Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 44 (2004) 557–562
Vision Research
www.elsevier.com/locate/visres

Combining depth cues: effects upon accuracy and speed of performance in a depth-ordering task

George Mather *, David R.R. Smith [1]

Laboratory of Experimental Psychology, Biology School, University of Sussex, Falmer, Brighton BN1 9QG, UK

Received 19 February 2003; received in revised form 29 September 2003

Abstract

Two experiments investigated how the number of available depth cues affected the speed and accuracy of depth-ordering judgements. A series of textured tiles was presented on a computer monitor, with relative depths defined by combinations of contrast, blur and interposition. Subjects were required to move a mouse pointer inside each tile in turn, starting with the tile that appeared nearest, clicking on each. Accuracy of depth-ordering was much higher than chance in all conditions, though performance using the interposition cue alone was worse than in all other conditions. The only difference in reaction time in different cue conditions was in the time elapsed before the first-click. Subjects responded substantially faster when three depth cues were present (0.84 s) than when only one depth cue was present (1.41 s). The improvement in reaction time with cue numerosity is consistent with probability summation between cues extracted by independent processes.

1. Introduction

Human depth perception is supported by a number of visual cues, ranging from binocular cues (convergence and retinal disparity) to a variety of monocular cues such as interposition and blur (see Howard & Rogers, 1995, for a review). Despite the multiplicity of cues, subjective impressions indicate that we form a single coherent estimate of the three-dimensional structure of the immediate visual environment. The multiplicity of depth cues and the apparent unity of depth perception have led researchers to ask how the information provided by different depth cues is integrated to yield a single depth estimate for each region of the visual image. The dominant view, inspired by work in computer vision (e.g. Bülthoff & Mallot, 1988; Johnston, Cumming, & Parker, 1993; Parker, Cumming, Johnston, & Hurlbert, 1995) is that perceived depth corresponds to the weighted sum of the depth values signalled by different cues. The relative importance of different cues is governed by their weights in the algebraic sum.

In a standard psychophysical technique used to draw inferences regarding relative cue weights, observers are shown a test stimulus containing multiple depth cues, which may provide inconsistent depth information. Subjects perform a depth-matching task, setting a probe to match the apparent depth of the test. Cue weights can be inferred from the way that depth settings vary with manipulation of cue values. A number of studies have provided empirical support for this general framework (e.g. Frisby, Buckley, & Horsman, 1995; Mamassian & Landy, 2001; Mather & Smith, 2000; Parker et al., 1995).

Experiments on depth cue integration typically involve tasks with extended viewing times and finely judged observations based on a rating scale or a depth match. However, in many real-world tasks requiring depth judgements observers may not have the time or inclination to make a carefully considered response. For instance, users of computer graphical interfaces may wish to navigate quickly between items arranged in different layers of a virtual desk-top, and may rely on rapid judgements based on available cues (e.g. Mori & Hayashi, 1995). In a more safety-critical context, car drivers who suddenly encounter changes in the visual environment (e.g. fog banks) may need to make rapid depth judgements in the face of marked changes in available depth cues. In this study, therefore, we sought

* Corresponding author. Tel.: +44-1273678342; fax: +44-1273678611.
E-mail address: georgem@biols.susx.ac.uk (G. Mather).
[1] Present address: Physiological Laboratory, Downing Street, University of Cambridge, Cambridge CB2 3EG, UK.

