NeuroReport 7, 825–829 (1996)

How perceived object dimension influences prehension

Umberto Castiello,[CA]
Claudia Bonfiglioli[1] **and**
Kerry M. B. Bennett

Department of Psychology, Monash University, Churchill, 3842, Victoria, Australia; [1]Dipartimento di Psicologia, Universita di Bologna and European Medical Centre, Pieve di Cento, Italy

[CA]Corresponding Author

A kinematic study assessed the effects of the perceived dimensions of an object upon the patterning of a prehension movement involving that object. If an apple was perceived as two-dimensional, subjects utilized a large precision grip between the index finger and thumb. If the apple was perceived as three-dimensional, whole hand prehension involving all the digits was utilized. A visual perturbation from perceived two-dimensional to three-dimensional at movement onset resulted in a transition from the 2D precision grip pattern to the 3D whole hand prehension. These results suggest that visual mechanisms for interpreting the dimensions of an object directly influence motor selection pathways, and do not necessarily access a three-dimensional central nervous system representation of the object.

Key Words: Object vision; Prehension; Kinematics; Two-dimensional; Three-dimensional; Recognition; Representation; Motor control; Reach; Grasp; Human; Visual pathways

Introduction

This study addresses the issue of how visual interpretative mechanisms influence motor control. It asks the question 'If an object is perceived as two-dimensional, will a motor act involving that object differ from that involving the same object when it is perceived as three-dimensional?'

Many theories about object recognition assume the existence of a strong associative memory structure, where visual input of an object is manipulated and matched to relevant stored models.[1,2] In object-centred theories, the description of a viewed object is derived on-line according to an intrinsic object-centred reference frame which is independent of the object's position.[3] Biederman[4] proposed that such viewer-centred invariance allows the definition of volumetric object components (geons) from the detection of the edge properties of an image. In viewer-centred theories, view-specific image descriptions of objects are matched to corresponding stored representations of these object views.[5] In this system, a network stores a certain number of different views associated with the object, these views being subdivided according to semantic and object category.[6,7]

Most object recognition theories presume the existence of three-dimensional (3D) models, where, presumably two-dimensional images of an object are matched to the relevant 3D representations. The issue of how two-dimensional (2D) and three-dimensional (3D) features interact was recently addressed by Ullman,[8] whose alignment method does not use three-dimensional models. Rather, the spatial coordinates of an object's contours are matched to those of pre-stored two-dimensional views; the identification process is built from a small collection of two-dimensional images.

Experimentation related to visual interpretation, including recognition, of an object has rarely been linked to motor outputs involving that object.[6] Conversely, movement studies, such as those related to prehension, have focused upon intrinsic object characteristics such as shape, size, function and category,[9–11] rather than upon the perceived dimensionality of the object to be operated upon. The current study addressed the link between the visual perception of an object's dimensions and the characteristics of the motor output which acts upon that object. It aimed to determine whether there are differences in the movement of reaching to grasp an apple according to whether the apple is perceived as two-dimensional or three-dimensional. If the movement pattern is different for the two-dimensional perception, is the motor output impermeable to a perturbation which changes the perceived dimensionality of the object, or will the originally recruited pattern be interrupted for the execution of a three-dimensional output? If the latter interruption occurs, is it possible to detect the point of computational passage from a 2D to a 3D output through a kinematic assessment?

Materials and Methods

Eight students (4 women and 4 men, aged 21–32 years) volunteered to participate. All were righthanded (Edinburgh Inventory[12]), reported normal or corrected-to-normal vision, and were ignorant as to the experimental purpose. Each subject attended an experimental session of approximately 1 h duration.

The subject was seated in front of a table (1 × 2m) in a dimly lit room. Prior to each trial, the ulnar border of the right hand was positioned on a pressure-sensitive switch 20 cm in front of the midline of the subject's thorax. The shoulder was flexed, the elbow flexed, the forearm semi-pronated, the wrist was in 10–15° of extension, and the index finger and thumb were held opposed. The target was an apple, presented on a tray in the mid-saggital plane, 30 cm from the switch. For the control/perturbed block, the apple appeared to be two-dimensional (2D) for 80 trials – this was achieved through a dark–light interplay with spotlights behind the apple. (These lights did not shine directly into the subject's eyes). For 20 randomly interspersed trials, a visuo-dimensional perturbation was introduced at the onset of the transport component; that is, upon release of the switch. With this spotlight shift, the apple became illuminated from above so that it appeared as a three-dimensional object. Two other blocks of 10 trials were also performed: 1) the silhouette apple; 2) the 3D apple. The order of these single blocks in relation to the perturbed/control block was counterbalanced.

