PERGAMON

Vision Research 39 (1999) 4333–4340

Vision Research

www.elsevier.com/locate/visres

Depth aliasing by the transient-stereopsis system

Mark Edwards *, Clifton M. Schor

School of Optometry, University of California, Berkeley, CA 94720, USA

Received 18 February 1999; received in revised form 23 June 1999

Abstract

A fundamental problem in stereo-processing is determining which images in the two eyes correspond to the same object. This problem is particularly pronounced with periodic stimuli where it is theoretically possible to binocularly match a given feature in one eye with any of the identical features in the other eye. One way to minimise the likelihood of the occurrence of such aliasing is to restrict the upper-disparity limit that a particular binocular cell can process to one-half of the spatial period to which the cell is sensitive. While such a restriction would not be a major problem for the sustained stereo-system (which processes small disparities) it would be for the transient system (which is capable of processing disparities as large as 10°). Large-field sinewave variations in luminance were used to compare the propensity of the sustained and transient systems to exhibit depth aliasing-that is to signal a depth sign that corresponds to a binocular match that is greater than the nearest-neighbour pairing. Results were that: depth aliasing was exhibited at short, but not at long durations; decreasing the disparity of the stimulus reduced the likelihood of depth-aliasing; and the critical disparity for this reduction in depth aliasing was dependent upon the spatial frequency of the stimulus, i.e. it was phase, not absolute disparity dependent. Based upon these results, we conclude that while the sustained system implements the half-cycle disparity-processing limit, the transient system does not.

Keywords: Transient; Sustained; Stereopsis

1. Introduction

The extraction of 3-dimensional depth from the two 2-dimensional retinal images requires determining which retinal image in each eye corresponds to the same object. That is, the correspondence problem needs to be solved. One popular class of model that has been proposed to solve this problem implements a coarse-to-fine processing strategy (Marr & Poggio, 1979). Within this approach, the disparity processed by any given cell does not exceed one-half the greatest spatial-period (i.e. 180° phase offset) to which each binocular cell is sensitive. Such an approach is particularly useful in the processing of periodic stimuli because it minimises the risk of aliasing: the matching of features in a periodic stimulus that are separated by more than 180° of phase angle. To process both small and large disparities, a system utilising such an approach would need to employ binocular cells whose overall spatial-frequency tuning, and hence spatial half-periods, encompass the required disparity range. There is good evidence that the stereo system does (at least in part) employ such an approach (e.g. Schor, Wood & Ogawa, 1984a; Blake & Wilson, 1991).

A drawback with the half-cycle-limit approach is that, by definition, it restricts the maximum disparity that a given binocular cell can process. Even if cells tuned to a relatively low spatial-frequency are used (e.g. 0.5 cpd) the upper disparity limit that can be processed by such a system employing such an approach is still relatively small (e.g. 1°). While this restriction upon the upper-disparity limit would not be such a great problem for the sustained stereo-system, it would be for the transient system (Schor, Edwards & Pope, 1998). Unlike the sustained system, which only processes relatively small disparities that are substantially within the fusible range (i.e. typically less than 1°) the transient system processes large disparities: up to 10° in size (Westheimer & Tanzman, 1956). Thus, while the half-cycle limit approach is appropriate for the sustained system, it is insufficient to account for the large dispar-

* Corresponding author. Fax: + 1-510-643-5109.

E-mail address: mark@hering.berkeley.edu (M. Edwards)

0042-6989/99/$ - see front matter
PII: S0042-6989(99)00149-2

ity range of the transient system. For example, in order to process a disparity of 10°, a cell tuned to a spatial frequency of 0.05 cpd would be needed.

It is therefore possible that the sustained system may employ the half-cycle-limit approach, and hence would not exhibit depth aliasing, while the transient system doesn't, and hence could exhibit aliasing. In the present study, depth aliasing is defined as the perception of depth in the opposite direction to that corresponding to the nearest-neighbour (< 180° phase shift) binocular match (See Fig. 1). Stimuli used were large-field periodic stimulus that contained many possible depth matches (see below). The starting point for the present study was the observation that while long-duration observation of such stimuli did not result in depth aliasing, brief observations could. Note that we have previously linked the processing of short duration stimuli to the transient system (Pope, Edwards & Schor, 1999). In the following experiment we determined the conditions under which transient depth aliasing would occur.

