When is Vision Useful During A Familiar Manipulatory Task?

Roberta L. Klatzky
Carnegie Mellon University

Susan J. Lederman
Queen’s University

Kimberly A. Purdy
Queen’s University

ABSTRACT

We asked at what point(s) vision is particularly useful for a simple and familiar peg-in-hole task. Experimental participants performed the task for 200 trials, 100 trials with full vision and 100 trials without vision or with vision at just one of four critical stages -- prior to reach onset, when reaching for the peg, when grasping the peg, or when transporting the arm and inserting the peg into the hole. Comparisons of the mean time to perform under these conditions showed a graded forward effect of early vision, such that vision at initial stages facilitated performance at later ones, but with less and less effect on stages that occurred progressively later following visual offset. In most cases, vision was useful in the stage at which it occurred, but an important exception was the grasping stage. Vision provided at the point of grasping did not facilitate grasping but did affect transporting the arm and inserting the peg, as did vision during the transport/insert stage itself. Implications for direct and remote manipulation are considered.

INTRODUCTION

In general terms, manipulation refers to contact with an object for functional purposes, usually involving the application of forces to move the object or one of its parts. Manipulation is a multistage event. Often, it requires transporting the arm to the object and grasping it appropriately, before force is applied (see MacKenzie & Iberall, 1994, for review). In this paper, we are concerned with the need for vision during a familiar manipulatory task. We ask, in particular, how well a task can be performed when vision is available only at certain stages, if that task has previously been performed repeatedly under full visual control. This paper presents data for a subset of the conditions in an extensive experiment; in a subsequent journal publication we will present the full data set.

The movement of an object or part during manipulation is often constrained with respect to external or object-centered space. In fact, we suggest that many common manipulatory tasks are variants on the simple peg-in-hole procedure. In tasks of this type, grasped objects or parts must be mated with others. Consider such operations as fastening a shirt button, tying a knot, plugging in an electrical cord, threading a needle, inserting a straw into a glass, capping a pen, or putting a cup into a saucer. In each of these cases, mating of objects or parts occurs.

Manipulation typically relies on sensory input from both vision and touch. Under visual guidance, contact with the object is preceded by some planning of the desired movement, localizing the object and reaching for it. Studies have shown that the time prior to initiating a movement is related to its complexity or the number of sub-movements to be performed (Henry & Rogers, 1960), and to constraints on force at contact (Klatzky, Fikes, & Pellegrino, 1995) or spatial placement (Sidaway, 1991; Spijkers, 1987). It has often been suggested that the pre-movement interval is spent planning the commands needed to execute the movement (Klapp, 1977; Rosenbaum, 1987). More complex or constrained tasks involve a greater number of motor instructions or more fully parameterized ones; they therefore require longer planning time than do simple actions.

Research on reaching indicates that it is composed of two stages: a high-velocity ballistic phase followed by a more controlled deceleration phase (Jeannerod, 1981; 1984). There is evidence that the initial phase of the reaching movement is planned in advance of any muscle activity (Marteniuk, Leavitt, MacKenzie & Athenes, 1987). The deceleration phase has been shown to be affected by the force precision required at contact (Marteniuk, MacKenzie, Jeannerod, Athenes & Dugas, 1987) and the spatial accuracy required to perform the task (Gentilucci, Chieffi, Scarpa & Castiello, 1991), as well as by the size of the target object and its distance from the observer (which affect spatial precision -- Fitts, 1954), and by knowledge about object properties derived from past experience (Marteniuk et al., 1987).
Reaching involves more than merely transporting the arm: By the end of the reach, the hand has formed a preshape appropriate to the object and task (Jeannerod, 1984; Pellegrino, Klatzky, & McCloskey, 1989; Wing, Turton, & Fraser, 1986), and the wrist is oriented appropriately to the position and orientation of the object in space.

In short, considerable manipulatory planning has preceded contact, much of it under visual control. Subsequent to contact, cutaneous and kinesthetic receptors provide information about joint positions, the orientation of fingerpads relative to object surfaces, net forces, and incipient and actual slip (Westling & Johansson, 1987). Vision may also play a role subsequent to contact, of course. Foveal vision is used particularly in tasks that require precise spatial positioning, as in threading a needle or mating irregularly shaped parts. For tasks that lack that level of spatial constraint, however, it is possible that vision plays little role beyond the point of contact.