0042-6989/$ - see front matter
doi:10.1016/j.visres.2003.09.036

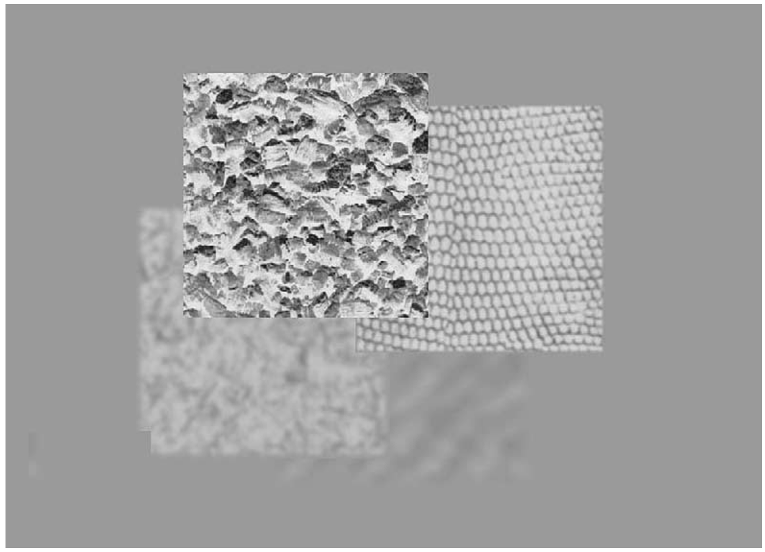

Fig. 1. Example of the texture displays used in the experiments.

to examine how multiple depth cues are employed in a task requiring rapid depth judgements.

We developed stimuli of the kind depicted in Fig. 1. The grey scale image contained four textured tiles. The apparent depth-order of the tiles was defined by a combination of (i) blur (increasing at greater depths; Mather & Smith, 2002 discuss how blur can be used to establish depth-order); (ii) contrast (decreasing at greater depths; Fry, Bridgman, & Ellerbrock, 1949, show how contrast decreases with distance due to atmospheric perspective); and (iii) interposition (nearer tiles occluding farther tiles). The example in Fig. 1 contains all three depth cues, but we also generated images containing single cues and all possible pair-wise combinations of cues. To assess the effectiveness of the images, we employed a task akin to navigating through layered windows in a graphical computer interface. Subjects were required to indicate the apparent depth-ordering of the tiles by moving the mouse pointer inside each tile in turn (starting with the nearest) and clicking once on each. The computer recorded errors in reported depth-ordering, and the time taken to register each click. The aim of the experiment was to determine how different cues, and different combinations of cues, affected observers' speed and accuracy in assigning depth-order.

2. Experiment 1

2.1. Methods

2.1.1. Subjects

Five observers took part in the first experiment, one author and four others näive to the purpose of the study. All observers were experienced in making judgements in psychophysical experiments. The display was viewed binocularly without head restraint and with natural pupils. Appropriate optical corrections were worn. Observers fixated a central fixation mark. The room was kept dark with the only source of illumination coming from the display.

2.1.2. Apparatus

Stimuli were displayed on a Silicon Graphics O2™ workstation and displayed on a GDM-17E21 colour graphic display. The frame rate was 75 Hz, with a horizontal line frequency of 79.8 kHz. The viewing distance was 114 cm. The display area of the monitor subtended 14.51° wide by 12.26° high. Each display pixel subtended 41 s arc.

The minimum and maximum luminance attainable on the monitor was 0.01 and 65.92 cd/m^2, respectively. Luminances were measured using a Minolta LS-100 photometer. The monitor was linearised by inverting ($r^2 = 0.998$) a 3rd-order polynomial fitted to the calibration data. A gamma-correcting lookup table was used to ensure stimulus linearity.

2.1.3. Stimuli and design

Stimuli contained image-processed versions of natural Brodatz textures (Broadatz, 1966). Each original Brodatz image was digitized to produce a 512×512 pixel image with 256 grey-levels. Our images were based on D3, D5, D18 and D110 (where for instance D3 refers to the photographed texture on page number 3 of Broadatz, 1966). Stimulus displays were generated by taking a pseudo-randomly determined square portion (256×256 pixels) of each Brodatz texture, and arranging the patches appropriately against a uniform grey background (33 cd/m^2). There were arbitrary differences in the mean luminance and distribution of grey-levels between the textures. These differences were removed prior to experimental manipulation so that all textures had the same Gaussian-like distribution of grey-levels centred on the same mean luminance (approximately 34 cd/m^2). The RMS and standard deviation of the grey-levels of the textures were the same. Each texture patch subtended an angle of $2.92° \times 2.92°$ at the viewing distance of 114 cm (this varied somewhat in that blurred stimuli were slightly larger because of the bleeding of the texture at the edges into the background, e.g. a stimulus blurred with a Gaussian blurring function with a space constant of 4 min arc, subtended an angle of $3.15° \times 3.15°$). Reductions in contrast due to spatial blurring were compensated by histogram-equalisation. [2]

