

Constraints and Principles for the Design of Human-Machine Interfaces: A Virtual Reality Approach

Yaoping Hu^{*,†}, Melvyn A. Goodale^{*}, and Roy Eagleson[†]

^{*}Vision and Motor Control Laboratory, Dept. of Psychology

[†]Dept. of Electrical and Computer Engineering

University of Western Ontario, London, Ontario, Canada, N6A 5C2

E-mail: {yaoping, mgoodale, eagleson}@julian.uwo.ca

ABSTRACT

We conducted two experiments to compare the visual frames of reference used to scale grasping movements directed at objects with those used to estimate the size of the same objects -- either immediately or after a 5-s delay. A virtual "workbench" was employed for presenting two different-sized objects in 3D. Subjects were instructed to pick up or estimate the marked one of the two objects. We found that the presence of the other object affected not only the estimate of the size of the target object when subjects made their estimates both immediately and after a 5-s delay, but also the scaling of grip aperture in flight when subjects picked up the target object after a 5-s delay. However, when subjects picked up the target object immediately, their grasp was scaled to the actual size of the target object and was not influenced by the presence of the other object. These findings suggest that the control of delayed motor actions utilizes the same relative metrics in allocentric frames of reference used by conscious perception, whereas the control of normal visually guided motor actions relies on absolute metrics in egocentric frames of reference. Implications of these findings for the design of human-machine interfaces are discussed.

1. INTRODUCTION

Humans not only design and create machines, but they are the undisputed users and operators of those machines. Often these machines are controlled from a distance. Modern examples of the latter are tele-robotics and tele-presence. The physical constraints of the hardware and software that are used to process information and to implement the operator's decisions make the task of improving the efficiency, stability, and usability of human-machine interfaces a complex task. Part of the problem is related to human factors. Humans interacting with objects in their surroundings do not always behave in the way the designers planned; instead they tend to get actively involved in exploring and acting upon goal objects [3]. In other

words, designers have not yet properly integrated human factors into their design, and have not considered the role of the functional and architectural structure of the cognitive systems that drive human behavior. A consideration of these factors could lead to the development of better human-machine interfaces.

It was this goal that motivated us to conduct two experiments in which we investigated the frames of reference used by human visuomotor systems mediating actions directed at objects and by human perception of those objects. The framework for this research was based on a recent theory of the functional and architectural structure of human cognitive systems¹. We presented two different-sized objects in a visual scene and varied the temporal delay between viewing the objects and the point at which the human subjects initiated their grasping movements (or perceptual estimates) -- simulating the unavoidable temporal delay in communication between a human operator and a machine in many human-machine systems. The objectives of this study were to explore how the temporal delay shifts the frame of reference used by subjects to generate their object-directed actions and to explore the implications of this shift for the design of human-machine interfaces.

The present paper is organized as follows: In Section 2, we briefly describe the theoretical framework for the distinction between the visual control of action and visual perception in the human brain. In Section 3 and 4, we describe the grasping and estimating experiments, respectively. In Section 5, we discuss the implications of these results for the design of human-machine (H-M) interfaces. Finally, in Section 6, we present our conclusions.

¹ The experiments were carried out at the University of Western Ontario in compliance with the SSHRC (Canada) Guidelines (1981) and supported by NSERC (Canada) Grant No. 6313 to M. A. Goodale.

2. ACTION AND PERCEPTION

Vision plays an essential role in guiding our actions and forming our thought about the world. Although it seems self-evident that our actions are directed by our conscious perception of objects, recent evidence from a broad range of studies on neurological patients and normal observers tells us otherwise. In 1992, on the basis of this work and complementary studies in the monkey, Goodale and Milner proposed that the visual control of actions directed at objects is mediated by neural processes that are functionally and anatomically distinct from those mediating the visual perception of those objects [7]. In particular, they proposed that the dorsal stream of visual processing, which arises from primary visual cortex and projects to posterior parietal cortex, transforms visual information for the control of skilled motor actions whereas the ventral stream of visual processing, which also arises from primary visual cortex but projects instead to infero-temporal cortex, transforms visual information for perceptual representation. Of course, vision for action and vision for perception work together in controlling our behaviour as we live our complex lives. But they play complementary roles in the production of adaptive behaviour.