Suppose vision is not available for a manipulatory task -- what might we expect? The answer depends in part on the predictability of the environment. For example, we commonly manipulate controls in an automobile without focusing on them. In such cases, we can capitalize on the fact that the controls are in predictable locations and orientations and that their shape is known. Lovelace (1989) found subjects were actually more accurate at relocate a target position when that target had been defined kinesthetically, as opposed to visually. In this study, the experimenter left the subject's finger to the target object while the subject was blindfolded, or the subject moved his or her own finger to the target object under full visual guidance. All subjects were then blindfolded and asked to move their finger to the target object. The task was consistently performed better when the target had been presented without vision. In a predictable environment, the cost of restricted visual input may not be as severe as one might expect.

When the locations and orientations of objects are less predictable, vision would ordinarily serve to localize the object and determine its orientation, which would guide the reach trajectory, wrist orientation and hand shaping. Without vision, the object would have to be localized by "blind" search, and no preliminary shaping of the hand or orienting of the wrist would occur. Once contact was made, subsequent placement of the fingers, wrist, and arm in the proper manipulatory posture could be guided top-down, by knowledge of the object, and bottom-up, by further search on the object's surface, as necessary. After the manipulator's hand(s) assumed the appropriate position on the object or part, however, the role of vision is less clear, and as we have noted, it is possible that manipulation could proceed under haptic control without penalty for its absence.

**PRELIMINARY STUDY**

A study preliminary to that reported here (Purdy, 1991) asked how well people could perform 40 manipulatory tasks in the absence of vision. (We chose tasks that did not obviously require foveal vision due to spatial constraints.) Each task was performed twice, either with or without vision, by a given subject; across subjects all tasks were performed with and without vision. Our performance measures were pre-contact time -- the time from a "go" signal to contact with the object, and post-contact time -- that required to perform the manipulation.

All tasks were performed successfully to completion, whether or not vision was present. The pre-contact time was actually faster without vision than with it, by about 100 ms. This may reflect the added demands of preshaping during the reach, when vision was present. The advantage for the no-vision condition also no doubt depends on the fact that each object was placed in the same predictable location and hence was relatively easy to locate without vision. With respect to post-contact manipulation time, there was an expected advantage when vision was present. The mean time to perform the manipulation on the first trial was 2.1 s with vision and 3.5 s without it. This difference was reduced on the second trial -- the mean performance time was 1.7 s with vision and 2.7 s without it.

The results of this preliminary study therefore point to a considerable ability to perform common manipulatory tasks in the absence of vision. The cost is on the order of a 50% increase in performance time -- or about 1 s -- for the range of tasks studied. Thus it appears that vision is useful for manipulation, but certainly not essential, when the task is familiar and does not typically require foveation.

**METHOD**

In the study reported in this paper, we turned to a related question: Given that vision is useful, at what points is it especially useful for a simple peg-in-hole manipulatory task? We are particularly interested in tasks with which a person is familiar and with which we have previously been executed under visual guidance. Our approach to this question was to require subjects to perform a peg-in-hole task for a series of trials, while varying the period of time in which they could use vision. For this purpose, the task was divided into 4 stages:

(i) **pre-reach**: from an auditory "go" signal to liftoff from a home position (measured by release of a microswitch);
(ii) **reach**: from liftoff to peg contact (measured with a piezoelectric sensor beneath the peg);
(iii) **grasp**: from contact to lifting the peg off the supporting surface (measured with optical sensors);
(iv) **transport/insert**: from liftoff of peg to insertion into the hole (measured with a reflective object sensor).

In our principal experimental groups, college-age subjects first practiced the task with full vision for 100 trials, then entered into a series of another 100 trials where they had vision during just one of these four stages -- pre-reach, reach, grasp, or transport/insert -- or not at all. There were 10 subjects per group, all right-handed and using the dominant hand to respond. Access to vision was implemented with a set of computer-controlled eyeglasses having liquid crystal shutters as lenses (10 ms to become transparent; 2 ms to become translucent).