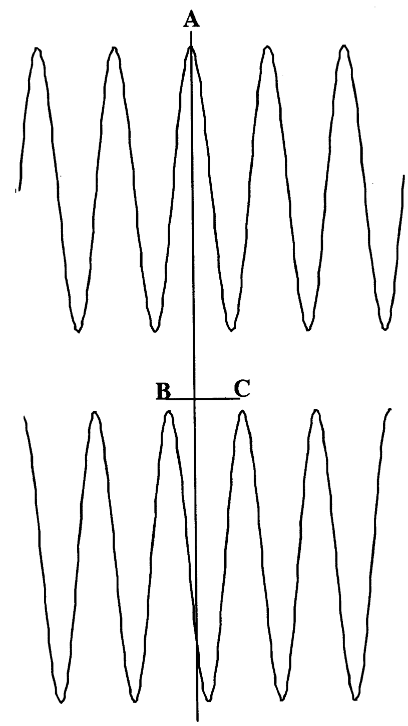

Fig. 1. Two of the many possible depth matches with a repetitive stimulus. Compared to the luminance sinewave presented to the left eye, the one presented to the right eye is has its phase offset by 90° (to the left). Matching feature A (left eye) with feature B (right eye) corresponds to matching the nearest-neighbour features (those separated by 90°). Matching features A and C corresponds to matching the next-to-nearest-neighbour features (those separated by 270°, i.e. 360 − 90°).

2. Experiment: transient-depth aliasing with periodic stimuli

The aim of this experiment was to determine the conditions under which transient-depth aliasing would occur. Parameters manipulated included the spatial frequency, phase offset and the spatial extent of the sinewave stimuli.

2.1. Method

2.1.1. Observers

Three observers were used. The two authors and an observer who was naive with respect to the aims of the experiment. All observers had corrected to normal visual acuity, normal stereopsis (as measured by a Randot Stereotest) and no history of any binocular visual-disorders.

2.1.2. Stimuli and procedure

Stimuli consisted of luminance sinewave-gratings. Spatial frequencies of 0.3, 0.6 and 1.8 cpd and disparity-phase offsets of 22.5, 45, 90 and 135° were employed. Two spatial extents (15 and 30°) were used for two of the observers (CS and ME) while the third observer (MS) was only tested with the largest stimulus extent (30°). To minimise the potential contributions of the edge (envelope) of the stimulus to the perceived depth-sign (near or far) it's disparity offset was kept at zero, i.e. the fixation depth.

The observer first maintained fixation on a pair of crosses and vertical nonius lines. Once the observer had established fixation, while perceiving the nonius lines aligned, they initiated the presentation of the test stimulus. The test stimulus replaced the fixation cross and nonius lines and consisted of two pairs of dichoptic sinewave gratings, with one placed above and the other below the original location of the fixation point. One of these pairs was presented at a crossed and the other at an uncrossed horizontal disparity relative to the depth that had been defined by the fixation point. The observer's task was to indicate which fused sinewave (upper or lower) was perceived to be nearest in depth to them. A vertical gap of 1.35° was maintained between the upper and lower stimuli to ensure that stimulus crowding did not occur (Westheimer & McKee, 1979). The viewing distance was 0.5 meters, the mean luminance of the display was 25 cd/m^2 and the contrast was 25%. Stimuli were presented in blocks of 20. Each data point reported represents the mean of 10 blocks of trials. In order to selectively drive the transient system, a stimulus duration of 140 ms was used (Pope et al., 1999).

With such a stimulus, a number of potential stereoscopic matches can be made. We will consider two of them. With a 90° phase offset between the two dichoptic images, matching the nearest features (those offset by 90°) will result in the perception of a depth whose sign is consistent with the direction of the physical displacement of the stimulus. However, matching the next-nearest feature (features offset by $360-90°$, i.e. 270°) will result in the precept of reverse depth: depth in a direction that is inconsistent with the physical displacement of the stimulus (see Fig. 1). A system that employs the anti-aliasing half-cycle limit would not be sensitive to the 270° phase offset (Fig. 1C) and would therefore only signal the 90° phase offset (Fig. 1B). However, a system that does not employ the half-cycle limit would, potentially, be sensitive to the larger disparity (Fig. 1C) and could therefore signal reverse depth.

