Evidence for surface-based processing of binocular disparity

A. Glennerster* S.P. McKee† M.D. Birch*

March 13, 2002

Summary

It is convenient to think of an object's location as a point within a Cartesian framework; the x-axis corresponds to right and left, the y-axis to up and down, and the z-axis to forward or backward. When an observer is looking straight ahead, binocular disparities provide information about distance along the z-axis from the fixation plane[1,2]. In this coordinate system, changes in disparity are treated as independent of changes in location along the orthogonal x and y axes. Does the human visual system use this three-dimensional coordinate system or does it specify feature location in a coordinate frame determined by other nearby visible

*University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK, (ag@physiol.ox.ac.uk)

†Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street San Francisco, CA 94115

features? Here we show that the sensitivity of the human stereo system is determined by the distance of points with respect to a local reference plane, rather than by the distance along the z-axis with respect to the fixation plane. There is a distinct advantage to using a local frame of reference for specifying location. It obviates the need to construct a complex three-dimensional space in either eye-centred or head-centred coordinates that must be updated with every shift of the eyes and head.

Results and discussion

Figure 1 illustrates the two hypotheses we tested. According to the first, the direction that the visual system treats as the 'depth axis' is along a line joining the fixation point and a point half way between the two eyes (often called the 'cyclopean point' since this is where the Cyclops' eye was located). This direction, shown by the white arrow, changes as the observer moves (figure 1), as does the locus of zero disparity, known as the Vieth-Müller circle and shown by the dashed line. Consequently, the disparity of points varies continually as the observer moves. The second hypothesis, which is supported by the data we present here, is that the visual system processes and represents the depth of points relative to a local surface. This type of depth representation would be much less prone to change as the observer moves.

Figure 1 about here

We used measurements of stereoacuity as a method of distinguishing these hypotheses. We compared three conditions. In each case we presented a regular grid of dots and measured subjects' sensitivity for detecting displacements of the central column of dots (see Methods). In one condition we introduced disparities to the target column by displacing each monocular half-image by an equal and opposite amount, in the traditional way. This is equivalent to a real 3-D displacement in the 'cyclopean' direction. In the other two conditions we displaced the target dots in either the right or the left eye's image leaving the other eye's image unchanged. This is equivalent to a real 3-D displacement directly towards one eye. If disparities are measured relative to the Vieth-Müller circle (or the fixation plane, which is approximately the same for the small stimuli we used) then stereo thresholds for detecting displacement of the target in these three conditions should be the same. However, if disparities are measured relative to a local surface then thresholds in these three conditions should vary systematically with the slant of the surface. Figure 2A illustrates the prediction of this 'surface' hypothesis for one direction of slant. The surface is drawn in plan view with the eyes shown artificially close to the surface. The filled circles mark the locations of three columns of the slanted grid. The prediction is that a smaller displacement should be required to reach a threshold level of performance on the detection task when the target is displaced more orthogonal to the surface (white arrow). The pattern of thresholds is predicted to reverse when the surface is slanted the other way.

Figure 2 about here

Figure 2B shows thresholds measured in two subjects for two opposite surface slants. In each case, thresholds are lowest when the target is displaced more orthogonal to the surface, as the 'surface' hypothesis predicts. For both subjects and for both slants the difference between thresholds measured for displacements towards the left and right eyes is statistically significant and in the predicted direction (one-tailed *t*-test, $p < 0.01$ in all cases). These results suggest that the influence of the slanted reference surface occurs early in stereo processing, affecting even the detectability of disparity differences between points.

Figure 3 about here

Even greater differences in depth sensitivity are predicted by the surface hypothesis when the target is displaced laterally (i.e. sideways) as well as in depth. As figure 3 shows, this is because shifting the target laterally brings it closer to the surface for one direction of slant but further away from the surface for the opposite slant. (For example, a target at the location marked by the left-most open circle is close to the reference plane in figure 3A but further away in figure 3B.) This distance from the surface is, according to the surface hypothesis, critical in determining how easily the visual system can detect a shift in the target's location.