Seven experimental conditions were defined in terms of the presence of one or more of three depth cues in the four texture patches: contrast, blur and interposition. Conditions were: contrast (C), blur (B), interposition (I), contrast + blur (C + B), blur + interposition (B + I), contrast + interposition (C + I), and contrast + blur + interposition (C + B + I). Details of the depth cues are as follows.

Contrast: Each texture was linked to a lookup table which allowed independent manipulation of Michelson contrast (C_M). The C_M level was set at 100%, 75%, 50% and 25% $(((L_{max} - L_{min})/(L_{min} + L_{max})) \cdot 100$ where L_{max} and L_{min} are the maximum and minimum luminances in cd/m^2 present in the stimulus).

Blur: The stimulus textures could be convolved by a two-dimensional separable Gaussian kernel with a space constant of 0 (sharp), 1, 2 and 4 min arc.

Interposition: This was achieved by laying each stimulus so that it partially occluded the stimulus below it. The amount by which each stimulus occluded the one below it was randomised and constrained so that all stimuli were clearly visible.

Each differently textured square was pseudo-randomly assigned to different levels of experimental manipulation. This reduces any biases that different textures might have towards being placed at particular points in a depth-ordering task. The combinations of cues were constrained so that the depth-orderings conveyed by each cue were in agreement, i.e. it is assumed that higher-contrast stimuli are perceived to be nearer than lower contrast stimuli (O'Shea, Govan, & Sekuler, 1997), that sharp textured stimuli are nearer than blurred stimuli (Mather & Smith, 2002) and that occluding stimuli are nearer than occluded stimuli. Thus in an experimental condition having three depth cues, the nearest textured stimulus was positioned on top of all other stimuli (interposition cue) and had a spatially unblurred texture with a contrast of 100% C_M. The farthest stimulus contained texture spatially blurred with a 4 min arc space constant at a contrast of 25% C_M.

2.1.4. Procedure

Each trial was initiated by pressing a mouse button. The stimulus display appeared and the mouse cursor immediately moved to the centre of the display. The display remained visible until the observer had finished their response. A central fixation mark was provided. The observer's task was to indicate the order of depth in which the texture patches appeared to lie. This was achieved by clicking on each patch in turn from nearest to furthest in perceived depth. The time from stimulus onset to each click was recorded. Observers were not given feedback as to the correctness or otherwise of their responses. In between trials the display was reset to a uniform mean luminance of 33 cd/m^2.

The stimulus condition displayed on each trial was selected pseudo-randomly from the set of seven available, with the constraint that no experimental condition would be presented for the $(n+1)$th time until all experimental conditions had been presented since the nth presentation. Each experimental condition was displayed ten times per experimental run. A computer controlled the selection of experimental conditions and recorded the responses. Data for each observer was pooled from two (usually) consecutive experimental runs providing twenty observations per experimental condition.

2.2. Results and discussion

Two main performance measures were derived from the data: (1) percentage of trials in each condition in which the subject reported the correct depth-ordering; (2) time after stimulus presentation at which the subject clicked on each square, for correct trials only. Fig. 2(a) and (b) show the means obtained using these two measures.

Percentage correct: Since there were 24 permutations of depth-ordering, the probability of reporting the correct order by chance was 0.0417 or 4.2%. Fig. 2(a) shows that responses in all stimulus conditions were far above chance level. The most obvious feature of the data is the relatively low percentage correct for the interposition cue compared to all other cues and cue combinations (37%; SE ±8.15%). To investigate this further, we calculated the percentage of errors made at each depth position in each stimulus condition. Results are shown in Fig. 3.