2.1.3. Apparatus

Stimuli were generated using a Cambridge Research Systems VSG 2/3 graphics card in a host Pentium computer and were displayed on a custom Model 3 Vision Research Graphics monitor. The dichoptic half-images were selectively presented to each eye via the use of Vision Research Graphics ferro-electric shutters. The fast decay rate of the monitor's P46 phosphor (0.1 μs to 10% of the phosphor's initial luminance value) ensured that there was no interocular cross-talk via the shutters. The frame rate of the monitor was 120 Hz so that the effective frame rate to each eye was 60 Hz. At this frame rate there was no noticeable flicker of the images. The observer initiated each trial and responded via a button box. A chin rest was used to stabilise the observer's head.

2.2. Results

The results for the three observers are shown in Fig. 2. Performance for the three spatial-frequencies is plotted for the two spatial-extents used (15 and 30°). Disparity is plotted as a function of the phase offset of the sinewave grating. Note that performance is quantified in terms of 'percent correct'. As noted above, given the multiple depths actually present in the stimulus, labelling the responses as correct or not is arbitrary. A correct response signifies that the observer perceived depth in the direction consistent with the displacement of the stimulus (i.e. a stereo match was made with the nearest neighbour feature: 90° offset). An incorrect response signifies that a match was made with a feature in the opposite (reverse) direction (e.g. 270° offset). Error bars represent mean $\pm$ 1 S.E. The basic pattern of responses is the same for all observers and a number of trends are clear. Consistent reverse depth (depth aliasing) was obtained for all observers. The propensity of seeing reverse depth increased as both the disparity offset and spatial extent of the stimuli were increased. Additionally, for two of the observers (CS and MS) reverse depth was more likely to be seen with the lower spatial-frequency stimuli. Thus for the large spatial-extent condition (30°) disparity offsets of 90° or more resulted in consistent depth-aliasing.

It is worth noting that when these same stimulus conditions were viewed for long durations, depth consistent with a nearest-neighbour match would be perceived. Fig. 3 compares performance for short (0.14 s) and long (9 s in a raised-cosine temporal-envelope) duration stimuli. For the long-duration condition, a fixation point was always present. The spatial frequency was 0.3 cpd and the disparity offset was 90° of phase (the upper limit of the sensory fusion range; Schor, Wood & Ogawa, 1984b). This disparity corresponds to a nearest-neighbour match of 0.83°, which is well within Panum's fusional area for this spatial frequency (Schor et al., 1984b). The next neighbour for an aliased match of 270° phase offset corresponds to 2.5°, which is well beyond Panum's fusional area. Observers were highly trained and could easily maintain fixation to within 1° of the fixation point. Thus binocular eye alignment was closer to the 90° match than to the 270° one for both the short and long duration conditions. For the long duration observations, if the stimulus was presented in a rectangular temporal-envelope, the initial depth precept was in the aliased direction. However, with the continued presentation of the stimulus, this precept faded and depth consistent with the nearest-neighbour match was then perceived. The initial precept of depth aliasing could be avoided by presenting the stimulus in a raised-cosine temporal envelope. Such a stimulus does not drive the transient system, driving only the sustained system. See Pope et al. (1999) for a full discussion of this issue.

We interpret the present results as indicating that, when presented with a periodic stimulus in which multiple, potential depths are present, the sustained stereo-system matches features that are closest to each other, while the transient system does not necessarily do so.

3. General discussion

The results of the present study are that: for a particular stimulus condition, observers perceived veridical depth (i.e. depth in the direction of the nearest-neighbour match) for long stimulus-durations and reverse depth short durations; decreasing the disparity of the stimulus decreased the likelihood that reverse depth would be perceived; and increasing the spatial extent of the stimulus increased the likelihood of reverse-depth perception.

Ogle was the first to propose that stereo-perception is mediated by two distinct systems (Ogle, 1952). He labelled them the quantitative and qualitative systems. Given that we have found that the magnitude of the perceived depth generated by the 'qualitative' system