We tested this hypothesis using a two-interval forced choice paradigm. Observers were asked to identify in which of two intervals the target column of dots was shifted from its central position (the display was static during each interval). We used a two-interval paradigm rather than the 'in front/behind' task of the first experiment in order to avoid pre-judging how subjects would perceive the displaced stimulus and whether they would use the perceived depth, lateral shift or some combination of the two to identify the displaced target column.

On separate runs we tested observers' sensitivity for detecting that the target had been shifted to each of the different locations shown in figure 3. We repeated the measurements for two opposite grid slants. The results show that increasing lateral displacements increase the detectability of the target (greater proportion correct) and that adding a disparity (filled and unfilled triangles) generally improves detectability over and above the zero disparity condition (crosses), as one would expect. There is also a clear asymmetry in the pattern of results, of the type predicted by the surface hypothesis. In those conditions when the target is more distant from the reference plane, performance is relatively good and when it is close to the reference plane performance is relatively poor. (A comparison of the 8 data points at lateral positions of $\pm$ 1 arcmin show this asymmetry most clearly.)

The predictions of the surface model are shown by the smooth curves in figure 3. In general, models of sensitivity take into account the visual system's sensitivity to one cue in isolation (e.g. disparity or lateral displacement of the target) and from these predict sensitivity to arbitrary combinations of cues. The assumption behind the model is that information from two cues are combined according to probability summation. The important element that makes this a 'surface' model is that the two cues here are displacement along, and disparity with respect to, the reference plane. We measured sensitivity to these two types of displacement in isolation and used the results to predict performance in the other, mixed-cue conditions. The model clearly predicts the direction of the asymmetries in the data. The predictions are close to the 95% confidence interval value for the data ($\chi^2 = 16.9$, 9 d.f.) for both slants ($\chi^2 = 17.1$ and 14.8 in figures 3A and b), with failures of the model tending to be underestimation of the magnitude of the asymmetry. A similar model based on disparity measured with respect to the fixation plane rather than the reference plane is much poorer at predicting the data. This is because it predicts a symmetric pattern of data with no effect of surface slant. (Equivalent χ^2 values for this model are 40.8 and 32.7 for figures 3A and 3B.)

Any model that will successfully predict these data must take into account both the disparity and the lateral position of points. One measure that does this is disparity gradient[3]. However, the disparity gradient between two points is not the same as the disparity of a point with respect to a local surface and we have not found a model based on disparity gradient that explains our data. The 'surface' model we have used is essentially the same as that proposed by Mitchison and McKee[4,5]. They investigated the correspondence rules used by the visual system when presented with ambiguous stereograms and found that the disparity of matched points is minimized with respect to an 'interpolation plane' through the surface (which for our stimulus is the plane of the grid). The salience model that Mitchison and Westheimer[6] proposed to account for the perceived depth of points is closely related. Glennerster and McKee[7] found, in addition, that the threshold for comparing the depths of two features was determined largely by their disparities with respect to a local reference plane. Thus, there is evidence that three central aspects of stereoscopic depth processing — correspondence, the magnitude of perceived depth and sensitivity to the relative depths of points — are determined by disparity with respect to a local interpolated plane. As discussed in relation to figure 1, there are obvious advantages to such a surface-based system, including the insensitivity of the representation to head movements.

Methods

Psychophysics

The stimulus was a regular, 7 by 7, 4° square grid of bright dots (55 cd/m^2, 2 arcmin width) presented on a dark background (0.4 cd/m^2) on two monitors viewed through front-silvered mirrors in a Wheatstone configuration and at a viewing distance of 2.65 m (see Andrews *et al.*[8] for details). Screen luminances were linearised and dot edges anti-aliased to allow accurate sub-pixel shifts. Exposure duration was 600 ms. A fixation marker was presented between trials. In experiment 1 (figure 2), subjects judged whether the central column was in front of or behind the plane of the grid. A method of constant stimuli was used, with 7 equally spaced disparities presented, centred on zero. Trials for three psychometric functions (displacements along the line of sight of the left, right and cyclopean eyes) were randomly interleaved in one run. The disparity gradient of the grid was $\pm$ 0.1 and constant within a run. The lateral displacements of the target were well below detection threshold, so subjects (two of the authors) could not determine from which of the three conditions any particular trial was drawn. The number of trials to be measured for each point on the psychometric function was determined in advance (200 in total, 10 per run). Data was fitted by probit[9] and the standard deviation of the fitted cumulative Gaussian taken as threshold (error bars show the standard error of this estimate).