Evidence from a number of studies of target-directed manual pointing and eye movements vs. conscious perceptual judgements about target location has shown that the actions use absolute metrics in egocentric frames of reference whereas the perceptual judgements use relative metrics in allocentric (world-based) frames of reference [2], [13]. Similar dissociations have also been observed in studies on scaling of grip aperture in manual prehension vs. conscious perception of object size in pictorial illusions [1], [4], [8].

Nevertheless, skilled motor actions are not entirely immune to the influences of conscious perception. Relative perceptual metrics *can* affect motor performance when pointing movements are not initiated immediately, but after a delay [2], [13]. Moreover, there is evidence that the kinematics of grasping movements are disrupted by such a delay [6], [9]. It is possible that grip scaling after a delay shows evidence of the effect of relative size judgments. To examine this possibility, we carried out a grasping experiment in which we compared grasping movements directed at objects in real time with grasping movements initiated to the same objects but after a 5-s delay. In a separate experiment, we examined the metrics used by perceptual judgements when subjects were asked to estimate how large these

objects were both in real time and after a 5-s delay. Overall, we expected that the frames of reference used to program grasping after a delay would be similar to those used to make perceptual judgements.

3. GRASPING EXPERIMENT

In the grasping experiment, we anticipated that subjects who grasped the target object in real time would not be influenced by the difference in size between the target object and its companion. In other words, we predicted that maximum grip aperture, a kinematic parameter that is highly correlated with object size [10], [11], would not be influenced by this relative size information – but would instead be correlated with the actual size of the object. In the case of delayed grasping, however, we expected that maximum grip aperture would be influenced by the relative difference in size between the two displayed objects. That is, we expected that after a delay subjects would open their fingers wider for a target object paired with a smaller object than they would for the same object paired with a larger object.

Methods

Subjects: The 28 subjects (mean age = 21.5 years) tested were strongly right handed, as determined by a modified version of the Edinburgh handedness inventory [12]. All subjects had normal or corrected-to-normal vision, with a stereoacuity of at least 40" of arc as determined by the Randot Stereotest (Stereo Optical, Chicago, USA). The subjects were students at the University of Western Ontario (UWO) and were paid for their participation.

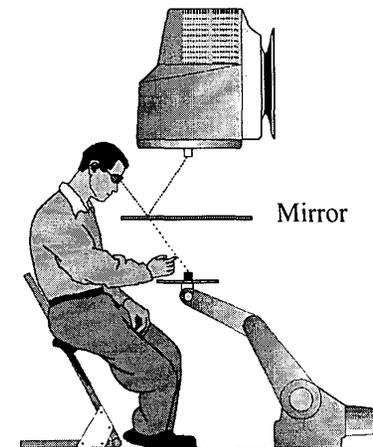


Fig. 1 Simple illustration of the experimental set-up.

Apparatus: Fig. 1 illustrates the experimental set-up. A SGI monitor was used to present two virtual objects on each trial. As shown in Fig. 2, one virtual object that was marked in the center of its top surface with a red dot served as the target. From trial to trial, the marked object was paired with either a larger, smaller or an identical non-target object, and was presented on the left or right in random order. These virtual objects were viewed through 3D stereo goggles (CrystalEyes) in a first-surface mirror placed below the monitor. A real object identical in size to the marked object was placed with a robot arm (251A, CRS Robotics, Burlington, Canada) at the corresponding location below the mirror. When subjects reached out to pick up the marked object, they ended up grasping the real object. A start button was located 80 cm above the floor at the subject's midline.

