*Behavior Research Methods, Instruments, & Computers*
*2002, 34 (3), 383-390*

# "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment

K. H. JAMES, G. K. HUMPHREY, and T. VILIS
*University of Western Ontario, London, Ontario, Canada*

B. CORRIE
*National Research Council of Canada, London, Ontario, Canada*

and

R. BADDOUR and M. A. GOODALE
*University of Western Ontario, London, Ontario, Canada*

We used a fully immersive virtual reality environment to study whether actively interacting with objects would effect subsequent recognition, when compared with passively observing the same objects. We found that when participants learned object structure by actively rotating the objects, the objects were recognized faster during a subsequent recognition task than when object structure was learned through passive observation. We also found that participants focused their study time during active exploration on a limited number of object views, while ignoring other views. Overall, our results suggest that allowing active exploration of an object during initial learning can facilitate recognition of that object, perhaps owing to the control that the participant has over the object views upon which they can focus. The virtual reality environment is ideal for studying such processes, allowing realistic interaction with objects while maintaining experimenter control.

The question of how active exploration of our environment affects behavior has been the topic of a number of research programs in the past (Held, 1965) and more recently (Christou & Bülthoff, 1999; Harman, Humphrey, & Goodale, 1999; James, Humphrey, & Goodale, 2001). A central problem in studying this question, however, is how to allow participants to explore their environments in a realistic way, without sacrificing experimental control. Virtual reality (VR) environments offer a unique opportunity to do this by providing realistic high-resolution three-dimensional (3-D) images that can be modified rapidly "on line" when subjects move a manipulandum. Moreover, the experimenter can monitor and quantify exactly how the observer is interacting with the virtual environment. Although VR technology has improved greatly in the past few years, only a few researchers have been exploring the usefulness of this technology for investigating questions about active exploration of the environment.

This research was supported by a Canadian Institutes of Health Research Student Fellowship to K.H.J. and by research grants from the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research to G.K.H., T.V., and M.A.G. We thank the editor, Jonathan Vaughan, and an anonymous reviewer for their helpful comments on an earlier version of this paper. Correspondence concerning this article should be addressed to G. K. Humphrey, Department of Psychology, University of Western Ontario, London, ON, N6A 5C2 Canada (e-mail: Keith@uwo.ca).

Tong, Marlin, and Frost (1995) investigated the role of active exploration versus passive viewing in the formation of spatial representations of a 3-D virtual environment. The active participants steered and peddled a stationary bike while they traveled through a virtual visual world presented through a head-mounted liquid crystal display. The passive participants were shown a video recording of what the active participants saw. The active participants developed more accurate representations of the spatial layout of this world than did the passive participants. Tong et al. suggested that the tight coupling that normally exists between motor output and visual input facilitates accurate representations of the environment.

Similarly, Christou and Bülthoff (1999) had active explorers control their own movement through a virtual environment, whereas passive observers watched a playback of the active explorers' route. To make sure that both the active and the passive participants were looking at the display carefully, they were required to respond to markers placed in the different scenes that were presented on the display. In a recognition test, all the participants were required to discriminate snapshots of the environment that they had just seen from snapshots of environments that they had never encountered before. The snapshots of the familiar environment were scenes that had contained either markers or unmarked scenes. The active explorers were able to identify unmarked scenes in the familiar environment better than were the passive observers, but there

 

was no difference between the two groups for the marked scenes. The researchers concluded that spatial encoding may be more complete in active explorers.

The studies outlined above have suggested that recognition of scenes and layouts in a VR environment is affected by whether initial familiarization is achieved by active exploration of the information in the scene or by passive observation of this same information. In previous studies (Harman et al., 1999; James et al., 2001), we have shown that the recognition and matching of individual objects on a computer monitor is also better after active exploration than after passive observation. In the present study, we will examine the role of active exploration in the learning of object structure in a VR environment. VR offered an opportunity to use much more realistic object displays. It was important to carry out this extension of our earlier work for two reasons. First, it enabled us to make sure that the active–passive difference was still present even when objects were well rendered and presented in stereo, as would be the case in the real world. Some earlier research that used real objects similar in structure to those used in the present report indicated that binocular (and thus stereo) viewing of the objects led to better generalization to new object views than did monocular viewing (Humphrey & Khan, 1992). Thus, it is possible that the additional information about object structure provided by stereo could eliminate differences in active and passive viewing of the objects. Second, we wanted to confirm that the patterns of exploration that we observed in our earlier studies would also be present in these rich 3-D displays. To this end, participants' reaction times to respond to test objects was recorded, as well as the way in which the participants manipulated the objects in the VR environment.