[2] Blurring the Brodatz texture with a given Gaussian blurring function necessarily reduces its luminance contrast. This can be quantified as a reduction in the standard deviation of the Gaussian-like grey-level distribution of the image. To restore the contrast of a blurred image we re-scaled all its grey-levels to widen its grey-level distribution, and match the distribution of the original Brodatz texture. The removes the potential confound of blurring and contrast variation.

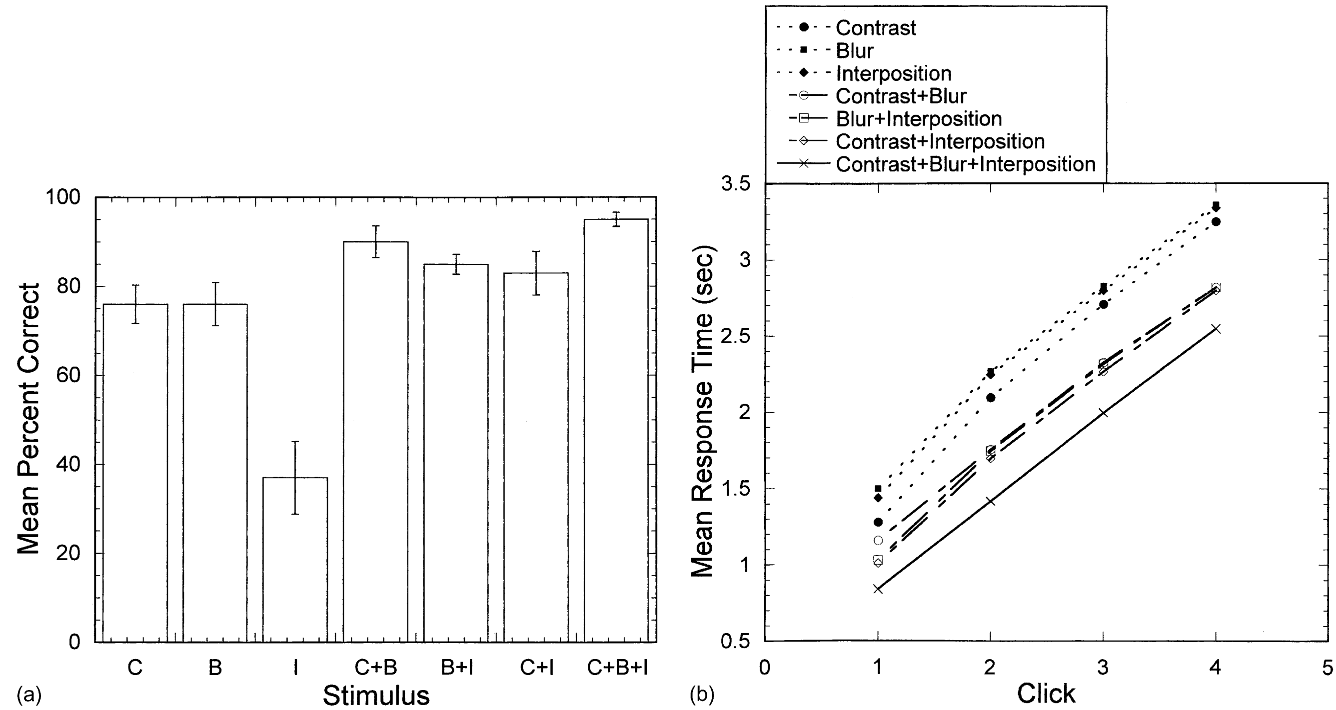

Fig. 2. Results of Experiment 1. (a) Mean percentage of correct depth-ordering responses for each of the seven stimulus conditions. Vertical bars represent SE of the mean. (b) Mean response time in seconds, as a function of stimulus condition and click number. This data was restricted to trials in which the subject reported the correct depth-order. SEs have been omitted for clarity, but were on average 0.283 s.