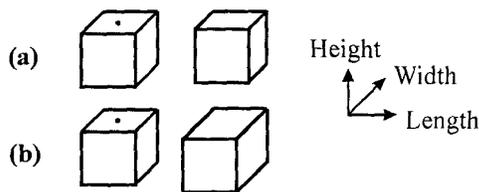


Fig. 2 Virtual objects in 3D used in grasping and estimating experiments. The target object was marked with a red dot. (a) the marked object was the larger object; (b) the marked object was the smaller object. The marked objects in (a) and (b) were actually identical in size.

Three infrared diodes (IREDs) were attached to the subject's index finger, thumb, and wrist, respectively. A WATSMART system (Northern Digital, Waterloo, Canada) was used to record the spatial position of each IRED at sampling rate of 100 Hz. The accuracy of the system and techniques of data analysis have been documented elsewhere [9].

Target objects were three blocks in size of $4 \times 4 \times 3$ cm, $4 \times 4 \times 4$ cm, and $4 \times 4 \times 5$ cm, respectively. Notice that only the dimension of object width was varied (see Fig. 2). The difference in width between the two virtual objects was 0.5 cm.

Procedure: Subjects were divided into two groups: one group was tested in real-time, the other group with a 5-s delay between viewing the objects and initiating a grasping movement. For the subjects tested in real-time (no-delay), subjects were instructed to reach out and pick up the marked object as soon as the virtual objects were presented. When their fingers left the start button, the virtual object display

disappeared. The subjects, therefore, executed the entire movement without seeing their hand or the target object. Data collection began when virtual objects were presented and stopped when the target object was picked up. For the subjects tested with a delay, there was a 5-s delay following a 500-ms viewing period. During the delay period, in which subjects could see neither their hand nor the target objects, they were required to keep their hand on the start button. An auditory signal at the end of the delay period signalled to the subjects that they should initiate the grasping movement. Data collection began when the auditory signal was presented and stopped when the target object was picked up.

Both groups of subjects were tested in one block of 72 trials in which they were instructed to pick up the marked object on each trial. Subjects were asked to bring the tips of their index finger and thumb together and to depress the start button before initiating a trial. Subjects were instructed to pick up the target object across its width using their index finger and thumb.

Dependent measures: The kinematic parameter, maximum grip aperture, was computed from the stored 3D position for each IRED. Maximum grip aperture was the maximum resultant distance between the index finger and thumb in flight as the subject reached out the hand to grasp the marked object.

Means of maximum grip aperture were calculated from grasping movements made to each of the target objects. Less than 3% of the trials were eliminated because of incomplete data. Each mean was based on a minimum of three trials.

Results and Discussions

Our predictions were borne out by the data in the grasping experiment. Subjects who picked up the marked object in real time were not affected by the presence of another larger or smaller object. As Fig. 3a illustrates, maximum grip aperture was the same for a marked object whether that object was larger or the smaller of the two objects in the visual array. Moreover, as the group data summarized in Fig. 4 indicate, there was no significant difference in grip aperture between the two kinds of trials [67.8 mm vs. 67.7 mm for 'larger' vs. 'smaller', $F(1,13) = 0.031$, *n.s.*]. However, the subjects' grip aperture was sensitive to the actual width of the object [$F(2,26) = 55.017$, $p < 0.001$], similar to what has been described earlier [10], [11].

As expected, the picture was quite different for the subjects who reached out to pick up the marked object after the delay. As can be seen in the individual