## METHOD

### Participants

Twenty-six participants 20–29 years of age (mean age, 23) volunteered to participate in the present experiment. All the participants reported normal or corrected-to-normal visual acuity. The participants were paid for their participation.

### Materials

Stimuli were the same computer-rendered grayscale images of 3-D objects that had been used in our previous studies (see Figure 1). When seen through the liquid crystal goggles, the objects subtended approximately 35º of visual angle when the axis of elongation of the object was perpendicular to the line of sight and 20º of visual angle when the axis of elongation was parallel to the line of sight. Therefore, the image size of the objects in the present experiment was substantially larger than it was in our previous experiments (e.g., see Figure 2).

### Apparatus

The VR environment was presented in a 3 × 3 × 3 m CAVE that was developed and integrated by Fakespace systems (Kitchener, ON, Canada).

**Computer**. The computer that controlled the virtual display was a Silicon Graphics Onyx2 with five pipes and 16 processors (4–6 used during application). One pipe drove the front wall and floor; another pipe drove the left and right walls.

**Goggles**. The goggles were Stereographics CrystalEyes liquid crystal goggles (see Figure 2).

**Projection system**. The images were presented in virtual 3-D via four Electrohome (Christie Digital) 9500s projectors. The images of the objects were projected onto the distant wall and floor of the CAVE. Two images were presented (the *left-eye* view and the *right-eye* view) in rapid succession, and each frame of the liquid crystal goggles alternated with the image alternation, producing the illusion of a 3-D object.

**Tracking device**. The tracking device was an Ascension Flock of Birds long-range magnetic tracking system.

**Image control by participant**. The participants were able to control object rotation by moving a hand-held box measuring 7.5 × 5 × 2.5 cm (see Figure 2). This box had four magnetic sensors attached to one 7.5 × 5 cm surface (designated the *top* of the box), allowing accurate tracking of the box's orientation in space. Movement of the box controlled the movement of the objects at a frame rate of 72 frames per second, which allowed accurate mapping of box rotation onto object image rotation. Tracking the rotation of the box allowed the collection and storage of data about how the participants moved the objects during the study phase of the experiment. This allowed the analysis of such data and permitted the *playback* of object rotation to other participants.

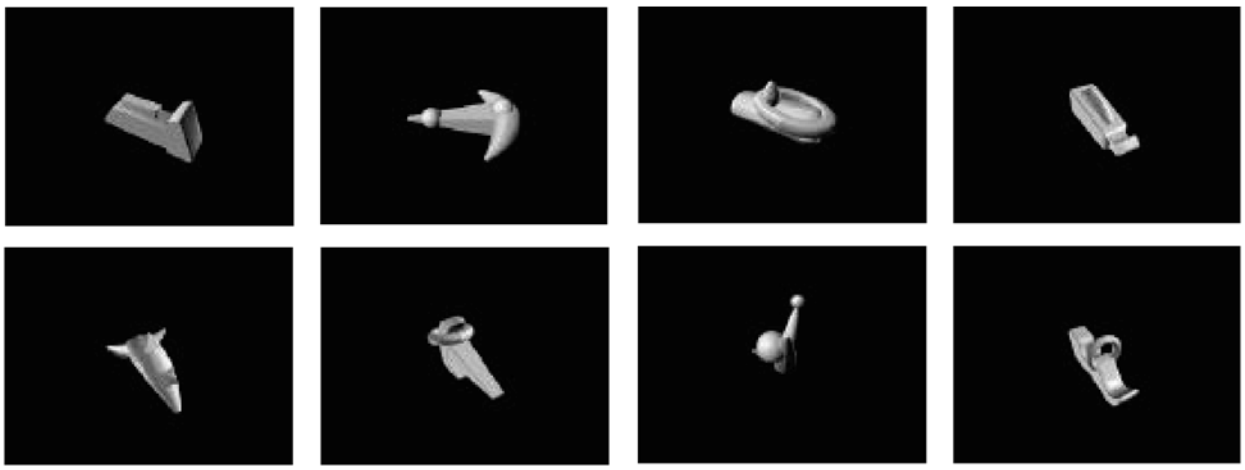

**Figure 1. Examples of the novel, computer-rendered, three-dimensional objects used in the present experiment.**

**Figure 2. Photograph of a participant seated in the CAVE, wearing goggles. The objects appeared in front of the participants as they rotated in three-dimensional space.**

**Image construction and presentation**. The images were initially designed and constructed using Specular Infini-D running on a Macintosh computer. Once designed, the images were converted to a VRML format for presentation. A Silicon Graphics Performer program was used to present the object image. A CAVElib (VRCO) program tracked and recorded the movement of the input device and then rendered the object images on multiple screens.