Reflective markers (0.25 cm diameter) were attached to the reaching limb on the radial aspect of the distal styloid process of the wrist; on the radial side of the nails of the index, middle, ring and little fingers; and on the ulnar side of the thumb nail. Movements were recorded with the Elite system,[13] consisting of two infra-red cameras (sampling rate 100 Hz) inclined at an angle of 30° to the vertical and placed 3 m in front of the table and 3 m apart. The spatial error measured from stationary and moving stimuli was 0.04 mm. Coordinates of the markers were reconstructed with an accuracy of 1/3000 over the field of view and sent to a host computer (PC 386). The SD of the reconstruction error was 1/3000 for the vertical (Y) axis and 1.4/3000 for the two horizontal (X and Z) axes.

An auditory tone (880 Hz; duration 250 ms, occurring at a randomized time of between 500 and 2000 ms after the subject had positioned the hand on the switch) signalled the subject to reach, grasp the apple and bring it to the starting switch. To promote a 'natural' movement, no instructions were given as to response speed, movement velocity or grasp type. Each subject performed five practice trials in the same manner as the subsequent experimental block. Later analysis of these practice trials gave the same results as those obtained in the experimental trials.

ELIGRASP software (B/T/S, 1994) was used to give a 3D reconstruction of the marker positions, and to filter the data (FIR linear filter; transition band = 1 Hz, sharpening factor = 2).[14,15] The transport component was assessed by analysing the trajectory, velocity and acceleration profiles of the wrist marker. The manipulation component was assessed by analysing the trajectory of each of the hand markers, the distance between each of these markers, and the opening and closing velocity of the digits. Movement completion was taken as the time when the digits closed upon the fruit, and there was no further change in the distance between the index finger and thumb. The period where the fruit was brought to the starting switch, was not assessed. Transport component dependent variables were movement duration, the times to peak velocity, peak acceleration, and peak deceleration of the wrist marker, and the amplitudes of these peaks. Manipulation component dependent variables were the times to peak grip aperture (between index finger and thumb) and to peak grip velocity, the amplitudes of peak grip aperture and of the velocity peak, the time at which the index finger deviated from the more ulnar digits for specification of a large precision grip (Break Detection Algorithm,[9] specification time), and the time at which the more ulnar digits rejoined the path of the index finger for specification of the whole hand prehension (recruitment time).

Results

The mean for each dependent measure was determined for each type of trial, and data were analysed using an analysis of variance (ANOVA) whereby the within-group factor was type of trial (perturbed/control). Relevant mean results are shown in Table 1. A further ANOVA with type of trial (control/blocked) as the within-group factor was performed so that the results of non-perturbed trials interspersed with perturbed trials could be compared with those of non-perturbed trials which were performed in blocks. For all trials, regression analyses were used to determine correlations between temporal events of the manipulation and transport components. The Fisher Z-transformation of data was used for homogeneity of variance and to counteract any non-normal distributions.

Qualitatively, the results were clear. The 2D apple was almost always grasped as if it were a disc; that is, with a large precision grip between the index finger and thumb (Fig. 1B,C). In contrast, the 3D apple was always grasped with a whole hand prehension, whereby all digits and the palm contacted its surface (Fig. 1A).

The 2D precision grip was typically performed with the ulnar fingers flexed out of the way (Fig. 1B), and the index finger deviating away from these digits (specification time[9]) at an average time of 189 ms after switch release for the control trials. During the

Table 1. Mean values of relevant dependent measures. Standard deviations are shown in parentheses

	Blocked trials		Control trials	Perturbed trials
	2D	3D	2D	2D→3D
Movement duration (ms)	716 (84)	722 (81)	725 (79)	793 (82)
Transport component				
Time to peak acceleration (ms)	213 (31)	216 (28)	210 (28)	156 (18)
Manipulation component				
Specification time (ms)	192 (25)	–	189 (25)	192 (25)
Termination of specification (ms)	–	–	–	280 (38)
Recruitment time (ms)	–	–	–	330 (22)

2D = apple perceived as two-dimensional. 3D = apple perceived as three-dimensional. 2D→3D = visual perturbation from a 2D to a 3D visual perception.