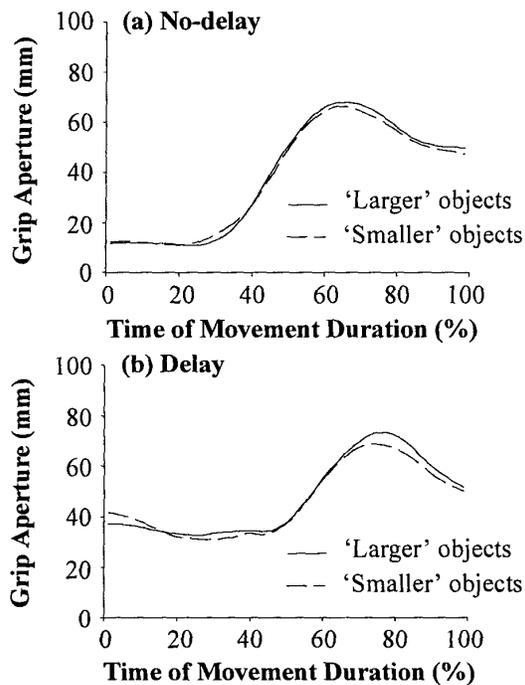


Fig. 3 Representative traces of grip aperture in the grasping experiment for grasping the larger and the smaller of two objects. In both cases, the target object was the same size; it was simply accompanied by a smaller object in one case and a larger object in the other. (a) No delay; (b) Delay.

data illustrated in Fig. 3b as well as in the group data summarized in Fig. 4, subjects opened their fingers significantly wider when the marked object was the larger of the two objects than when it was the smaller one [67.3 mm vs. 65.9 mm for 'larger' vs. 'smaller', $F(1,13) = 7.175$, $p < 0.05$] – even though the marked object was the same size in both kinds of trials. Again, superimposed on this relative size effect was the effect of differences in the real size of the target object that was used: the wider the target object, the wider the subjects opened their hand [$F(2,26) = 81.361$, $p < 0.001$].

The fact that grip scaling becomes sensitive to the relative size of an object after a delay reflects a shift from absolute to relative metrics. It appears that even a delay as short as 5 s forces the subject to rely, not on the real-time visuomotor transformations that would normally be used to program an object-directed grasping movement, but on stored perceptual memories of the target object and the array in which it was embedded. As we reviewed in the Introduction,

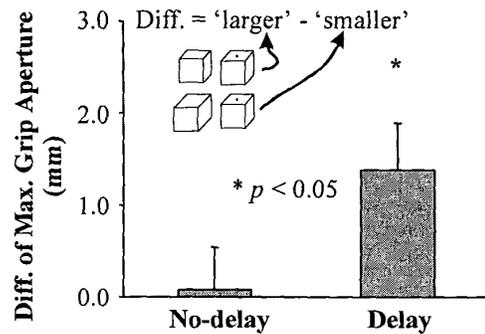


Fig. 4 Mean values of the difference in maximum grip aperture in the grasping experiment between grasping the marked objects paired by a smaller object and the marked objects paired by a larger object. (Error bars represent the standard error of the mean difference).

because conscious perception uses allocentric frames of reference, the relative size of the two objects in the array stored in memory can influence the scaling of the grasp when the object is no longer in view.

4. ESTIMATING EXPERIMENT

We expected that conscious judgements of the targets' size would show the same sensitivity to accompanying objects as we observed in delayed grasping. To examine this possibility, we conducted an estimating experiment, in which subjects were requested to make conscious perceptual estimations of the size of a target object by opening their thumb and index finger a matching amount while keeping their hand on the start platform.

Methods

Subjects: The 10 subjects (mean age = 22.0 years) tested were examined for handedness and stereoacuity as those in the grasping experiment. These subjects were students at UWO and were paid for their participation.

Apparatus: Same as in the grasping experiment.

Procedure: In contrast to the grasping experiment, subjects in the estimating experiment were asked to estimate the width of the marked object by opening their index finger and thumb a matching amount. In one block of trials, subjects estimated the width of the marked object as soon as the display was turned on. In another block of trials, they estimated its width after a 5-s delay. The viewing time for the delay trials was identical to that used in the delayed grasping; and the viewing time for the no-delay trials

was comparable to that used in the real-time grasping. The recording of the opening between the finger and thumb began when the subject's fingers left the start button and lasted for 2.5 s. Finally, an auditory signal indicated that the subject should reach out and pick up the marked object using their index finger and thumb. This last step ensured that subjects would receive the same haptic feedback as subjects in the grasping experiment.

Dependent measures: The dependent measure was the aperture (resultant distance) between the IREDs on the index finger and thumb when subjects estimated the width of the marked object. The aperture was measured when no changes had occurred for at least ten consecutive frames.