Our design is shown in Table 1, which indicates whether vision was present during a trial, and if so, at what stages.

Performance measures included the mean time to execute each stage and the variability in time of performance. These measures were observed for each stage of the task: pre-reach, reach, grasp, and transport/insert. By comparing trials within groups, and by comparing different groups at the same point in the trials, we used these measures to address two questions: (1) Is there a disruption...
Table 1. Design of the study.

<table>
<thead>
<tr>
<th>Group</th>
<th>Trials 1-100</th>
<th>Trials 101-200</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Vision present at all task stages)</td>
<td>(Vision present during some stages</td>
</tr>
<tr>
<td></td>
<td>vision present</td>
<td>but not necessarily all)</td>
</tr>
<tr>
<td>1. Vision -&gt; No vision</td>
<td>vision present</td>
<td>vision absent</td>
</tr>
<tr>
<td>2. Vision at pre-reach</td>
<td>vision present</td>
<td>vision present during pre-reach</td>
</tr>
<tr>
<td>3. Vision at reach</td>
<td>vision present</td>
<td>vision present during reach</td>
</tr>
<tr>
<td>4. Vision at grasp</td>
<td>vision present</td>
<td>vision present during grasp</td>
</tr>
<tr>
<td>5. Vision at transport/insert</td>
<td>vision present</td>
<td>vision present during transport/insert</td>
</tr>
</tbody>
</table>

of performance at the point of transition from trials with full vision to trials in which it is largely or entirely absent? (2) During the trials where vision is supplied for just one stage of the task, what is the impact of having vision at Stage i on performance during Stage j? For example, what is the impact of having vision at the pre-reaching stage on performance during the grasping stage?

Our design allows us to ask how vision at stage i affected performance during stage j for combinations of 5 values of i and 4 values of j. That is, there are 5 stages at which vision is present (none, pre-reach, reach, grasp, transport/insert) and 4 stages at which performance is measured (pre-reach, reach, grasp, transport/insert). We can also ask this question about both mean performance time and variability. In this paper we will concentrate on the effects on mean performance time, but similar effects tend to hold when variability is measured.

At each stage of the task, certain groups would be expected to perform equivalently. First, for the initial 100 trials, any differences between groups performing with vision should be artifacts of individual differences, because vision is present throughout the task during those trials for all the groups. Differences may arise, however, once groups shift to performing with vision during just one stage of the task. Even then, any differences may be present only at particular stages of the task. That is, within a single trial, two groups may be expected to perform equivalently for some subset of stages (where the stages are pre-reach, reach, grasp, and transport/insert). Specifically, two groups should perform equivalently during a trial up to the point where vision is introduced for one and not the other, because up to that point, neither group has vision. (This assumption might be violated if over time, the two groups learned different strategies that affected all stages. However, it is a reasonable starting assumption that is borne out by the data.) Let us now consider the stages of the task at which the groups would be expected to perform equivalently, as shown in Table 2.

As the table indicates, if a measure is taken during the pre-reach interval, then the only group that has vision available is the Vision-at-Pre-reach group. All others prepare for the reach without vision and should perform equivalently. If the measure is taken during the reach interval, then both the Vision-at-Pre-reach group and the Vision-at-Reach group experience vision before reaching, whereas the others reach "blind" and should perform equivalently. The Vision-at-Pre-reach group and the Vision-at-Reach group may or may not perform the reach equivalently, since one has vision prior to the reach and the other at the point of its initiation. If they do perform similarly, it would suggest that having vision just before the reach serves a function similar to that of vision during it. Continuing down the table, it should be apparent that the later within the task that a measure is taken, the more groups will have had some vision prior to the measured stage, and the fewer groups who will be performing without a visual history and hence should be performing similarly.