In experiment 2 (figure 3), subjects judged in which of two intervals the central grid column was shifted (in the other interval it was always presented in the centre of the grid with zero disparity). The exposure duration was again 600 ms and the inter-stimulus interval was 500 ms. The shifted location of the central, target column in the signal interval was constant for one run of 50 trials. The different locations tested are shown in figure 3, i.e. 0 and $\pm$ 0.4 arcmin disparity; 0, $\pm$ 1 and $\pm$ 2 arcmin lateral displacement and all combinations of these displacements. Data show the result for at least 200 trails. Error bars show the standard deviation of the binomial distribution.

Model

The curves shown in figure 3 were calculated as follows. For each grid slant, we calculated (i) k_1, the detectability (d') per arcmin of target disparity when the target had no lateral displacement and (ii) k_2, the d' per arcmin of target displacement along the plane of the grid (in both cases, by measuring performance for displacements that are not shown in figure 3). The d' versus displacement plots were fitted with a straight line constrained to pass through the origin. Then, for each target position, we computed the expected d' contribution from disparity ($d'_d = k_1 d_r$, where d_r is target disparity with respect to the reference plane) and from the component of lateral displacement ($d'_l = k_2 l_r$, where l_r is target displacement along the reference plane). By probability summation, the expected detectability of the target, d'_t, is

$$d_t'^2 = d_d'^2 + d_l'^2.$$

We adjusted these d' estimates to account for cue-independent errors (as if the subject made a random response on a small proportion of trials[10]). The best fit for this error rate, λ, was computed once for the entire data set for each subject ($\lambda = 0.05$ for the data shown, with no significant change in this value when computed using a model based on disparity with respect to the fixation plane rather than the reference plane). λ is the only free parameter in the model. In figure 3, the d' predictions have been converted to proportion correct. In the symmetric, fixation plane model referred to in the text, $d'_d = k_1 d_f$ and $d'_l = k_2 l_f$, where d_f is disparity with respect to the fixation plane and l_f is lateral displacement along it.

References

1. J. J. Koenderink. Fundamentals of bicentric perspective. *Lecture notes in computer science*, 653:233–251, 1992.

2. I. P. Howard and B. J. Rogers. *Binocular Vision and Stereopsis.* OUP, Oxford, 1995.

3. P. Burt and B. Julesz. A disparity gradient limit for binocular fusion. *Science*, 208:615–17, 1981.

4. G. J. Mitchison and S. P. McKee. Interpolation in stereoscopic matching. *Nature*, 315:402–404, 1985.

5. G. J. Mitchison and S. P. McKee. The resolution of ambiguous stereoscopic matches by interpolation. *Vision Research*, 27:285–294, 1987.

6. G. J. Mitchison and G. Westheimer. The perception of depth in simple figures. *Vision Research*, 24:1063–1073, 1984.

7. A. Glennerster and S. P. McKee. Bias and sensitivity of stereo judgements in the presence of a slanted reference plane. *Vision Research*, 39:3057–3069, 1999.

8. T. J. Andrews, A. Glennerster, and A. J. Parker. Stereoacuity thresholds in the presence of a reference surface. *Vision Research*, 42:3051–3061, 2002.

9. D.J. Finney. *Probit Analysis.* CUP, Cambridge, 3rd edition, 1971.

10. F. A. Wichmann and N. J. Hill. The psychometric function I:fitting, sampling and goodness-of-fit. *Perception and Psychophysics*, 63:1293–1313, 2001.