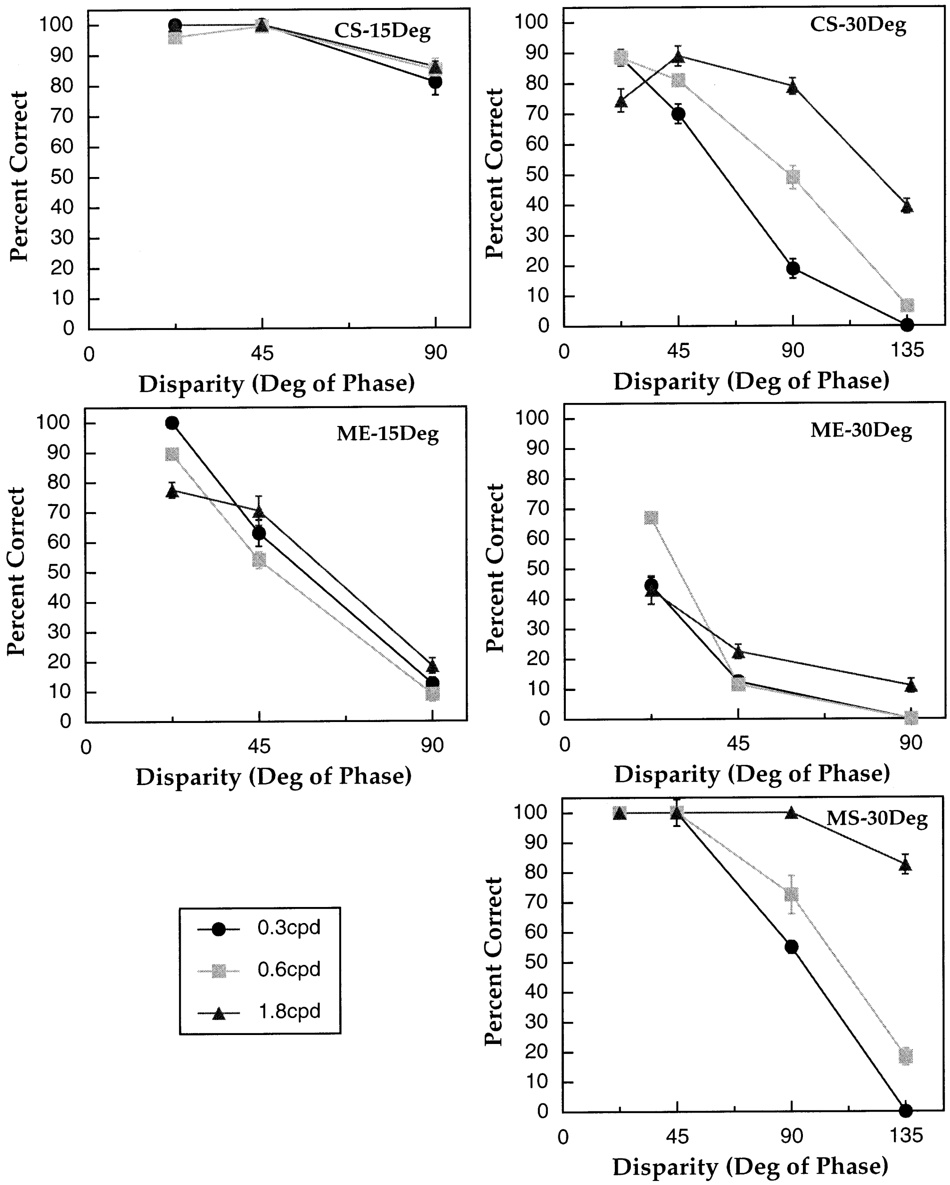

Fig. 2. Results for the main experiment. Performance, percentage of responses that were in the nearest-neighbour direction, is plotted against disparity (degrees of phase) for three spatial-frequency conditions (0.3, 0.6 and 1.8 cpd). For two of the observers (CS and ME) graphs are given for two stimulus sizes (15 and 30°) while for MS, only results for the 30° stimulus size are given.

(that is, the perceived depth of highly diplopic stimuli) can be varied (also see Richards & Kaye, 1974) and that a defining difference between these two systems appears to be their respective temporal sensitivities, we have relabelled the 'quantitative' and 'qualitative' systems the sustained and transient systems, respectively (Schor et al., 1998; Pope et al., 1999).

As noted by Ogle (1952), one of the defining differences between these two systems is the disparity range over which they operate. The quantitative system processes small disparities (those that are substantially with Panum's area) and the qualitative system can process larger, highly diplopic disparities (Ogle, 1952). While, in an earlier study, we found that the transient (quantitative) system can also process small (fusible) disparities (Pope et al., 1999) the difference in the upper-disparity limit of these systems remains an important difference between the two systems. The upper-disparity limit of the sustained system is around 1 degree (Ogle, 1952; Schor et al., 1982b) while that of the transient system is around 10° (Westheimer & Tanzman, 1956). This difference is important when considering the utility of employing a half-cycle-limit approach to depth processing. Since such an approach restricts the maximum disparity that can be processed, it would not significantly impede the sustained-system's (small-disparity) processing, while it would impede the transient-system's (large-disparity) processing. Thus it would be logical for the sustained system to employ a half-cycle-limit approach but not for the transient system to do so. We interpret the present results as supporting this notion. The propensity of observers to perceive 'reverse depth' at short, but not long stimulus durations we attribute to the transient system binocularly matching features that are separated by more than one-half the spatial period of the stimulus. In other words, the transient system is not implementing a half-cycle limit. The fact that we did not perceive reverse-depth with the long duration stimulus we attribute to the sustained system implementing half-cycle-limit processing (see Fig. 1).