Table 2 indicates when groups should perform similarly; more important is whether the groups that do not have equivalent visual histories will perform differently from one another. We are particularly interested in the point when subjects shift from trials that are performed entirely with vision to trials where vision is available at just one stage. That point is just after the 100th trial; we will call the trials immediately preceding the 100th the "pre-shift" period and those immediately following it the "shift"

Table 2. Measures for which groups should perform equivalently (V symbolizes the presence of vision).

<table>
<thead>
<tr>
<th>Measure taken at Stage:</th>
<th>Groups that should perform equivalently</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-reach</td>
<td>V-&gt;No-V, V at Reach, V at Grasp, V at Transport/Insert</td>
</tr>
<tr>
<td>Reach</td>
<td>V-&gt;No-V, V at Grasp, V at Transport/Insert</td>
</tr>
<tr>
<td>Grasp</td>
<td>V-&gt;No-V, V at Transport/Insert</td>
</tr>
<tr>
<td>Transport/Insert</td>
<td>(none predicted)</td>
</tr>
</tbody>
</table>
period. To isolate these points, we computed each of our response measures in 25-trial blocks. (This was chosen to provide a reasonably stable data base within each block but still to differentiate blocks; the results are essentially equivalent under a variety of block sizes, and whether or not blocks are treated discretely or with a moving average.) We then compared the mean times during the pre-shift period (trials 75-100) to those during the shift period (100-125). We say that there was a shift effect on the mean performance time if it was greater during the shift period than the pre-shift period.

We also looked at the average within-subject standard deviations, which acted similarly to mean performance time; these data will be reported in a later paper. Also to be reported later are results regarding the end of the series of trials, when subjects might be expected to have recovered from the shift (trials 175-200); we will call this the "recovery" period. We say that there was complete recovery if the mean during the recovery period was equivalent to that during the pre-shift period, there was no recovery if the mean during the recovery period is equivalent to that during the shift period, and partial recovery if the mean is lower during the recovery period than the shift period but does not descend to pre-shift levels.

RESULTS

Figure 1 shows the mean shift effect by group and performance measure. For ease of assessing the effect, in this figure all five conditions are anchored at the pre-shift period to a common value, namely, the mean pre-shift time across the conditions. For a given group, the difference between the two values shown on the graph is set to the shift effect for that particular group. Thus the graph shows the magnitude of the shift effect for each group, with the function displaced vertically for a particular group to the extent that the group's mean pre-shift value differs from the mean across groups. Note that variability around the group mean was typically small relative to the shift effect: The standard error of the mean at pre-shift, computed across the five group means, was 6.4 ms for pre-reach time, 8.0 ms for reach, 2.2 ms for grasp, and 14.1 ms for transport/insert. And as was noted above, any differences between the groups prior to shift would simply reflect individual-subject differences, as there was no treatment differentiating the groups at that point.

Figure 1. Mean time to perform each stage (pre-reach, reach, grasp, transport/insert), during trials 75-100 (pre-shift period) and 100-125 (shift period). The pre-shift time is anchored for each condition at the mean across conditions, and the magnitude of the difference between y-values for a condition corresponds to the observed shift effect. The function labels indicate the group, as defined by the stage with vision after the shift (if any); the y-axis label indicates the measured time period.
The principal findings of this study can be summarized as follows.

(1) Shifting from performance of a peg-in-hole task with vision to performance with no or partial vision has a general tendency to increase the mean performance time for each stage, that is, to show a shift effect. (Often, people show at least partial recovery from the effects of visual withdrawal over time.) There are, however, some notable exceptions to this pattern.

(2) Early stages of the task -- pre-reach and reach -- are minimally perturbed by the shift, if vision is provided just during or before those stages.

It is not surprising that if vision is provided during the pre-reach interval, then the pre-reach time is not affected by withdrawing vision at later stages. After all, the pre-reach is the first stage of the task.

Less obvious is a similar effect on the second stage of the task, reaching. That is, providing vision just at the point of the reach perturbs its performance only slightly. Apparently, people can use the vision provided at the reaching onset to guide the reach itself. It also appears that providing vision just prior to the reach, during the pre-reach interval, effectively guides reaching.

(3) In contrast, the grasp stage is perturbed if vision is provided only during the act of grasping, but less so if it is provided at earlier stages.