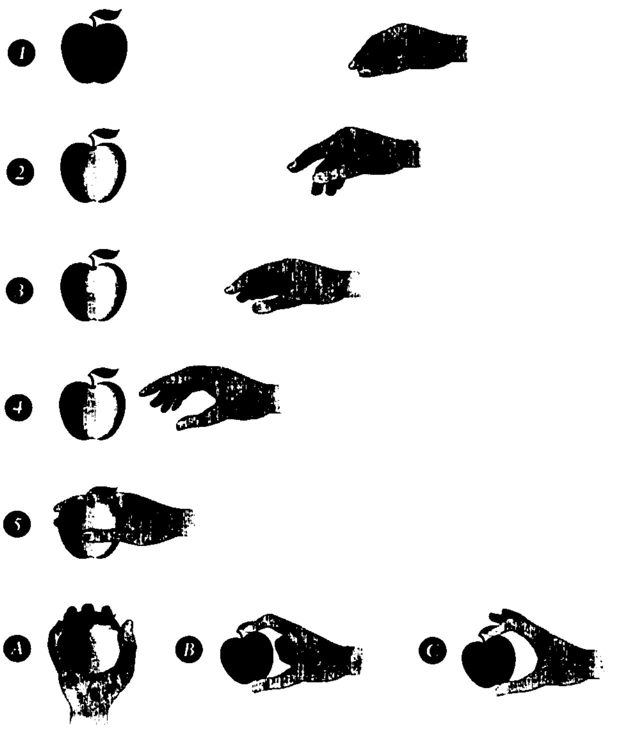

FIG. 1. Top five rows. Sequence of hand postures during a perturbed trials where, at movement onset to the apparently two-dimensional apple, the lighting changed so that the apple appears to be three-dimensional. The apple is shown as it appeared to the subject. **(1)** Starting position. **(2)** Spotlighting changes so that apple appears as a three-dimensional object. The index finger and thumb show specification for precision grip in response to the initial two-dimensional presentation. **(3)** Ulnar fingers show recruitment for whole hand prehension in response to the three-dimensional presentation. **(4)** Hand position towards completion of the movement. **(5)** Whole hand prehension grasp of the apple. Bottom row. **(A)** Typical whole hand prehension grasp of the three-dimensional apple. **(B)** Typical large precision grip of the two-dimensional apple; ulnar fingers flexed. **(C)** Less typical large precision grip of the two-dimensional apple; ulnar fingers extended.

perturbed/control blocks, 89.53% of the 640 control trials (8 subjects × 80 trials) were performed in this way. For 35 cases (5.47%) the middle finger was also involved. For 22 cases (3.44%) a whole hand prehension was utilized, and for 10 cases the ulnar digits were held extended rather than flexed (but the object was grasped with a precision grip; Fig. 1C). Results for the blocked trials mirrored those obtained for the control/perturbed trials. Regardless of the block presentation order, the apparently 2D apple was usually grasped with a large index finger/thumb opposition with ulnar fingers flexed (70/80 trials; 87.5%), and the 3D apple was grasped with a whole hand prehension (80/80).

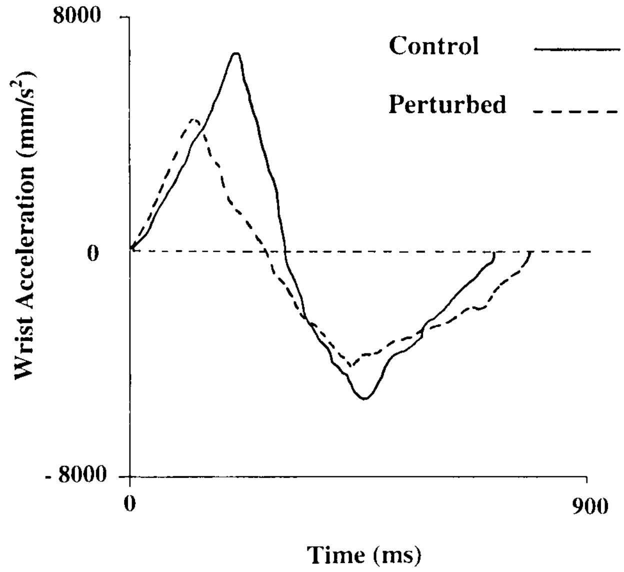

FIG. 2. An example of the acceleration profiles for the Control 2D trials (solid line) and the perturbed (2D 'becoming' 3D) trials (dotted line) of one subject. Note that the peak of acceleration for the perturbed trials is earlier than that of the control trials.

Remarkably, a precision grip for the 2D apple was utilized even for most of the control trials performed immediately after the perturbed 3D trials where a whole hand prehension had been utilized. Of the 160 cases, 152 (95%) showed the typical precision grip pattern.

