Exp Brain Res (2003) 148:196–201
DOI 10.1007/s00221-002-1295-2

RESEARCH ARTICLE

Paul B. Hibbard · Mark F. Bradshaw

Reaching for virtual objects: binocular disparity and the control of prehension

Received: 6 April 2002 / Accepted: 23 September 2002 / Published online: 19 November 2002

Abstract Although, in principle, binocular cues provide veridical information about the three-dimensional shape of objects, our perception on the basis of these cues is distorted systematically. The consequences of these distortions may be less serious than they first appear, however, since in everyday life we rarely are required to judge the absolute shape, size or distance of objects. An important exception to this is in the control of prehension, where veridical information about an object to be grasped is required to plan the transport of the hand and to select the most appropriate grip. Here we investigate whether binocular cues provide accurate depth information for the control of prehension using disparity-defined, virtual objects and report that whilst binocular disparity can support prehensile movements, the kinematic indices, which reflect distance-reached and perceived size, show clear biases. These results suggest that accurate metric depth information for the control of prehension is not available from binocular cues in isolation.

Keywords Prehension · Binocular disparity · Distance perception · Size perception

Introduction

We reach for and grasp objects many times during the day. The apparent simplicity of this task, however, belies the complexity of the computational problem faced by the visuo-motor system, which must determine the distance, size and shape of the target object, select the appropriate motor program to transport the hand and pre-configure the fingers and, finally, guide and refine the movement to its completion (Jeannerod 1988). Although there is, typically, a range of information sources available to the visuomotor system, the prevalent view is that binocular cues (binocular disparity and angle of convergence) are paramount in the specification of distance, size and three-dimensional shape for the control of prehension. Several convergent lines of evidence support this idea. First, unlike many other visual cues, binocular cues, in principle, can afford a full reconstruction of the three-dimensional metric structure of the scene. This information is provided primarily by the horizontal component of binocular disparity which must be "scaled" by taking the viewing geometry (i.e. the orientation of the two eyes relative to the head) into account. Second, Marotta et al. (1997) have found that patient DF, who continued to reach for objects successfully despite profound visual form agnosia, could not form appropriate grip apertures when deprived of binocular cues. Finally, the kinematic indices of monocular reaches made by subjects with normal vision reveal that reaching performance is affected by the lack of binocular information (Servos et al. 1992; Jackson et al. 1997) with the grasp component being particularly affected (Watt and Bradshaw 2000).

It is typically assumed that the goal of human stereopsis is the recovery of metric scene structure, requiring both the measurement and scaling of binocular disparity. The degree to which binocular cues support veridical judgments of an object's properties, however, is questionable (e.g. Johnston 1991; Todd et al. 1995; Bradshaw et al. 1996; Glennerster et al. 1996; Bradshaw et al. 1998, 2000). The general conclusion of these experiments is that depth constancy is considerably less than perfect, perceived shape is distorted and absolute distance is misestimated. For example, objects appear larger and relatively stretched in depth when presented at near distances and smaller and relatively compressed in depth when presented at far distances (Bradshaw et al. 1998; Brenner and van Damme 1999; Johnston 1991). These distortions of size and shape are consistent with perceptual misestimations of distance used in the scaling

---

P.B. Hibbard (✉)
School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JU, UK
e-mail: pbh2@st-andrews.ac.uk
Tel.: +44-1334-462053
Fax: +44-1334-463042

P.B. Hibbard · M.F. Bradshaw
Department of Psychology, University of Surrey, Guildford, GU2 7XH, UK

of disparity information, which is overestimated at distances of less than around 1 m and underestimated at further distances (Foley 1980).