If the transient system does not employ a half-cycle limit approach, then what sort of model could account for both its depth-aliasing behaviour and the finding that reducing the disparity of the stimulus reduced the likelihood of the observer perceiving reversed-depth? The model proposed by Richards (1971, see also Poggio, Motter, Squatrito & Trotter, 1985) offers one possibility. This model essentially employs two types of disparity detectors: those that are narrowly tuned to disparity (either in an excitatory and inhibitory manner), and those that are broadly tuned to disparity. The narrowly-tuned detectors are sensitive to a small disparity range that is centred around the fixation point. That is, these detectors are sensitive only to small disparities. The broadly-tuned detectors cover a wide range of either near or far disparities, so they are sensitive to large disparities (see Fig. 4).

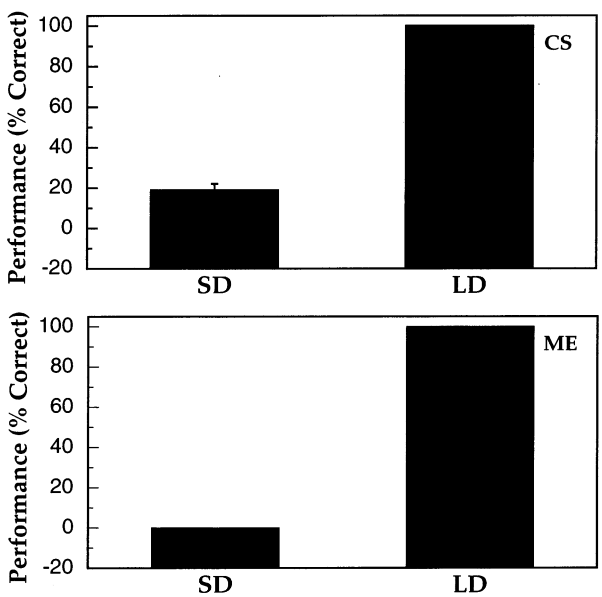

Fig. 3. Performance, percentage of responses that where in the nearest-neighbour direction, is plotted for the two stimulus conditions: short duration (SD) and long duration (LD). Both observers obtained 100% performance for the long-duration condition while performance for the short-duration condition was near 0%. That is, for the long-duration condition, observers perceived depth that was in the nearest-neighbour direction, while for the short-duration condition they perceived depth that was in the next-to-nearest-neighbour direction.

Within such a model, the narrowly-tuned detectors would implement the nearest-neighbour match, and hence signal the 'correct' depth. To account for the present results within such an approach, the disparity corresponding to phase offsets of 90 or more would have been represented more strongly by the broadly-tuned detectors than by the narrowly-tuned ones. Additionally, the response of the broadly-tuned detector/s sensitive to the 'reverse depth' (360–90°, i.e. 270°) direction would need to have been greater than that of the broadly-tuned detector/s sensitive to nearest neighbour match (90°). See Stimulus 1 in Fig. 4. As the disparity of the stimulus was decreased, the stimulus would have come within a more sensitive region of the disparity range covered by the narrowly-tuned detectors and a point would be reached where the response of these detectors was greater than that of the broadly-tuned detectors sensitive to the opposite depth. See Stimulus 2 in Fig. 4.

It is worth noting that an alternative to the above described most-active detector solution is theoretically possible. It could be argued that if any of the potential disparity matches fell within the range of the narrowly-tuned small-disparity detectors, then that depth sign would be perceived, with the outputs of the broadly-tuned detectors only being used if the narrowly-tuned detectors are not driven by the stimulus. That such a notion is incorrect, however, is demonstrated by the effect that varying the spatial extent of the stimulus had on depth perception. For a given small-disparity offset, increasing the spatial extent of the stimulus increased the likelihood that the observer would perceive reverse depth (see Fig. 2). Since the narrowly-tuned detectors would have continued to be driven as the spatial extent of the stimulus was increased, this finding shows that merely driving these detectors does not necessarily lead to the observer perceiving the direction of depth signalled by them. Note that the present results could also be accounted for by proposing that the transient system employs only broadly-tuned detectors (i.e. no narrowly tuned ones) and that the response of these detectors rolls off more quickly at large disparities than it does at small disparities. However, such a possibility is unlikely, since the largest disparity used in the present study was only about 1° and the system remains sensitive up to 10° (Westheimer & Tanzman, 1956).