Analysis of the perturbed trials showed that the subject began a large precision grip but then rapidly changed the output to a whole hand prehension. Mirroring the results of the control trials, specification for precision grip occurred at a mean of 192 ms after switch release, indicating initial programming for precision grip in response to the apparently 2D apple. An obvious halting of this output was revealed by a dip in the grip opening velocity profile at a mean of 280 ms. Changes in the trajectories of the ulnar digits (middle, ring, little) showed that whole hand prehension was then recruited (Fig. 1) at an average of 330 (± 10 ms) after switch release (and thus after perturbation). The transport component showed an even earlier recognition of the perturbation (Fig. 2). The time to peak wrist acceleration was earlier for perturbed (156 ms) than for control trials (210 ms; $F(1,7) = 28.04$, $p < 0.0001$). For six subjects, this peak time was strongly correlated to the time of recruitment for whole hand prehension ($r_s = 0.78–0.91$). A further significant difference between the perturbed and control trials was that movement duration was longer for the former ($F(1,7) = 17.05$, $p < 0.001$). When comparing blocked trials with 10 randomly chosen control trials to the apparently 2D apple, no differences were found.

Discussion

In this study, an apple was presented in front of a subject, who was instructed to reach for and grasp this everyday target. During the experiment, nothing about the physical characteristics of the apple changed. The only manipulation was to the way in which the dimensions of the apple were perceived – two-dimensional or three-dimensional. One remarkable finding was that the subject grasped the perceived 2D apple using a large opposition grasp between the index finger and thumb. In contrast, the 3D apple was grasped with a whole hand prehension, involving more digits.

Previous research has demonstrated that humans normally use a precision type grasp for smaller, more delicate objects, and whole hand prehension for larger objects.[9,16] Thus the use of a precision grip for an apple may appear inappropriate if related to an apple *per se*. However, it is not inappropriate if the apple is considered as a disc. For example, Marteniuk *et al*[17] assessed the transport and manipulation components of prehension when subjects were instructed to reach and grasp for discs of 1–10 cm diameter. The type of prehension utilized was consistently a precision grip which showed an increase in index finger/thumb aperture with disc size. The results of the current study indicate that the apple was perceived as two-dimensional (or as a disc), and that this online visual interpretation of the target's dimensions clearly influenced the motor output to that object. The motor output to an object which is perceived as two-dimensional (or $2\frac{1}{2}$ D[3]) is clearly different from the motor output to an object which is perceived as three-dimensional.

The strength of the effects of this on-line visual interpretation was confirmed with the perturbation results. For perturbed trials, the subject initially perceived the apple as two-dimensional. Upon initiating the reach to the apple, it suddenly 'became' three-dimensional. Surprisingly, the initially recruited 2D pattern of a large precision grip was curbed and a whole hand prehension was used to grasp the apple. This rapid modification was evident in both the manipulation and transport components of the action, and is consistent with other perturbation study results. For example, and as previously found,[18,19] the transport component showed an earlier acceleration peak for perturbed than for control trials, indicating recognition of the dimension change and early changes to facilitate those necessary in the manipulation component. The changes in the manipulation component are consistent with an arrest of the precision grip and with the recruitment of the 3D whole hand prehension pattern. Intuitively, this change seems unnecessary, given that the apple is still the same apple. However, it demonstrates the potent effect of the interpretation of an object's dimensions in commandeering a defined motor output.

Do these results demonstrate that a stored model of the object has been utilized somewhere in the process between visual interpretation of its dimensions and the motor output? After all, an apple is a familiar object, one which is often grasped, and one which could be presumed to have a central nervous system representation. The answer is of course difficult, but it can be assumed that even if the silhouette is recognized as an apple, and even if exactly the same apple has been previously grasped when it was perceived as three-dimensional, we do not appear to scan a canonical mental dictionary of object models that, presumably, would suggest the use of whole hand prehension. Alternatively, it maybe that we scan this model, but that its influence is over-ridden by the powerful effects of on-line visual interpretation of perceived dimension.

It could be suggested that the associative memory system is divided into more specialized stores where two-dimensional object shape systems are separated from their equivalent three-dimensional representations. This would be consistent with Marr's computational approach to vision,[3,20] which is divided into three stages: (1) the primal sketch – transformation of the image into a primitive but rich description of the way its intensities change over the visual field; (2) the $2\frac{1}{2}$ dimensional sketch – the giving of explicit information about an image so that it can be brought to a state consistent with a three-dimensional representation; (3) a three-dimensional representation which assumes that a single stored model of an object is accessed no matter what view of the object is presented.