These somewhat surprising psychophysical results may have limited behavioural significance because there are many tasks, undertaken in everyday life, which do not require the recovery of veridical depth structure (Glennerster et al. 1996; Bradshaw et al. 2000). Indeed, tasks requiring the recovery of the full metric structure of the scene are relatively rare. The computational expense involved in measuring and calibrating disparity information to recover metric structure accurately may therefore only be undertaken for tasks such as prehension, for which it is necessary. This view is consistent with the emerging notion of task-dependent processing (Glennerster et al. 1996; Bradshaw et al. 2000). One prevalent everyday task that does require veridical information about objects in the world is prehension. To select successfully the appropriate motor programs to transport the hand to the correct location and to pre-configure the fingers to form an appropriate grip, precise information about distance and size is required. Indeed, in distinguishing between perception and action systems Milner and Goodale (1995) have speculated that the action system is characterised by its recovery of metric information about the world (see also Aglioti et al.1995).

The goal of the current study, therefore, was to establish whether binocular information can be exploited to provide veridical information about the distance, size and shape of objects when they are to be grasped in a natural prehension task. Previous studies investigating the role of binocular cues in the control of prehension typically have failed to isolate disparity cues, as they have used real objects in real scenes. In the current study, virtual, disparity-defined objects were used to assess the role of binocular disparity in the control of prehension. In any study of natural prehensile movements, the participant usually reaches for, and grasps, the target object successfully and so a kinematic analysis is required to reveal the effects of the experimental manipulations. In the present study we therefore determined the peak wrist velocity and peak grip aperture, which are indirect indices of perceived distance and perceived size (e.g. Watt and Bradshaw 2000; Jeannerod 1984; Jeannerod and Decety 1990). To relate these indices to the physical dimensions of distance and size (and so we can make explicit comparisons with the perceptual tasks used in other studies) we also included real objects that are specified by the full range of visual cues. Real objects were grasped in near-identical experimental conditions as those for the virtual, disparity-defined objects. Results from this condition could then be used as a reference to compute a “notional distance” and “notional size” based on the relative velocities and grip-apertures exhibited for the real and virtual objects.

It is important at this stage to emphasise that, although the current study was clearly inspired by perceptual studies showing biases in the perception of three-dimensional size and shape from binocular disparity, the goal was not to make a direct comparison between performance in the two domains. Rather, the aim was to establish the extent to which binocular cues provide accurate metric depth information for the control of prehension.

Methods

Participants

Nine right-handed adults completed the experiment. All participants had normal or corrected-to-normal vision and had stereo acuity scores <40 arc s (Randot stereo-test, Stereoptical, Chicago, Ill., USA). Participation was voluntary, and the experiments were performed in accordance with appropriate ethical standards of the ethics committee of the University of Surrey. All participants gave their informed consent to their inclusion in the study.

Apparatus and stimuli

Participants sat at a matt black table with their head position maintained using a chin rest. Eye-level was fixed at 17 cm above the tabletop. The start point for each trial was a 2-cm diameter start button mounted on the table top along the body midline. The task was to reach out and pick up wooden objects that were placed on the tabletop. The objects were three elliptical cylinders with a height of 9.0 cm, and diameters of 3.2×5.0, 5.0×5.0 and 7.4×5.0 cm. Objects were placed with their shorter diameter either along or orthogonal to the line of sight, giving three object widths and three object depths. Objects were placed at two distances (30 and 50 cm).

In the real object viewing condition, objects were illuminated by a desk lamp in an otherwise dark room and viewed through a semi-silvered mirror set at an angle of 45° to the median plane (Fig. 1). The objects were painted black and covered in randomly positioned, white blobs with a diameter of ~3 mm and a density of 1 dot/cm^2.

