–vertical stereo-pairs at disparities beyond the fusion range where, because of diplopia, rivalry could not be induced. Although the patches were diplopic, the disparity pedestal of 1 deg was well below the upper disparity limit for stimuli of this size (Wilcox & Hess, 1995). Apart from the addition of the disparity pedestal, the apparatus and procedure was identical to that described above. Spatially broad bandwidth (1.89 octaves) noise patches were used, and the psychometric functions plotted in Fig. 8 represent the average of two functions for each subject. Again it was not possible to make reliable depth judgements with these orthogonally oriented stimuli, thus confirming that rivalry was not responsible for our subjects' inability to see depth with such stimuli.*

What is the relationship between 2-D and 3-D envelope-based localization?

Given that the stimulus envelope used for stereopsis is extracted before binocular combination, it is possible that the same operations are used to identify it that are used for monocular localization. If this is true then we would expect 3-D and 2-D localization performance to be similar for a task in which both versions require the use of the stimulus envelope. Toet *et al.* (1987) and Hess and Holliday (1992) have demonstrated that for spatial frequency band limited stimuli that are a fixed amount above contrast threshold and non-abutting, 2-D localization is limited by the stimulus envelope. These conditions are identical to those used in our preceding stereoacuity experiments. Therefore, to compare 3-D and 2-D localization, we measured *monocular* localization using the same stimuli, subjects and procedure described above. The only difference between the tasks was that, instead of making a depth judgement, subjects were required to indicate if the central noise patch was to the left or to the right of the vertically aligned reference patches. Monocular localization performance as a function of viewing distance is represented by the filled symbols in Fig. 9, for spatially broad (triangle) and narrow (inverted triangle) bandwidths. Stereoacuity results are represented by dashed (narrow bandwidth) and dotted (broad bandwidth) lines and open circles (uncorrelated) and squares (correlated). Monocular localization is a factor of 10 worse than stereoacuity for

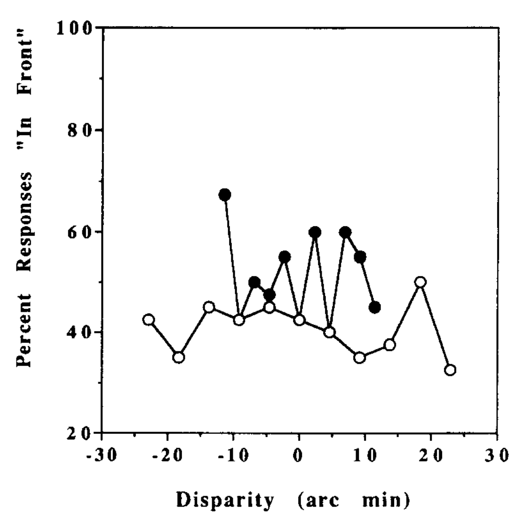

FIGURE 8. Psychometric functions obtained by measuring stereoacuity for horizontal–vertical stereo-pairs with a 1.0 deg disparity pedestal. Results (average of two psychometric functions) are shown for two subjects: LMW (○) and JH (●).

*One subject (LMW) also attempted this task at a pedestal of 0.5 deg. Again it was not possible to make reliable depth judgements at any disparity.

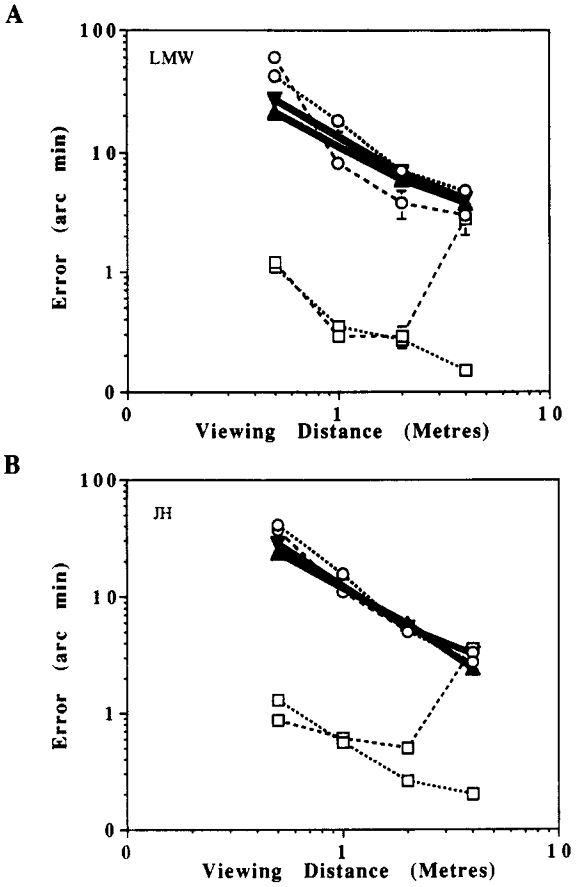

FIGURE 9. Monocular localization results are indicated by solid lines and filled symbols for broad (▲) and narrow (▼) bandwidth noise patches. For comparison, data from Figs 3 and 4 are replotted as - - - (narrow bandwidth) and · · · (broad bandwidth) lines and open symbols (□ represent correlated noise and ○ represent uncorrelated noise). Error bars indicate ±1 SEM.