Acknowledgements

Supported by the Royal Society, the Wellcome Trust and NEI EY06644.

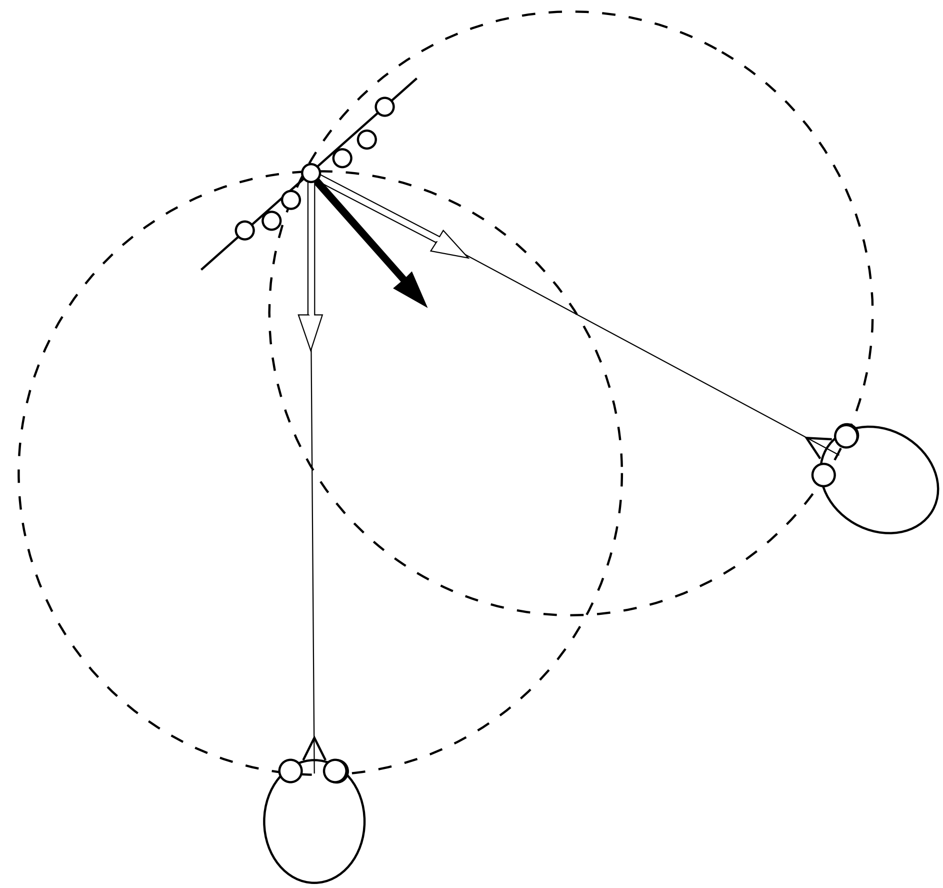

Figure 1: **Possible coordinate frames for processing disparity.** An observer, seen from above, views a surface of points (open circles). The binocular disparity of these points is commonly defined with respect to the Vieth-Müller circle (dashed line) which passes through the two eyes and the fixation point. When the observer moves, so does the Vieth-Müller circle, changing the disparity of all the points. There is, however, no change in the depth of points measured relative to the local surface (in the direction of the filled arrow).