Virtual objects were defined by presenting random dot stereograms on a 19” flat-screen computer monitor positioned 46.7 cm from the observer, orthogonal to the body midline, and viewed through the semi-silvered mirror (Fig. 1). The resolution of the monitor was 800×600 pixels, and the refresh rate was 120 Hz. The left and right eye’s images were presented separately using CrystalEyes LCD shutter-glasses. Cylinders were defined by Gaussian blobs, with a standard deviation of 1 mm, placed on the surface of the virtual cylinder with a density of 1 dot/cm^2. The position of each dot in the left and right eye’s image was determined using a standard ray-tracing technique. The cylinders

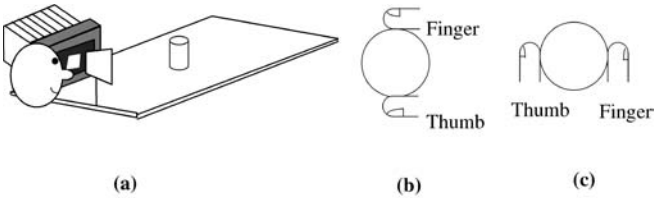


Fig. 1 **a** Participants sat at a tabletop which they viewed through a semi-silvered mirror that was placed at their eye-height, at an angle of 45° to their body midline. In separate blocks of trials, participants either viewed real objects presented on the table top, or an occluder was placed behind the mirror and participants viewed virtual objects, presented on a computer monitor, that were reflected in the mirror. Participants were asked to grasp the objects with their thumb and forefinger, grasping either **b** the front and back of the object or **c** the left and right sides of the object

were viewed for 2 s, before an audible beep was heard, and the images were removed from the computer screen as soon as the participant's hand was moved. A real wooden cylinder was placed on the table (but could not be seen) so that appropriate haptic feedback was available when the cylinders were picked up.

Three spherical infrared reflective markers were attached to the thumbnail, the nail of the forefinger and the head of the radius of the wrist of the right hand. Positions of the wrist, forefinger and thumb markers were recorded by a three-camera Macreflex motion analysis system operating at 120 Hz. The accuracy of the Macreflex system was assessed using a procedure based on that of Haggard and Wing (1990). The standard deviation of measurements was found to be <0.3 mm.

Design and procedure

Two blocks of trials were run in which participants viewed either the real objects or virtual, binocularly defined objects. Each block consisted of 36 trials (two distances×three widths×three repetitions, plus two distances×three depths×three repetitions). Observers were instructed to pick up the objects with the thumb and forefinger of their right hand, grasping either the left and right side of the object or the front and back of the object, as instructed on each trial. Objects were viewed for 2 s, after which time a short beep was heard. Participants reached out and picked up the objects as soon as they heard the beep. When the participant's hand moved off the start switch, the desk lamp was switched off, so that all reaches were performed open loop, with no visual feedback. The binocularly defined virtual object condition was performed in a dark room and an occluder was positioned behind the semi-silvered mirror so that the real objects could not be seen.

Data analysis

The position of the wrist, thumb and index finger were recorded by the Macreflex system at a sampling rate of 120 Hz. These data were filtered using a zero-phase filtering algorithm with a cut-off frequency of 12 Hz (Oppenheim and Shafer 1989), and the peak velocity of motion of the wrist, and the peak grip aperture (greatest separation between the thumb and index finger) were derived. These kinematic indices were chosen as they have been shown in a number of previous studies to scale with the distance and size of objects, respectively (e.g. Jeannerod 1988; Watt and Bradshaw 2000).

Results

Individual mean values were calculated for each object by distance combination for both conditions.

Figure 2a depicts peak velocity elicited for the real and virtual objects at both viewing distances. The data were entered into 2×2×3×2 [stimulus type×object distance×object size×direction of grasp (width or depth)] repeated-measures ANOVA. A significant main effect of viewing distance was found ($F_{1,8}$=98.0; P<0.001), but there was no significant main effect of stimulus type ($F_{1,8}$=4.4; NS). Thus, participants' wrist velocity scaled with object distance and there was no significant tendency for reaches to be faster or slower overall for the disparity defined stimuli. Clearly, subjects reached more slowly for the virtual objects placed at the furthest distance, which was reflected in a significant interaction ($F_{1,8}$=15.2, P<0.01) indicating that wrist velocity scaled differently with object distance for real and disparity-defined stimuli.