## Design and Procedure

**Study session**. The design of the present study was the same as that used in Harman et al. (1999). The participants studied half of the objects actively and half passively. The active and passive trials were run in separate blocks, and the order of the blocks was counterbalanced across participants. Within the blocks, the order of object presentation was randomized across participants. During active exploration, the participants were told to study each object carefully from all angles, so that they had a good idea of the object's 3-D shape. They were also told to keep their heads still throughout the study session, and this was monitored by the experimenter. The participants were seated approximately 75 cm from the distant wall of the VR CAVE, wearing goggles to view the objects in apparent 3-D and holding the input box with which they were to control the object movement (see Figure 2). After 40 sec of practice with two objects, the study phase of the experiment began. If the participants started with the *active* exploration condition, an image of a 3-D object would appear in front of them. The initial orientation of the object at the beginning of each exploration trial was determined by the orientation in which the participant was holding the input device when the trial began. Inspection of these orientations indicated considerable variation across participants, as well as within participants across trials, in the *selection* of the object's initial orientation. Throughout a trial, the participants were free to move the input box in any orientation in three dimensions. The rotation of the input box was reflected directly in the rotation of the object (e.g., if the participant rotated the box counterclockwise at a particular speed, the object rotated in the same direction with the same speed). The participant was therefore able to rotate the 3-D object in any orientation in three dimensions for 20 sec. Any translation movements made by the participant were not incorporated into the visual display, whose axes of rotation remained fixed. The temporal lag between movement initiation by the participant and rendering of the new position of the object was never greater than 25 msec. This short lag gave the impression of coincident movement.

The mapping of the input box onto the object was such that the top of the object corresponded to the top of the box and the principal axis of the box was perpendicular to the principal axis of the object. This meant that when the participant held the input box with two hands, the foreshortened view of the front of the object was directly facing the participant. The participants rotated 10 objects actively, after which they were presented with an additional 10 objects that were viewed passively (or vice versa). In the passive condition, the participants did not move the objects; rather, the objects moved on their own. In fact, this movement was a recording of a previous participant's active exploration of that particular object. During this passive condition, the participants were instructed to simply watch the object and try to remember its appearance. There was a 5-sec interval between each 20-sec study trial in both the active and the passive conditions. After all 20 objects had been studied, the test session began.

**Test session**. The test session also took place within the VR CAVE. The participants were still required to wear the goggles but were not required to manipulate the input device. During the test session, static, virtual 3-D images of the 20 study objects and 20 similar, unstudied objects were presented from four different viewpoints individually (Figure 3). Each object was presented as follows: from the front, with the axis of elongation parallel to the line of sight; from the side, with the axis of elongation perpendicular to the line of sight; from an intermediate front view, which was halfway between the front and the side views; and from an intermediate back view, which was halfway between the back and the side views. Therefore, the participants were required to respond to 160 images

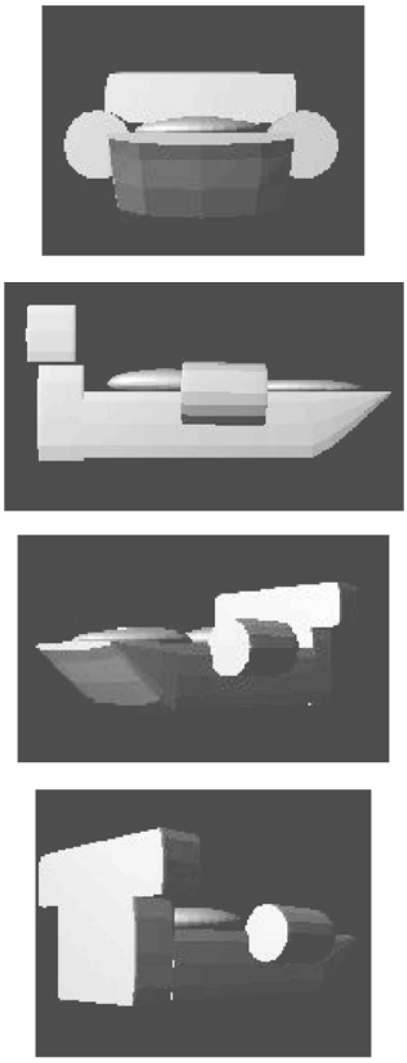

**Figure 3. Examples of the test angles of the objects used in the present study. From top to bottom: front "foreshortened" view, side view, intermediate front view, and intermediate back view.**

of objects (40 objects from four angles) that were presented in random order. As soon as an object appeared, the participants were required to indicate, as quickly and accurately as possible, whether they had seen that object in the study session or not. The participants responded by pressing buttons on a wireless mouse, responding with the left thumb for *not seen* and with the right thumb for *have seen*. The object remained in view until they responded.