correlated noise patches, in spite of the fact that the stimulus arrangement viewed by each eye was identical in the two conditions. Instead, monocular localization is similar to stereoacuity when the left and right eye stereo-pairs are uncorrelated. These results are consistent with the proposal that early non-linear operations which extract the stimulus envelope are common to 3-D and 2-D localization.

Relation to previous work

Orientation tuning. We have shown here that envelope-based stereopsis requires input from like-oriented filters. Liu, Tyler, Stevenson and Ramachandran (1992) have reported that depth identification is possible using ±45 deg Gabor stereo-pairs. It is tempting to explain this apparent discrepancy by differences in methods and stimuli between the two studies (e.g. contrast, noise/gratings, absolute orientation etc.). However, a more parsimonious explanation is that, with their stimuli, disparity information is available to linear filters which are broadly tuned for orientation.

Second-order stereopsis. There are several previous reports in the literature of depth perceived in stereograms in which there is no luminance edge defining the disparate region (see Frisby & Mayhew, 1978; Ramachandran, Rao & Vidyasagar, 1973a; Ramachandran *et al.*, 1973b). For example, Ramachandran *et al.* (1973a) reported reliable depth discrimination for patterns whose disparate regions were defined by changes in texture, line orientation, and contrast polarity. The results of such experiments have been classified as "second-order" because the disparate region is defined by texture, orientation or contrast polarity rather than a more traditional luminance-defined edge. However, the disparity in these stereograms is available to a *set* of linear filters having different optimum frequency and/or orientation responses. Therefore, such stimuli do not produce unequivocal evidence for the operation of the non-linear mechanism of the type defined here.

Another class of experiments have measured stereopsis for uncorrelated patterns which have monocular motion-defined contours (Lee, 1970; Prazdny, 1984; Halpern, 1991). As noted by Halpern (1991), these experiments demonstrate that the monocular extraction of contours can precede binocular matching. This statement is consistent with our results. However, since the experiments listed above depend critically on the processing of motion signals prior to processing disparity information, it is not clear what their relationship might be to our results. It is possible that the non-linear mechanism that we have identified is also used to interpret the disparity signals provided by motion. It is important to note that a number of different non-linear operations have now been defined for different types of visual processing (e.g. motion, texture, spatial alignment and stereopsis); their inter-relationships are not yet understood.

More relevant to the results presented here are those of Carney and Shadlen (1984), who found that depth identification was possible when a disparity signal was provided by contrast modulation of uncorrelated noise in the two eyes. Comparison of the two experiments is difficult because of stimulus differences; for example, we used enveloped patches while their stimuli were full-field. However, it is possible that performance in their task was mediated by the same non-linear mechanism as discussed here.

2-D vs 3-D localization with Gabor patches. The similarity of monocular localization and stereoacuity in our third experiment is somewhat surprising. In a previous study (Hess & Wilcox, 1994) we found that monocular localization performance for three vertically separated Gabor patches was poorer than stereoacuity, even when the envelope was limiting performance (narrow bandwidth patches). In these experiments, for broad bandwidth Gabor patches stereoscopic localization was approximately 20 times better than monocular localization. This difference dropped by half for narrowband Gabor patches; stereopsis was better by a factor of 10.

When we used Gabor stimuli to measure stereoacuity (Hess & Wilcox, 1994), the envelope appeared to limit performance for narrow bandwidth stimuli. We suggested that when the bandwidth was <0.6 octaves the disparity signal provided by the carrier grating was unreliable, and therefore not used to make the depth judgement. However, we cannot be certain that the carrier did not make *any* contribution to the accuracy of the depth percept. It is possible that the improvement in stereopsis relative to monocular localization with Gabor patches can be attributed to some contribution of the interior spatial frequency content (carrier grating) as we proposed in the preceding discussion of the experiment of Liu *et al.* (1992). However, in the experiments reported here, the noise carrier does not provide a disparity signal; therefore performance must depend on the envelope-based disparity signal. Under these limiting conditions, monocular and stereoscopic localization performance are very similar.