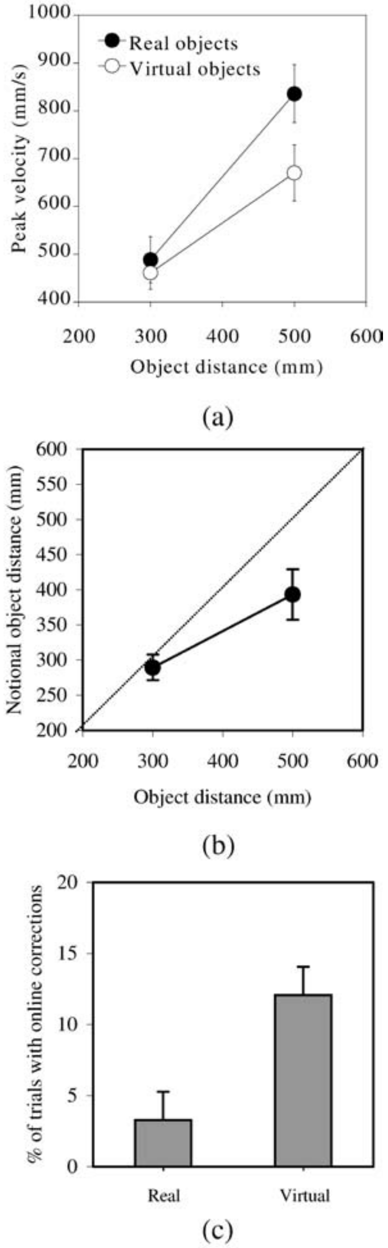


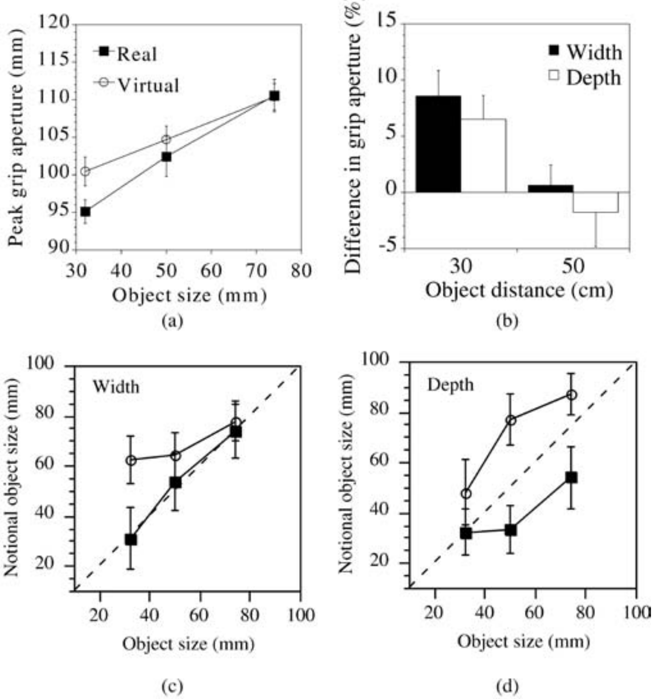
Fig. 2 a Peak wrist velocity (here averaged over the different object sizes and grasps across the widths and depths of objects) scaled with object distance for both real and virtual objects. **b** Compared with reaches made to physically viewed objects, peak wrist velocities were consistent with an underestimation of the further object distance under virtual viewing conditions, as shown by the calculated notional object distance (see text for details) **c** This conclusion is supported by the fact that participants made significantly more corrections to their prehensile movements under virtual viewing conditions, again consistent with a misestimation of object location. Means±SEM

This interaction arose because peak wrist velocity scaled less with object distance for virtual objects than for real objects, resulting from the fact that participants reached more slowly for disparity-defined objects at the further distance. Peak velocity was also faster when reaching across the depth of the object than across its width ($F_{1,8}=19.7$; $P<0.005$). A significant interaction between orientation and object size was also observed ($F_{2,16}=3.9$; $P<0.05$), resulting from a decrease in peak velocity with increasing object size when reaching across the width but not the depth of objects.