Because the motor outputs differ according to perceived dimension, the results from the current study support the idea of stages, or different channels, of visual interpretation. A single-level representation of an object is not supported because this would predict identical outputs to the same object irrespective of whether it was perceived as two- or three-dimensional. Similarly, the results do not support the idea that presentation of a normally perceived 3D object in a novel 2D form elicits a matching process with stored views of the 3D object. Adopting Marr's model,[3,20] it would appear that under certain conditions a visual stage related to the composition of 2-$2\frac{1}{2}$D attributes accesses motor outputs using channels which bypass, or pass through but ignore, a stage related to the composition of 3D attributes. The very low incidence of trials in which whole hand prehension was used for the 2D apple could, however, indicate a default level of inter-channel cross-talk, or that a 3D representation is not entirely ignored.

Ullman's theory,[8] that combinations of two-dimensional views contribute to recognition, is perhaps best approximated by these results. In particular, support is given to the hypothesis that the direct use of an associative memory for two-dimensional motor output patterns does not require an obligatory match with a three-dimensional model. A point requiring further investigation is whether the transition from a two-dimensional to a three-dimensional motor output occurs through a matching of small sets of stored patterns.[8]

Conclusion

This study presents two notable findings. First, the planning of a motor output involving an object is related to the on-line visual interpretation of the object's dimensions, and can be independent of the object's semantic meaning. Thus, though an object, presented from different views and in different dimensions, may share a common structure for name, category and so on, the choice of the motor output relies on a hypothetical motor associative memory structure[21] that takes into account the actual characteristics of the viewed object. Second, kinematic corrective mechanisms show that there is a transition from a 2D to a 3D output. This could reflect a simple shift from one independent dimension-motor channel to another. Alternatively, it may reflect an adaptive compensatory process whereby small sets of specialized 2D stores are combined to form a 3D representation[8] which then influences the motor outcome.

References

1. Lawson R, Humphreys GW and Watson DG. *Perception* **99**, 593–613 (1994).
2. Treisman A. Representing visual objects. In: Meyer DE and Kornblum S, eds. *Attention and Performance*. Cambridge, MA: MIT Press, 1993: 162–174.
3. Marr D and Nishihara HK. *Tech Rev* **1**, 2–23 (1978).
4. Biederman I. *Psychol Rev* **94**, 115–147 (1987).
5. Tarr MJ and Pinker S. *Cogn Psychol* **21**, 233–283 (1989).
6. Castiello U, Scarpa M and Bennett KMB. *Nature* **374**, 805–808 (1995).
7. Tulving E. Episodic and semantic memory. In: Tulving E and Donaldson W, eds. *Organization of Memory*. New York: Academic Press, 1972: 381–403.
8. Ullman S. The visual recognition of three-dimensional objects. In: Meyer DE and Kornblum S, eds. *Attention and Performance*. Cambridge, MA: MIT Press, 1993: 79–98.
9. Castiello U, Bennett KMB and Stelmach GE. *Exp Brain Res* **94**, 165–178 (1993).
10. Jeannerod M. *J Mot Behav*, **16**, 235–254 (1984).
11. Klazky RL, McCloskey B, Doherty S *et al. J Mot Behav* **19**, 187–213 (1987).
12. Oldfield RC. *Neuropsychologia* **9**, 97–113 (1971).
13. Ferrigno G and Pedotti A. *IEEE Trans Biomed Eng* **32**, 943—950 (1985).
14. D'Amico M and Ferrigno G. *IEEE Trans Biomed Eng* **28**, 407–415 (1990).
15. D'Amico M and Gerrigno G. *IEEE Trans Biomed Eng* **30**, 193–204 (1992).
16. Castiello U. *J Exp Psychol: Human Percept Perf* (in press).
17. Marteniuk RG, Leavitt JL, MacKenzie CL *et al. Hum Mov Sci* **9**, 149–176 (1990).
18. Paulignan Y, MacKenzie C, Martenuik R *et al. Exp Brain Res* **83**, 502–512 (1991).
19. Castiello U, Paulignan Y and Jeannerod M. *Brain* **114**, 2639–2655 (1991).
20. Marr D. *Vision*. San Francisco: WH Freeman 1982.
21. Kosslyn SM, Alpert NM, Thompson WL *et al. Brain* **117**, 1055–1071 (1994).

ACKNOWLEDGEMENTS: We thank B/T/S/ s.r.l. and, in particular Drs R. Assente, E. D'Amico and M. Bulgheroni for technical support. Professor Razzaboni and Dr Colonna are also thanked for allowing us to use the excellent facilities at the European Medical Centre. Alessandro Allodi is acknowledged for his technical expertise in the spotlighting procedure, and Sharon Harrup is thanked for her assistance with the figures. Funding for this research has been obtained from the Monash Research Fund and the Australian Research Council to U. Castiello.

**Received 18 December 1995;
accepted 23 January 1996**