Caveat

In answering the first of our questions we assumed that binocular combination for stereopsis involves cross-correlation of the monocular inputs (Tyler & Julesz, 1978; Poggio *et al.*, 1988; Cormack *et al.*, 1991; Stevenson *et al.*, 1994). We will use $CC(\tau)$ to denote the standard cross-correlation operation and assume that stereoacuity is some function of the peak of the cross-correlation operation. Let us further use the functions $f(x)$ and $g(x)$ to denote the signals from the two eyes. In our first experiment we distinguished between the possibility that the (rectifying) non-linearity precedes the cross-correlation stage [Fig. 10(A)], such that

$$CC(\tau) = \int |f(x)| \cdot |g(x+\tau)| \, dx; \quad (3)$$

and the possibility that the (rectifying) non-linearity follows the cross-correlation stage [Fig. 10(B)], such that

$$|CC(\tau)| = \left| \int f(x) \cdot g(x+\tau) \, dx \right|. \quad (4)$$

Assuming the canonical form of cross-correlation, where summation directly follows multiplication, our results are consistent with equation 3. However, Fig. 10(C) illustrates a modified form of a cross-correlation operation where a rectification operation is inserted between multiplication and subsequent summation:

$$CC(\tau) = \int |f(x) \cdot g(x+\tau)| \, dx. \quad (5)$$

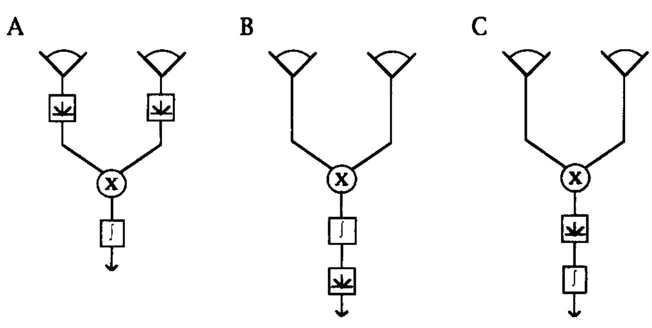

FIGURE 10. Shown here are examples of the potential order of envelope extraction (rectification) and cross-correlation operations. (A, B) The two possibilities assessed in Experiment 1, where envelope extraction either precedes (A) or follows (B) combination. In (C) we present a modified form of cross-correlation in which the signal is rectified after the outputs from the two eyes are multiplied but before they are summed.

Given the equality

$$\int |f(x)| \cdot |g(x+\tau)| = \int |f(x) \cdot g(x+\tau)| \, dx, \quad (6)$$

models A and C in Fig. 10 are indistinguishable. Therefore, for forms of binocular combination which do not involve a standard cross-correlation, our results suggest that the non-linearity must *at least* precede summation.

Computational modeling and neurophysiology

The vast majority of computational models of stereopsis have described the initial stereoscopic processing stage as a bank of linear, spatial frequency and orientation selective disparity detectors. Recent models have included non-linear operations that might make other forms of stereoscopic information available. These models fall into two categories. Either they derive their disparity from quadrature spatial filters in different eyes (Bowne, McKee & Tyler, 1990; Jacobson, Gaska & Pollen, 1993; Qian, 1994) or the same eye (Jacobson *et al.*, 1993). In the first case, the computation is based on disparity energy, whereas in the second case it is based on the disparity of the local contrast energy. The results of experiment 1 are consistent with both of these however, the similarity in 2-D and 3-D localization shown in Fig. 9 favours the contrast energy model.

Poggio and Fischer (1977), Ferster (1981) and others have shown that cortical neurons can be divided into categories roughly similar to the crossed, uncrossed and zero disparity classes proposed by Richards (1970). To summarize, one class of cell is tuned to near zero disparities and responds either with excitation or inhibition (tuned excitatory or inhibitory). The other two classes respond positively to either crossed or uncrossed disparities, these are termed near/far cells. While the tuned cells are considered optimal for detecting small disparities near the horopter, near/far cells would be well-suited for processing large disparities off the horopter.

Poggio *et al.* (1988) measured responses of striate cortical cells in the monkey to correlated and uncorrelated random dot stereograms (RDS). The responses of cells tuned to near zero disparities were suppressed when viewing uncorrelated RDS. In contrast, near/far cells responded equally to both correlated and uncorrelated stereograms. The apparent resilience of the response of near/far cells to uncorrelated local contrast information suggests they are candidates for subserving stereopsis in the uncorrelated noise conditions tested here.

Note added in proof

In a recent book chapter [Papathomas, Chubb, Gorea & Kowler, Eds, *Early vision and beyond* (1995)], Tyler described an experiment in which he found that stereoacuity for gabor patches oriented at ± 45 deg was similar to that obtained for Gaussian luminance blobs. Since disparity can be derived from each of these stimuli using *linear* filters broadly tuned for orientation, their result provides further support for the position that we outline in the section on Orientation Tuning.

REFERENCES

Bowne, S., McKee, S. P. & Tyler, C. W. (1990). A disparity energy model for early stereo processing. *Investigative Ophthalmology & Visual Science (Suppl.)*, *31*, 303.