As introduced above, the peak velocities exhibited in reaches to virtual objects can be expressed in terms of a notional object distance by indexing them relative to velocities exhibited in response to real objects presented under "full-cue" conditions. This was achieved by performing a regression of peak velocity onto object size for the real objects, and using the resulting equation to convert peak velocities into corresponding notional-object distances. Separate regressions were performed for each object size, and for reaches across the width and depth of objects. Notional object distance therefore refers to the distance at which a real object would need to be placed such that it was reached for with the same peak velocity as the virtual object. Expressed in this way (Fig. 2b) the results suggest that the distance to the virtual objects placed at 50 cm was underestimated, relative to the a real object placed at this distance, by ≈20%. This interpretation is supported by an analysis of the number of trials (see Fig. 2c) containing online corrections [i.e. trials containing more than one peak in both wrist velocity and grip aperture, which suggests that the initial preprogrammed reach was to the wrong location (Marrotta and Goodale 1998)]. Errors reflected in this measure are evident to the participants through a lack of haptic feedback (i.e. they fail to find an object in the place where it is expected, and so make another attempt to find the object). Significantly more corrections were made for the virtual objects than for the real objects ($t_{1,8}=3.97$; $P<0.01$).

Figure 3 shows the results based on the peak-grip aperture exhibited when participants reached for the target object. The scaling of peak grip aperture to take account of object size was analysed using a 2×2×3×2 (stimulus type×object distance×size of object×direction of grasp) repeated-measures ANOVA. Grip apertures scaled with object size for both the real and virtual objects (Fig. 3a). A significant main effect of object size ($F_{2,16}=35.7$; $P<0.001$) was found, demonstrating that grip aperture increased significantly as object size increased. No other main effects were observed. A significant interaction was found between stimulus type and object distance ($F_{1,8}=10.0$; $P<0.05$), indicating that peak grip aperture was differently affected by viewing distance for the real and virtual objects. Specifically, grip apertures for virtual objects were larger than those for real objects at the closer object distance, but approximately the same at the far viewing distance as illustrated in Fig. 3b. A significant interaction between condition and size ($F_{2,16}=5.0$; $P<0.05$) was also found, indicating greater scaling of grip aperture with object size for real than for virtual objects.

Peak grip aperture, of course, is only an indirect index of apparent size. However, it is possible to interpret differences in grip aperture under the assumption that they arise from differences in apparent size. We calculated notional object widths and depths by performing regressions of peak grip aperture against object size for the real objects. This was done separately for each distance and for grasps across the width and the depth of the objects. Notional object size indicates the size of a real object that would elicit the observed peak grip aperture for a virtual object. Notional object widths and depths are plotted in Fig. 3c and d respectively. These results show clearly that grips were appropriate for objects larger than those presented at the close but not the far distance i.e. the use of binocular disparity and retinal image size as cues to object depth and size was affected by viewing distance.



(a) (b)

(c) (d)

Fig. 3 a Peak grip apertures (here averaged over the different object distances and grasps across the widths and depths of objects) scaled with object size for both real and virtual objects. **b** Differences between maximum grip apertures for real and virtual objects (*positive values* indicate that grips apertures were greater for the virtual objects). Peak grip apertures were greater for virtual objects than for real objects at the closer distance, but not at the further distance. Notional object sizes (see text for details), for **c** object width and **d** object depth, emphasize that grip apertures scaled appropriately for object size for disparity-defined objects, and were consistent with an overestimation of object size at the closer distance. *Symbols*: ○ 30 cm; ■ 50 cm distance

A final analysis was performed to assess any distortions in object shape. This was done by calculating a notional depth:width ratio for each object size. This index reflects distortions in object shape, and has been used in perceptual tasks to assess the accuracy with which disparity information is scaled to take account of viewing distance (Johnston 1991). The results suggest that objects appeared stretched in the depth dimension relative to the