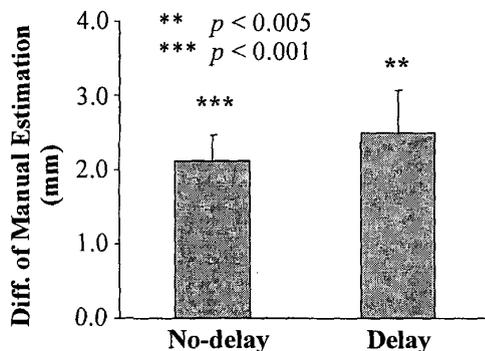


Fig. 5 Mean values of the difference in manual estimation in the estimating experiment between estimating the marked objects paired by a smaller object and the marked objects paired by a larger object. (Error bars represent the standard error of the mean difference).

Results and Discussions

Overall, subjects estimated the size of the marked object to be significantly larger in the delay condition than in the no-delay condition [41.5 mm vs. 38.1 mm for delay vs. no-delay; $F(1, 9) = 22.598$, $p < 0.001$]. This result agreed with observations in a number of earlier studies [9], [10], although it is not clear why the delay results in this tendency to overestimate the size of objects. Importantly, however, the size estimates of the target object were influenced by the other object in scene, no matter whether subjects gave the estimates immediately or after a 5-s delay. As Fig. 5 illustrates, in both delay and no-delay conditions, subjects opened their fingers significantly wider when the target was the larger of the two objects than when it was the smaller one [Delay condition: 42.7 mm vs. 40.2 mm for 'larger' vs. 'smaller'; $F(1,9) = 18.167$, p

< 0.005 ; No-delay condition: 39.2 mm vs. 37.1 mm for 'larger' vs. 'smaller'; $F(1,9) = 37.034$, $p < 0.001$] – even though the target was the same size in both kinds of trials. Again, superimposed on this relative size effect was the effect of the real width of the object: in all conditions, the wider the target object, the wider the subjects opened their fingers in manual estimations [Delay condition: $F(2,18) = 98.952$, $p < 0.001$; No-delay condition: $F(2,18) = 52.478$, $p < 0.001$]. The results of the estimating show that even in real time our conscious perception is influenced by the relative size of objects in the visual scene. Moreover, the order of magnitude of the difference in estimation is similar to that seen in grip scaling after a delay.

5. IMPLICATIONS IN H-M INTERFACES

Real-time Action and Perception

Taken together, the results of the two experiments described above provide convincing evidence for a dissociation between the metrics used by the visual mechanisms mediating the control of action and those used by the visual mechanisms mediating conscious perception. Even though perceptual judgments of an object's size are influenced by the presence of other objects in the visual array – resulting in a size-contrast effect – the scaling of real-time grasping movements directed at those objects reflect the absolute, not the relative, size of the target. In other words, even when the same visual display is used, the motor actions directed at target objects differ from the conscious perception of those same target objects.

It is important to note that the stimulus array that we used in the present experiments was an example of a situation in which human operators confront multiple goal objects in a visual array. Yet even with a two-object display, we found a clear difference between the scaling of grasping movements made in real time and the manual estimates of those objects. The former (vision for action) was based on absolute metrics in egocentric frames of reference while the latter (vision for perception) was based on relative metrics in allocentric frames of reference.

These factors should be considered when designing human-machine interfaces. If a human operator manipulates certain tools to interact with objects in a visual display (real or virtual), a reference point should be provided with respect to the operator, helping him/her establish an extended egocentric frames of reference for executing his/her motor actions. Certainly, this scenario is different from monitoring a visual display for understanding the

scene – as watching television in allocentric frames of reference.

Delayed Action and Real-time Perception

Although the visual control of real-time actions depends largely on absolute metrics in egocentric frames of reference, our results show that the scaling of actions made after a delay can be influenced by the presence of other objects in the same visual array. This scaling, like perceptual judgements made in real time, appears to be based on the relative, not the absolute, metrics of the target object.