Carney, T. & Shadlen, M. (1984). Stereopsis with amplitude modulated gratings: A role for form? *Investigative Ophthalmology & Visual Science (Suppl.)*, *82*, 294.

Cormack, L. K., Stevenson, S. B. & Schor, C. M. (1991). Interocular correlation, luminance contrast and cyclopean processing. *Vision Research*, *31*, 2195–2207.

Ferster, D. (1981). A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. *Journal of Physiology*, *311*, 623–655.

Frisby, J. P. & Mayhew, J. E. (1978). The relationship between apparent depth and disparity in rivalrous–texture stereograms. *Perception*, *17*, 661–678.

Halpern, L. (1991). Stereopsis from motion-defined contours. *Vision Research*, *31*, 1611–1617.

Hess, R. F. & Holliday, I. (1992). The coding of spatial position by the human visual system: Effect of spatial scale and contrast. *Vision Research*, *32*, 1085–1097.

Hess, R. F. & Wilcox, L. M. (1994). Linear and non-linear contributions to stereopsis. *Vision Research*, *34*, 2431–2438.

Jacobson, L., Gaska, J. P. & Pollen, D. A. (1993). Phase, displacement and hybrid models for disparity coding. *Investigative Ophthalmology & Visual Science (Suppl.)*, *34*, 908.

Landy, M. S., Cohen, Y. & Sperling, G. (1984). HIPS: Image processing under UNIX. Software and applications. *Behaviour, Research Methods, Instruments and Computers*, *16*, 199–216.

Lee, D. N. (1970). Binocular stereopsis without spatial disparity. *Perception and Psychophysics*, *9*, 216–281.

Liu, L., Tyler, C. W., Schor, C. M. & Ramachandran, V. S. (1992). Positional disparity is more efficient in encoding depth than phase disparity. *Investigative Ophthalmology & Visual Science (Suppl.)*, *33*, 1373.

Poggio, G. & 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., Gonzalez, F. & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity. *Journal of Neuroscience*, *8*, 4531–4550.

Pointer, J. S. & Hess, R. F. (1989). The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. *Vision Research*, *29*, 1133–1151.

Prazdny, K. (1984). Stereopsis from kinetic and flicker edges. *Perception and Psychophysics*, *36*, 490–492.

Qian, N. (1994). Computer stereo disparity and motion with known binocular cells. *Neural Computation*, *6*, 390–404.

Ramachandran, V. S., Rao, M. V. & Vidyasagar, T. R. (1973a). The role of contours in stereopsis. *Nature*, *242*, 412–414.

Ramachandran, V. S., Rao, M. V., Sriram, S. & Vidyasagar, T. R. (1973b). The role of colour perception and "pattern" recognition in stereopsis. *Vision Research*, *13*, 505–509.

Richards, W. (1970). Stereopsis and stereoblindness. *Experimental Brain Research*, *10*, 380–388.

Sato, T. & Nishida, S. (1993). Second-order depth perception with texture-defined random-check stereograms. *Investigative Ophthalmology & Visual Science (Suppl.)*, *3642*, 1438.

Sato, T. & Nishida, S. (1994). Does an envelope-detecting mechanism mediate stereopsis for band-limited stimuli? *Investigative Ophthalmology & Visual Science (Suppl.)*, *35*, 1916.

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

Spitzer, H. & Hochstein, S. (1985). A complex-cell receptive-field model. *Journal of Neurophysiology*, *53*, 1266–1286.

Stevenson, S. B., Cormack, L. K. & Schor, C. M. (1994). The effect of stimulus contrast and interocular correlation on disparity vergence. *Vision Research*, *34*, 383–396.

Toet, A. & Koenderink, J. J. (1988). Differential spatial displacement discrimination of Gabor patches. *Vision Research*, *28*, 133–143.

Toet, A., von Eekhout, M. P., Simons, H. L. & Koenderink, J. J. (1987). Scale invariant features of differential spatial displacement discrimination. *Vision Research*, *27*, 441–452.

Tyler, C. W. & Julesz, B. (1978). Binocular cross-correlation in time and space. *Vision Research*, *18*, 101–105.

Westheimer, G. & McKee, S. (1980). Stereoscopic acuity with defocused and spatially filtered retinal images. *Journal of the Optical Society of America*, *70*, 772–787.

Wilcox, L. M. & Hess, R. F. (1995). D_{max} for stereopsis depends on size, not spatial frequency content. *Vision Research*, *35*, 1061–1069.

Acknowledgements—We would like to thank Eric Fredericksen and David Field for many helpful discussions, James Elder for his constructive and insightful criticism, and Joey Hum for her unfailing participation as a subject in these experiments. This work was supported by the Medical Research Council of Canada (grant No. MT 10818).