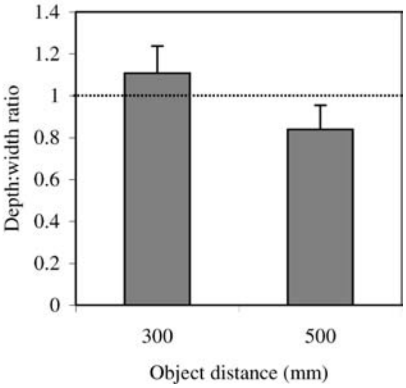


Fig. 4 Object shape (depth:width ratio) appeared to be affected by distance, in a manner consistent with that observed in perceptual studies

width dimension at the close distance, and that the reverse was true at the far distance (Fig. 4).

Discussion

For both real and disparity-defined, virtual objects, peak wrist velocity scaled with object distance, and peak grip aperture scaled with object size (both width and depth) showing that information about both location and size was readily available for these objects. Information about object distance was provided by both vergence and the height of the object in the visual scene, both of which have been shown to be important in the control of prehension (Mon-Williams and Dijkerman 1999; Marotta and Goodale 1998; Watt and Bradshaw 2002). Size information was provided by the retinal size of the image (width) and binocular disparity (depth). Here, we show that this information is sufficient to allow for the scaling of reaching and grasping in the control of prehension. The use of binocular information in this way is consistent with findings that prehension is more efficient under binocular than monocular viewing conditions (Servos et al. 1992; Marotta et al. 1997; Jackson et al. 1997; Watt and Bradshaw 2000).

Despite this scaling on the basis of binocular cues, there were clear biases in the use of this information. Specifically, participants acted as if the objects were smaller at the further distance than at the closer distance, and distance appeared underestimated at the further distance. Interestingly, this pattern of results is similar to that seen in many perceptual studies as discussed above, where the explanation also suggests inaccurate or incomplete scaling in taking account of fixation distance. It has been argued that the perception of metric structure arises by the application of a default set of viewing parameters, or by the combination of estimated and default parameters (Kontsevich 1998). A similar explanation may be forwarded in the context of the scaling of information for the control of prehension. However, there is no a priori reason to expect the same defaults, and therefore exactly the same biases, in each case. Another interesting feature of the data is that the biases for reaching and grasping, while both demonstrating insufficient scaling to take account of object distance, were not mutually consistent. Thus, estimated size appeared to be incorrect at the closer distance, while estimated distance appeared to be incorrect at the further distance. These results are consistent with relatively independent control of reaching and grasping (Marotta et al. 1997; Watt and Bradshaw 2000) and independent estimation of object size and distance by the visual system (Brenner and van Damme 1999). However, care should be taken before making such a conclusion. Alternatively, an overall difference in grip aperture or wrist velocity combined with consistent scaling of distance and size might explain these discrepant results. Either an overall increase in grip aperture, or an overall slowing of the reaching movement, for virtual versus real objects, could explain the apparent discrepancy between the effects of distance on reaching and grasping. While no firm conclusions can be drawn either way on the basis of these results, the latter possibility is consistent with the adoption of a “conservative strategy”, providing both an increase in grip aperture and a slowing of the reach movement.

The present results demonstrate that the visual system is not able to recover metric depth information accurately for the control of prehension on the basis of binocular cues alone. It is important to emphasise that these results do not question the importance of binocular cues in the control of prehension; rather, they suggest that, while this information is important, its use is subject to biases similar to those found in perceptual studies. While these results have important implications for our understanding of the control of prehension, they are probably of little consequence for the actual control of reaching in everyday conditions. Firstly, these results were obtained in a reduced cue environment, to assess the role of binocular cues in the control of prehension. In more natural situations, many cues to three-dimensional shape will be available, and it is likely that the accurate perception of depth relies on the integration of a number of sources of information, rather than relying on a single source in isolation (Landy et al. 1995; Bradshaw et al. 2000). Secondly, visual feedback will normally be available that will allow movements to be adjusted online (Morgan 1989). This latter possibility is of particular importance, since it suggests again that the visual system might rely on relatively simple information (in this case, the relative disparities between the digits and the object) and that the metric depth structure of the world need not be recovered accurately even in the control of prehension.