The demonstration that the programming of delayed grasping movements uses relative metrics to scale grip aperture parallels earlier work showing that the programming of the spatial endpoint of delayed saccades and manual aiming uses allocentric rather than egocentric frames of reference [2], [5]. Actions directed at remembered objects cannot rely on the same motor programs used by real-time actions. After all, the location and disposition of the goal object in egocentric frames of reference could change dramatically as the observer moves with respect to the object (or vice versa). It would make little sense, therefore, to store the motor programs more than a few milliseconds. Nevertheless, on occasion, we do make movements towards objects that we have seen earlier. The evidence from the present study suggests that, when we do this, we base the motor program on stored perceptual information – and in doing so incorporate the relative metrics of conscious perception into the programming of grip aperture.

As we described in the Introduction, the delayed action simulates the case of human-machine systems, in which a temporal delay in communication between a human operator and a machine is unavoidable. Our study would suggest that when a human operator encounters such a delay, his/her motor actions would fall into the same allocentric frames of reference used by conscious perception – even though the operator receives proper haptic feedback from his/her actions. As a result, his/her motor actions are different from those executed in real time. Thus, care should be taken when designing human-machine interfaces to deal with the inevitable delay between command and action. One way to avoid interpreting visual information in relative metrics would be to convert visual information from relative metrics into absolute metrics in using a pre-calibrated frame of reference with respect to human operators in cases where a temporal delay occurs.

6. CONCLUSIONS

In sum, relative information derived from stored conscious percepts influences the scaling of grasping made after a 5-s delay between viewing target objects and initiating a grasp -- even if no such influence occurs in the scaling of real-time grasping. These findings have important implications for the design of more efficient and stable human-machine interfaces.

REFERENCE

- [1]. Aglioti, S., Desouza, J. F. X., & Goodale, M. A., Size-contrast illusions deceive the eye but not the hand. *Curr. Biol.*, 5, 679-685, 1995.
- [2]. Bridgeman, B., Peery, S., & Anand, S., Interaction of cognitive and sensorimotor maps of visual space. *Percept. & Psychophys.*, 59, 456-469, 1997.
- [3]. Carroll, J.M. & Mack, R.L., Metaphor, computing systems, and active learning. *Intl. J. Man-Machine Studies*. 22, 39-57, 1985.
- [4]. Gentilucci, M., Chieffi, S., Daprati, E., Saetti, M.C., & Toni, I., Visual illusion and action, *Neuro-psychologia*, 34, 369-376, 1996.
- [5]. Gnadt, J. W., Bracewell, R. M., & Andersen, R. A., Sensorimotor transformations during eye move-ments to remembered targets. *Vision Res.*, 31, 693-715, 1991.
- [6]. Goodale, M. A., Jakobson, L. S., & Keillor, J. M., Differences in the visual control of pantomimed and natural grasping movements. *Neuro-psychologia*, 32, 1159-1178, 1994.
- [7]. Goodale, M. A., & Milner, A. D., Separate visual pathways for perception and action. *Trends in Neurosci.*, 15, 20-25, 1992.
- [8]. Haffenden, A. M., & Goodale, M. A., The effect of pictorial illusion on prehension and perception, *J. Cogn. Neurosci.*, 10, 122-136, 1998.
- [9]. Hu, Y., Eagleson, R., & Goodale, M. A., The effects of delay on the kinematics of grasping. *Exp. Brain Res.*, 126, 109-116, 1999.
- [10]. Jakobson, L. S., & Goodale, M. A., Factors affecting higher-order movement planning: a kinematic analysis of human prehension. *Exp. Brain Res.*, 86, 199-208, 1991.
- [11]. Jeannerod, M., *The neural and behavioural organization of goal directed movements*, Oxford: Oxford University Press, 1988.
- [12]. Oldfield, R. C., The assessment and analysis of handedness: the Edinburgh inventory. *Neuro-psychologia*, 9, 97-112, 1971.
- [13]. Wong, E. & Mack, A., Saccadic programming and perceived location. *Acta Psych.*, 48, 123-131, 1981.