In fact, providing vision at the time of the grasp produces performance that is little better than providing it later (or not at all). In contrast, providing vision prior to the grasp -- during the pre-reach or reach interval -- considerably facilitates grasp performance relative to conditions with later or no vision. One reason may be that the hand adopts too wide a preshape when reaching is done blindly as compared to under visual guidance (Wing et al., 1986), hence taking longer to close during the grasp itself. We do not have preshape data to test this interpretation, however.

(4) The transport/insert stage is minimally perturbed if vision is provided during that period; provision of vision at earlier stages produces more perturbation -- but less than if no vision is available at all.

In short, one can see a graded forward effect of early vision, such that vision at initial stages facilitates performance at later ones, but with less and less effect on stages that occur progressively later in the period following the advent of vision. In most cases, vision is useful in the stage at which it occurs, but an important exception is the grasp stage. Vision provided at the point of grasping does not facilitate grasping but does have some carryover to transporting the arm and inserting the peg. Vision during the transport/insert stage also considerably aids performance relative to early vision or none at all. This may reflect visual guidance of the loaded transport component, more than insertion per se.

It appears that grasping is the only stage where vision beginning at the onset of that stage is of little use. This is not surprising, when we consider that the grasp stage is also the fastest, with a mean grasp time pre-shift of approximately 100 ms, as compared to values ranging from 250 to nearly 700 ms for the other stages. A simple reaction time to a light takes about 180 ms (Woodworth & Schlosberg, 1954) -- it seems, then, that reacting to the visual information at the onset of the grasp would take longer than the grasp usually takes to perform. It is not surprising, under the circumstances, that vision at grasp onset cannot compensate for its absence earlier on.

IMPLICATIONS FOR DIRECT MANIPULATION

The present data indicate points at which vision is most valuable in a peg-in-hole task. If one could provide only limited vision, it seems clear that the first and last stages of the task are critical points of application. Provision of vision prior to reaching stabilizes reach onset and duration, and it facilitates grasping as well. (In fact, our recovery data indicate that given practice with the limited-vision version of the task, there is relatively little penalty for having only a brief period of vision prior to grasping. Transport of the arm and application of forces would suffer, however, were vision not provided.)

One might question the generality of our findings with the simple peg-in-hole task with regard to other manipulatory tasks. Given that so many everyday tasks are variants on the peg-in-hole, as was described above, it seems likely that the present findings have considerable generality. That is, the utility of vision at early and late stages of the task might apply to a variety of pre-familiarized tasks in which parts must be localized, grasped, and mated, without undue force or the need for foveation.

IMPLICATIONS FOR REMOTE MANIPULATION

Determination of the contribution of vision during various stages of manipulation is of obvious relevance to remote manipulatory tasks. In a teleoperational environment, vision is high-bandwidth and subject to communication delays. Multicamera systems may be needed to prevent the end-effector from occluding targets for manipulation, increasing the volume of data to be transmitted. If vision is found to be inconsequential for some stages of manipulation, transmission of visual data could be eliminated without decrements, reducing system load.

Again, one might ask whether these results would generalize to tasks in which manipulation is performed by teleoperation, rather than directly by the hand. To the extent that the teleoperator relies on visual sensing to localize objects, reach for them, and control the orientation of the end-effector, we would expect the importance of early visual control in teleoperation to be similar to that of manual manipulation. We would similarly expect late vision to be important for guiding the mating of parts.

If the timecourse of teleoperational events differs substantially from that of direct manual manipulation, however, then effects might differ, but in predictable ways. Consider that we have seen a gradient in the utility of vision, such that events close to a period of vision are helped more than those further from it. Due to time-based decay in visual memory, the graded effect might be more extreme if the task is slowed down, as is likely with teleoperation. For example, vision during the pre-reach might not serve to facilitate the grasp, if the period between these events (the reach) was extended. We have also found an exception to the graded utility of vision, such that vision at the point of grasping did not facilitate the grasp. However, vision at grasping might be more useful for slower grasping, that is, if the teleoperated grasp extended beyond the 100-ms duration of our peg-in-hole task.
References