Acknowledgements This work was supported by the Wellcome Trust. We are grateful to Simon Watt for helpful comments on the manuscript.

References

Aglioti S, Goodale MA, DeSouza JFX (1995) Size-contrast illusions deceive the eye but not the hand. Curr Biol 5:679–685

Bradshaw MF, Glennerster A, Rogers BJ (1996) The effect of display size on disparity scaling from differential perspective and vergence cues. Vision Res 36:1255–1264

Bradshaw MF, Parton AD, Eagle RA (1998) The interaction of binocular disparity and motion parallax in determining perceived depth and perceived size. Perception 27:1317–1331

Bradshaw MF, Parton AD, Glennerster A (2000) The task-dependent use of binocular disparity and motion parallax information. Vision Res 40:3725–3734

Brenner E, van Damme WJM (1999) Perceived distance, shape and size. Vision Res 39:975–986

Foley JM (1980) Binocular distance perception. Psychol Rev 87:411–434

Glennerster A, Rogers BJ, Bradshaw MF (1996) Stereoscopic depth constancy depends on the subject's task. Vision Res 36:3441–3456

Haggard P, Wing AM (1990) Assessing and reporting the accuracy of position measurements made with optical tracking devices, J Motor Behav 22:315–321

Jackson SR, Jones CA, Newport R, Pritchard C (1997) A kinematic analysis of goal-directed prehension movements executed under binocular, monocular and memory guided viewing conditions. Visual Cognit 4:113–142

Jeannerod M (1984) The timing of natural prehension movements. J Motor Behav 16:235–254

Jeannerod M (1988) The neural and behavioural organization of goal-directed movements. Clarendon Press, Oxford

Jeannerod M, Decety J (1990) The accuracy of visuomotor transformation. An investigation into the mechanisms of visual recognition of objects. In: Goodale MA (ed) Vision and action: the control of grasping. Ablex, Norwood, pp 33–48

Johnston EB (1991) Systematic distortions of shape from stereopsis. Vision Res 31:1351–1360

Kontsevich LL (1998) Defaults in stereoscopic and kinetic depth perception. Proc R Soc Lond B Biol Sci 265:1615–1621

Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: in defence of weak fusion. Vision Res 35:389–412

Marotta JJ, Goodale MA (1998) The role of learned pictorial cues in the programming and control of grasping. Exp Brain Res 121:465–470

Marotta JJ, Behrmann M, Goodale MA (1997) The removal of binocular cues disrupts the calibration of grasping in patients with visual form agnosia. Exp Brain Res 116:113–121

Milner AD, Goodale MA (1995) The visual brain in action. Oxford University Press, Oxford

Mon-Williams M, Dijkerman HC (1999) The use of vergence information in the programming of prehension. Exp Brain Res 128:578–582

Morgan MJ (1989) Stereopsis – vision of solid objects. Nature 339:101–103

Oppenheim AV, Shafer RW (1989) Discrete-time signal processing. Prentice Hall, Englewood Cliffs

Servos P, Goodale MA, Jackobson LS (1992) The role of binocular vision in prehension: a kinematic analysis. Vision Res 32:119–127

Todd JT, Tittle JS, Norman JF (1995) Distortions of 3-dimensional space in the perceptual analysis of motion and stereo. Perception 24:75–86

Watt SJ, Bradshaw MF (2000) Binocular cues are important in controlling the grasp but not the reach in natural prehension movements, Neuropsychologia 38:1473–1481

Watt SJ, Bradshaw MF (2002) Height in the field and binocular cues support natural prehensile movements. Spatial Vis 15:253–254