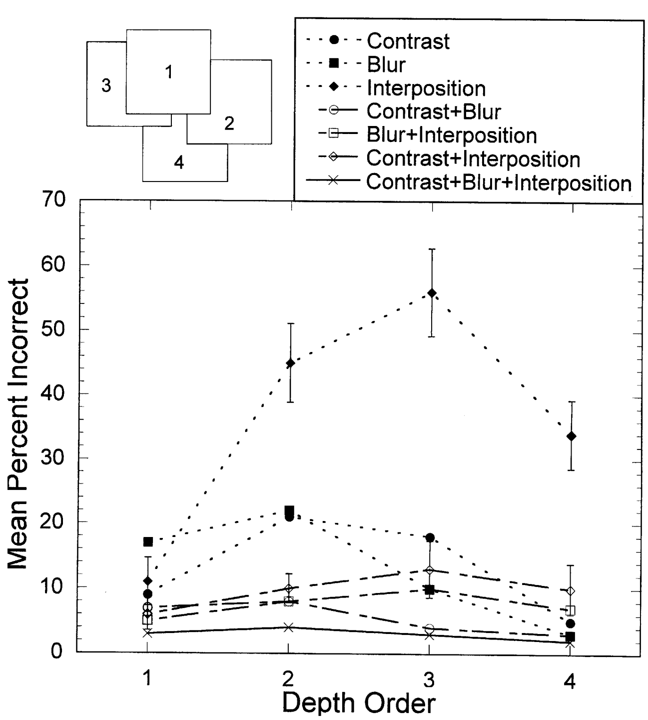

Fig. 3. Mean percentage of *in*correct responses to each tile as a function of stimulus condition and tile position in depth. As shown in the inset at top-left, tiles at position 1 were depicted as nearest the viewer, and tiles at position 4 were depicted as farthest from the viewer.

Conditions containing two or three cues yielded very few errors at any position. It is clear that the high number of errors that occurred using the single interposition cue arose most often in the middle-depth positions (2 and 3). This may reflect a possible limitation in the availability of the interposition cue relative to the other cues. The inset of Fig. 3 provides an example of the limitation. Four tiles are labelled in depth-order from nearest (1) to farthest (4). On the basis of interposition alone, the ordering of tiles (2) and (3) in this particular arrangement is ambiguous, since they have the same pattern of T-junctions. Arrangements in which the two tiles partially overlapped would obviously not suffer from this ambiguity. Tile arrangement varied randomly from trial to trial, so the ambiguity shown in Fig. 3 (inset) would not be present in every trial. On the basis of our stimulus dimensions, we calculated that 28% of trials in the interposition condition suffered the ambiguity shown in Fig. 3. If observers were incorrect in half of these trials, the ambiguity would account for an error rate of 14%, much lower than that actually obtained. The ambiguity cannot therefore account for all the errors recorded in the interposition condition.

Click time: Fig. 2(b) shows that the only differences in reaction time between conditions reflect the time elapsed before the first-click is executed. The time interval between later clicks is constant both within and across conditions, as shown by the straight and parallel lines in the graph. It is also clear that the time elapsed before the first-click depends on the number of cues present in the stimulus. Reaction time is slowest for stimuli containing single cues, and fastest for stimuli containing all three cues. The reduction in reaction time is substantial, from 1.41 to 0.84 s.

3. Experiment 2

The textured tiles used in the first experiment were normalised for luminance before the application of depth cue manipulations, even though the original Brodatz textures had different luminance values. The removal of natural variations in texture luminance may have had an undesirable effect on depth judgements. For example the edges between luminance-matched tiles do not contain any differences in mean luminance, only differences in second-order textural properties. We therefore repeated the first experiment using stimuli which retained their original luminance values: D3 (42.14 cd/m^2), D5 (35.2 cd/m^2), D18 (27.72 cd/m^2) and D110 (45.91 cd/m^2). All other experimental details were the same as in the previous experiment. Five observers took part (the second author and four experimentally-naïve others), two of whom had also participated in the first experiment.

3.1. Results and discussion

Fig. 4(a) shows data on the percentage of correct depth-ordering in each cue condition.