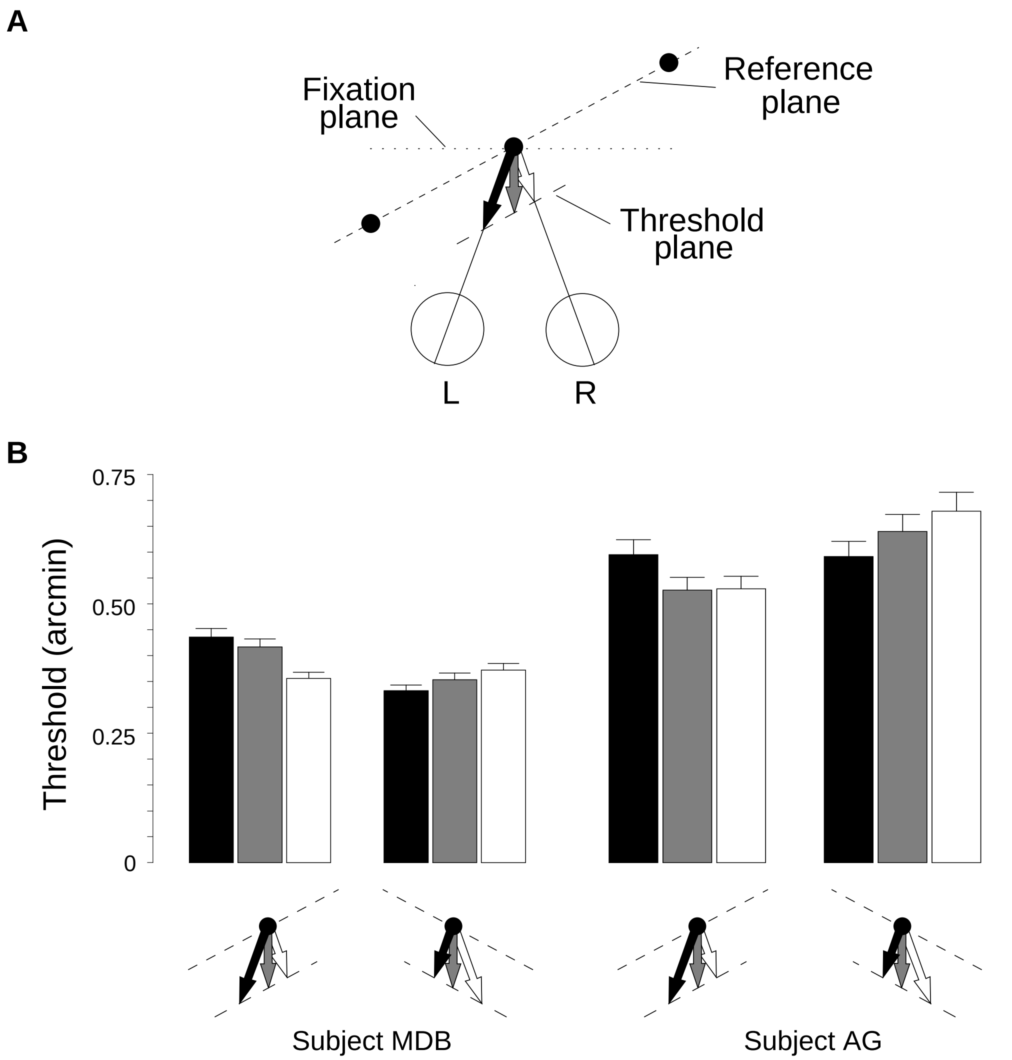

Figure 2: **Stereoacuity is affected by surface slant.** A top view of a slanted surface is shown with a reference plane connecting three visible features (A). When the central feature is shifted towards the left eye (L) the 'surface' hypothesis predicts that the threshold should be higher (black arrow) than for a shift towards the right eye (white arrow). The thresholds shown in (B) confirm this prediction in two subjects. The grey bars show thresholds measured for displacements towards or away from the 'cyclopean' eye. The pattern of thresholds is reversed for the opposite surface slant.