As to the question of why observers are more likely to perceived reverse depth as the spatial extent of the stimulus is increased, there are at least two possible reasons. The first possibility is that there is a variation in the density of the two types of detectors (narrowly-tuned and broadly-tuned) as a function of eccentricity. Specifically, there could be proportionally more broadly-tuned detectors than narrowly-tuned ones at large eccentricities, e.g. Richards and Regan (1973)

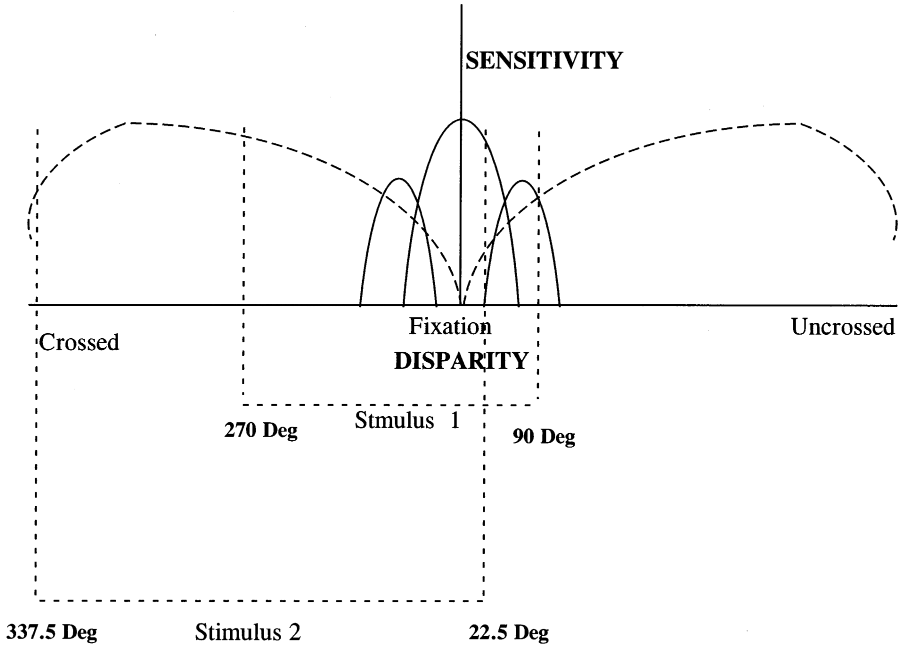

Fig. 4. Schematic representation of two hypothetical types of disparity-tuned detectors. Narrowly-tuned detectors (solid lines) are sensitive to only a small range of disparities that are near the fixation depth. Broadly-tuned detectors (dashed lines) are sensitive to a wide range of disparities. Note that this broad tuning also makes these detectors sensitive to large disparities that are outside the sensitivity range of the narrowly-tuned detectors. Relative responses to two stimuli are shown. Stimulus 1 shows the response to a 90° nearest-neighbour match and the corresponding 270° next-to-nearest-neighbour match. Note that the response of the narrowly-tuned detector is less than that of the broadly-tuned one. Stimulus 2 shows the response to a 22.5 and the corresponding 237.5° match. In this case the response of the broadly-tuned detector to the 337.5° match is greater than that of the narrowly-tuned one to the 22.5° match.

showed idiosyncratic retinotopic variations in sensitivity to large-disparity transient small-disparity sustained stimuli. Such a situation could result from the narrowly-tuned detectors being present only at small eccentricities, or due their population density being relatively constant (or decreasing) as a function of eccentricity, while that of the broadly-tuned detectors increases with increasing eccentricity. The result of such a situation would be that as the spatial extent of the stimulus was increased, proportionally more broadly-tuned cells than narrowly-tuned cells would have been activated, thus biasing the response of the system towards the depth sign carried by the broadly-tuned cells.