The contrast and blur cue conditions show worse performance than in the first experiment, as shown by the lower percentages correct in Fig. 4(a) relative to Fig. 2(a). Interposition was unaffected. It seems that introducing arbitrary variations in the mean luminance of the textured tiles made it more difficult to isolate the information provided by contrast and blur variation.

Fig. 4(b) plots mean first-click times in the two experiments averaged across conditions containing one, two, or three cues. Both experiments show a fall in RT as cue numerosity increases, though the effect was larger in Experiment 1 than in Experiment 2. Subjects viewed stimuli binocularly without head restraint. As a result, stereo and motion parallax depth cues were available that conflicted with the cues specified in the stimuli. To test whether removal of this cue conflict would substantially alter the results, one naïve subject who performed in Experiment 2 repeated the experiment with monocular viewing and a chin rest. [3] These supplementary observations are shown in Fig. 5, along with

[3] We are grateful to two anonymous referees, who suggested these observations.

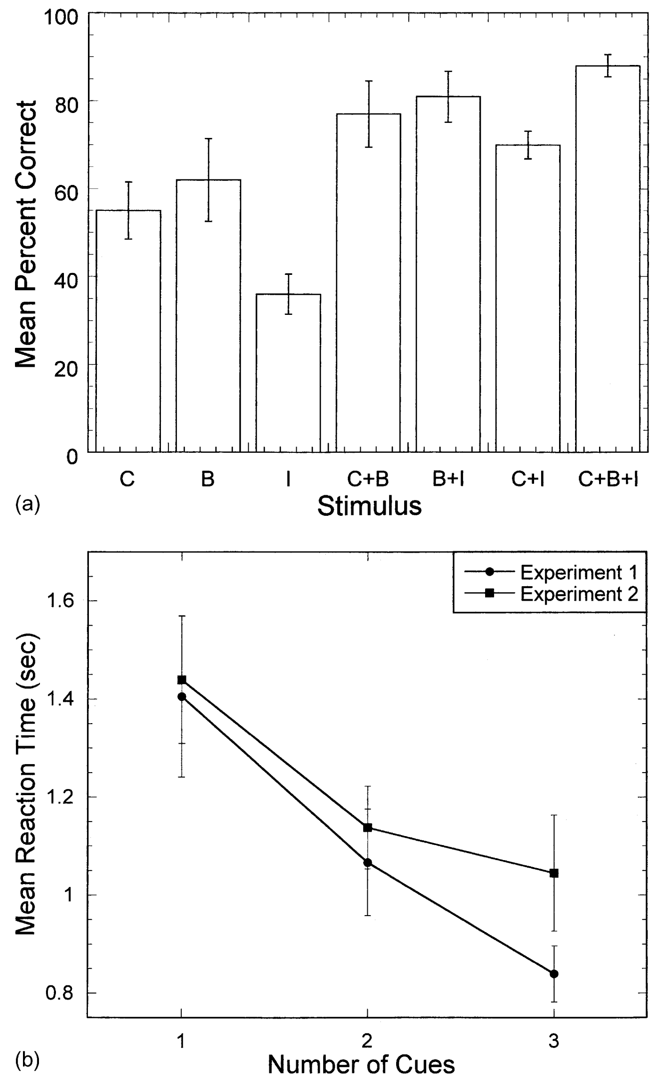

Fig. 4. (a) Results of Experiment 2, showing mean percentage of correct depth-ordering responses for each of the seven stimulus conditions. Vertical bars represent SE of the mean. (b) Mean first-click times obtained in both experiments. Correct reaction times were collapsed across the seven stimulus conditions into three values according to whether each condition presented 1, 2, or 3 depth cues. Circles show data from Experiment 1, and squares show data from Experiment 2.

results for the same subject in Experiment 2. All response times are shorter, perhaps reflecting either a practice effect or the removal of depth cue conflicts between stereo and motion parallax cues and other depth cues. But the dependence of response time on cue numerosity was still obtained.