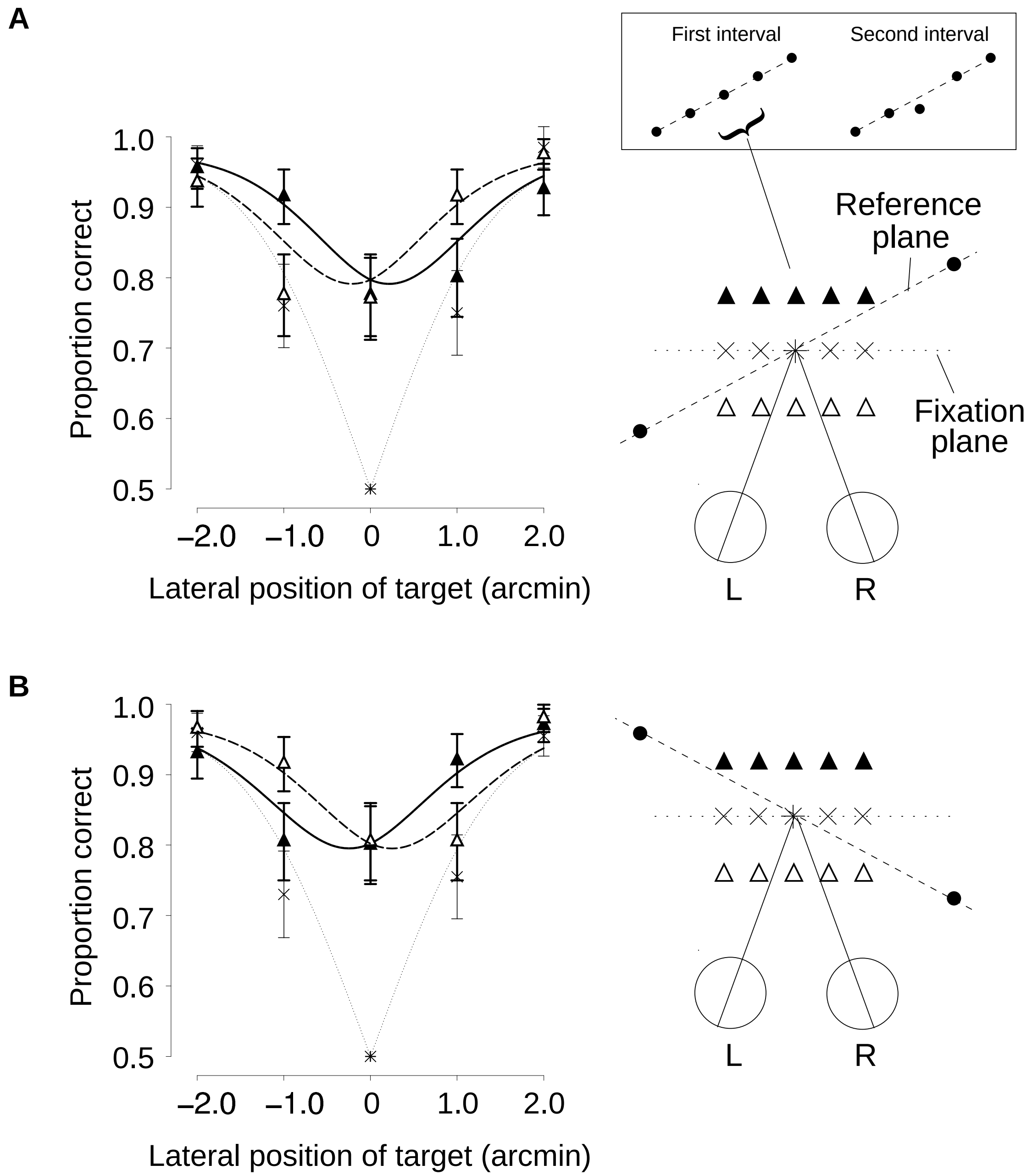

Figure 3: **Lateral displacements increase the asymmetry.** The task was to detect in which of two intervals the target had been shifted to a location away from the centre (which is marked by a star). An example of one trial is shown in the box. The diagrams on the right show the target locations we tested, for two opposite grid slants. (A) Uncrossed (i.e. far) target disparities (filled triangles) are more easily detectable than crossed (near) disparities (unfilled triangles) when the target is shifted to the left. The reverse is true for rightward displacements. (B) The pattern is reversed for the opposite grid slant (same subject). Crosses show results for target shifts in the fronto-parallel plane. Curves show the predictions of the 'surface' model (see text). Data for other subjects show a similar pattern.