The second possibility is that the spatial extent of the stimulus affects perceived depth via its effect upon the location of the edges of the stimulus. It is possible that the 0° edges drive (in part) the narrowly-tuned cells sensitive to the small disparities and that the effectiveness of the edges in driving these cells decrease as their eccentricity is increased. This interaction is analogous to the adjacency principle described by Gogel (1972) in which depth solutions are biased by neighbouring stimuli. Thus, as the edges of the stimulus were moved to a higher eccentricity by increasing the spatial extent of the stimulus, the activity of the narrowly-tuned cells would have decreased relative to that of the broadly-tuned cells, thus biasing the response of the system towards the depth sign carried by the broadly-tuned cells. In a number of pilot studies, we have observed that the depth sign of the edge can influence the perceived depth of the stimulus. If the disparity of the edge in a 15° field is set to be the same as the nearest-neighbour match, then that is the depth that is perceived. However, if the edge disparity is set to equal that of the next-to-nearest-neighbour match, then that is the depth (i.e. reverse depth) perceived. The interaction between the depth carried by the carrier and the edge (envelope) is a topic of ongoing research (also see Wilcox & Hess, 1997). The edges serve to constrain depth solutions within the large range of the broadly-tuned detectors in the transient system.

Another question of interest concerns the critical disparity at which depth aliasing starts to occur? Specifically, does it purely depend upon the physical displacement of the stimulus or does the disparity interact the spatial-frequency being used. In other words, is there a disparity position or phase dependency? The use of the multiple spatial-frequencies allows us to resolve this issue. Fig. 5 presents the results with disparity measured in absolute degrees, as opposed to degrees of phase used in Fig. 2. Comparison of the two figures shows that the decrease in the likelihood of perceiving reverse-depth as the disparity of the stimulus was decreased does not depend purely upon the absolute magnitude of the disparity. Rather, the effect of a particular magnitude of disparity depends upon the spatial frequency of the stimulus, i.e. the effect is phase dependent, not absolute size dependent. The 15° data for CS and ME and the 30° data for ME clearly conform to a this phase dependency. While the pattern of results for the different spatial-frequency for the 30° condition for CS and MS are not precisely the same, when plotted against phase angle, they are, however, much closer than when plotted against positional disparity. This finding suggests that performance was being mediated by a number of channels, tuned to different spatial frequencies. Furthermore, this finding adds further support to the notion that, within the transient system, small disparities are processed by cells that are tightly tuned to a small range of disparities (in a phase-disparity linked manner) while the larger disparities are processed by cells that have no phase-disparity linking and are thus susceptible to depth aliasing.

Finally, it is interesting to note that the increased tendency to see reverse depth as the spatial extent of the stimulus was increased does not appear to be based simply upon the increase in the number of cycles present in the stimulus. In fact, for the large field (30°) conditions, all observers were least likely to perceive reverse depth in the condition that contained the most number of cycles (1.8 cpd) as compared to the other two conditions (0.3 and 0.6 cpd).