The decrease in response time with cue numerosity must reflect some form of facilitation created by the presence of multiple cues. Raab (1962) studied the effect of presenting multi-modal (visual and auditory) stimuli on simple reaction time. He developed a statistical model of probability summation to explain the improvement in response time found when both modalities are presented together rather than singly. The

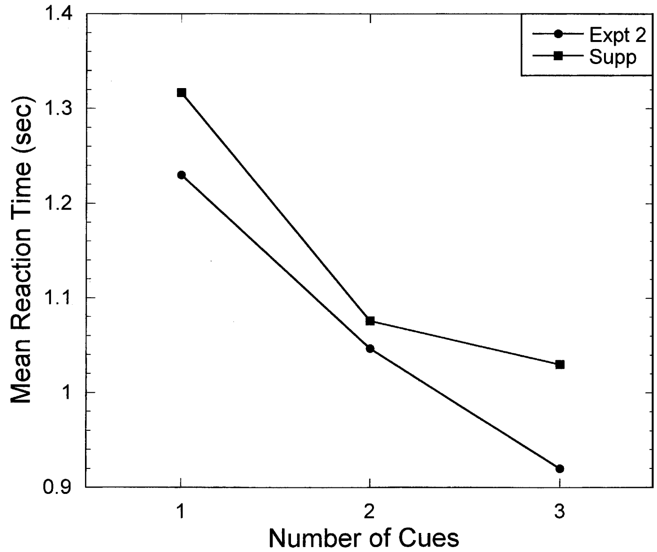

Fig. 5. Data from supplementary observations on a näive observer, who repeated Experiment 2 with monocular viewing and a chin rest to remove stereo and motion parallax cues.

improvement in response time shown in Fig. 4 is consistent with the improvements predicted by Raab (1962).

4. Conclusions

We found that increasing the number of depth cues present in a depth-ordering task led to marked improvements in both accuracy and speed of performance. The improvement in both speed and accuracy with cue numerosity shows that there is no speed-accuracy trade off. When observers are required to make rapid depth-ordering judgements in the presence of varying numbers of cues, differences in RT due to cue numerosity are confined to the time required before an initial response is made. Observers are substantially faster to initiate a response when more depth cues are present. There are practical implications of this research. When an observer is required to make rapid depth judgements in complex scenes, reactions will be faster (and more accurate) when the number of available cues is higher. This has obvious implications in assisting users to navigate between the different layers/windows of a virtual desk-top where order of importance/priority can be manipulated using multiple depth cues.

Acknowledgements

This research was supported by a grant from the Engineering and Physical Sciences Research Council, UK.

References

Broadatz, P. (1966). *Textures: A photographic album for artists and designers*. New York: Dover.

Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America A, 5*, 1749–1758.

Frisby, J., Buckley, D., & Horsman, J. (1995). Integration of stereo, texture, and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception, 24*, 181–198.

Fry, G. A., Bridgman, C. S., & Ellerbrock, V. J. (1949). The effect of atmospheric scattering on binocular depth perception. *American Journal of Optometry, 26*, 9–15.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford: Oxford University Press.

Johnston, E., Cumming, B., & Parker, A. (1993). Integration of depth modules: Stereopsis and texture. *Vision Research, 33*, 813–826.

Mamassian, P., & Landy, M. S. (2001). Interaction of visual prior constraints. *Vision Research, 41*, 2653–2688.

Mather, G., & Smith, D. R. R. (2000). Depth cue integration: Stereopsis and image blur. *Vision Research, 40*, 3501–3506.

Mather, G., & Smith, D. R. R. (2002). Blur discrimination and its relation to blur-mediated depth perception. *Perception, 31*, 1211–1219.

Mori, H., & Hayashi, Y. (1995). Visual interference with users' tasks on multiwindow systems. *International Journal of Human–Computer Interaction, 7*, 329–340.

O'Shea, R., Govan, D., & Sekuler, R. (1997). Blur and contrast as pictorial depth cues. *Perception, 26*, 599–612.

Parker, A., Cumming, B., Johnston, E., & Hurlbert, A. (1995). Multiple cues for three-dimensional shape. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 351–364). Mass: MIT Press.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences, 24*, 574–590.