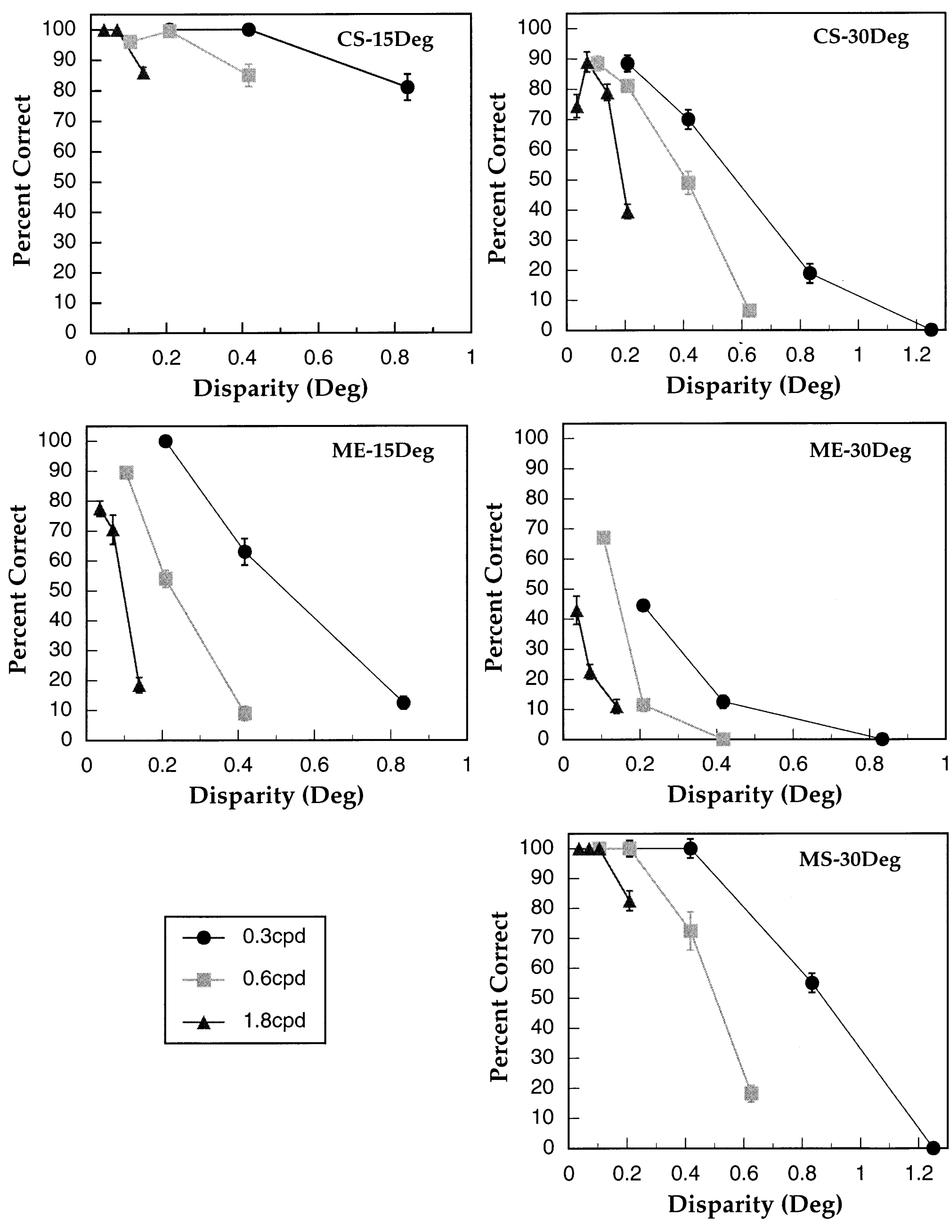

Fig. 5. The results shown in Fig. 2 are replotted here in terms of absolute degrees of disparity, as opposed to degrees of phase.

Acknowledgements

This research was supported by NEI grant EYO-8882.

References

Blake, R., & Wilson, H. R. (1991). Neural models of stereoscopic vision. *Trends in Neurosciences*, *14*(10), 445–452.

Gogel, W. C. (1972). Depth adjacency and cue effectiveness. *Journal of Experimental Psychology*, *92*(2), 176–181.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society London*, *204*, 301–328.

Ogle, K. (1952). On the limits of stereoscopic vision. *Journal of Experimental Psychology*, *44*, 253–259.

Poggio, G. F., Motter, B. C., Squatrito, S., & Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert Macaque to dynamic random-dot stereograms. *Vision Research*, *25*, 397–406.

Pope, D.R., Edwards, M., & Schor, C. M. (1999). Extraction of depth from opposite-contrast stimuli: transient system can, sustained system can't. *Vision Research*, *39* (24) 4010–4017.

Richards, W. (1971). Anomalous stereoscopic depth perception. *Journal of the Optical Society of America*, *A*, *61*, 410–414.

Richards, W., & Kaye, M. G. (1974). Local verses global stereopsis: two mechanisms? *Vision Research*, *14*, 1345–1347.

Richards, W., & Regan, D. (1973). A stereo field map with implications for disparity processing. *Investigative Ophthalmology and Visual Science*, *12*, 904–909.

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

Schor, C. S., Wood, I., & Ogawa, J. (1984a). Spatial tuning of static and dynamic local stereopsis. *Vision Research*, *24*, 573–578.

Schor, C. S., Wood, I., & Ogawa, J. (1984b). Binocular sensory fusion is limited by spatial resolution. *Vision Research*, *24*, 661–665.

Westheimer, G., & McKee, S. P. (1979). What prior processing is necessary for stereopsis. *Investigative Ophthalmology and Visual Science*, *18*(6), 614–621.

Westheimer, G, & Tanzman, I. J. (1956). Qualitative depth localization with diplopic images. *Journal of the Optical Society of America*, *46*(2), 116–117.

Wilcox, L. M., & Hess, R. F. (1997). Scale selection for second-order (non-linear) stereopsis. *Vision Research*, *37*, 2981–2992.