## RESULTS

### Test Session

Three 2 (study condition, active or passive) × 4 (test angle, foreshortened, intermediate front, side, or intermediate back) repeated measures analyses of variance (ANOVAs) were run on the data. One ANOVA was run on the response latencies of studied objects that had been correctly recognized. A second ANOVA was performed on accuracy (a correct response to a target object, or a hit) as a check to make sure there were no accuracy differences that would be indicative of a speed/accuracy trade off. The final ANOVA was conducted on a measure of sensitivity [(hits + correct rejections)/2; see Snodgrass & Corwin, 1988]. Four participants were excluded from experimental analysis on any measure because their sensitivity scores did not reach our criterion of 60% correct responses (see Snodgrass & Corwin, 1988).

**Response latencies**. The ANOVA revealed a significant main effect of study condition [$F(1,20) = 5.15$, $MS_e = 207{,}947.3$, $p < .05$]. Objects that were studied actively were recognized faster than the objects that were studied passively (see Figure 4). There was no significant effect of test angle and no test angle × study condition interaction. An analysis was also run using object as the unit of analysis, and this ANOVA revealed a significant effect of study condition [$F(1,18) = 5.49$, $MS_e = 199{,}754.1$, $p < .05$]: The participants who studied a given object actively ($M = 2{,}441.51$ msec, $SE = 139.20$ msec) responded faster than the participants who studied that object passively ($M = 2{,}587.61$ msec, $SE = 188.90$ msec).

**Accuracy**. This analysis revealed no significant effects of either study condition or test angle on response accuracies for the participant analysis or for the object analysis.

**Sensitivity**. The ANOVA run on the sensitivity data revealed no significant effect of study condition on this variable. There was, however, an effect of test angle [$F(3,57) = 8.49$, $MS_e = 73.3$, $p < .001$]. The intermediate front angle was recognized most accurately, using this sensitivity measure ($M = 76.05$, $SE = 1.94$), followed by the intermediate back angle ($M = 72.52$, $SE = 1.94$), the front angle ($M = 69.25$, $SE = 2.04$), and the side angle ($M = 67.38$, $SE = 2.22$; see Figure 5). Post hoc analyses (Tukey, $p < .05$) indicated that the intermediate front view was recognized with greater sensitivity than were both the front and the side views, and the intermediate back view was recognized with greater sensitivity than was the side view.

### Exploration Session

The exploration data were collected during the study session for the objects that the participants actively manipulated. To investigate how the participants explored these objects, the amount of time the participants spent viewing the objects from particular angles was plotted. As is depicted in Figure 6, the participants spent most of their time exploring the objects from specific viewpoints, largely ignoring other viewpoints. In general, the participants focused on the *plan* views of the objects, which can be defined as the front, side, and back views of the object, rotating the objects about the vertical axis (for a detailed description of object views, see James et al., 2001).

The sequences of exploration that the participants used were also investigated. Although there was no consistent order in the sequence of views that the participants generated, when the participants looked at some of the views, they made small movements around that view, sometimes for extended periods of time (Figure 7). For example, if a participant was looking at a 0° (front) view, he or

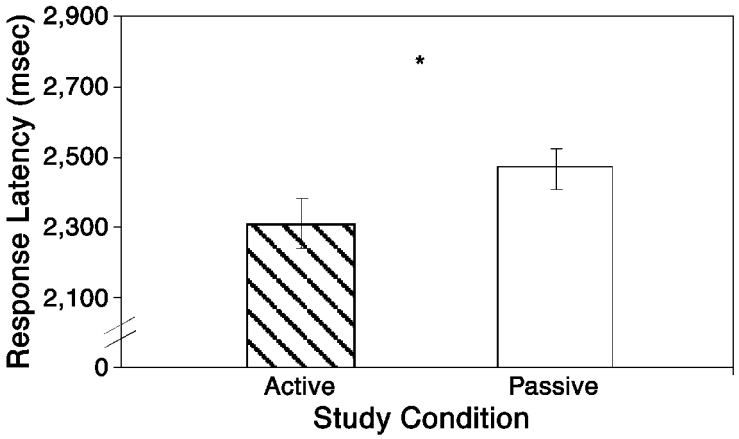


**Figure 4. Main effect of the exploration condition. Objects explored actively are recognized faster than objects viewed passively. The asterisk indicates significance at the $p < .05$ level, and error bars indicate the standard error of the mean.**

she actually tended to move the object back and forth around the 0° point. We called these small movements *wobbles* (defining them as movements with a range of less than −22.5° to + 22.5° around a central point for at least 4 sec). No individual wobble lasted longer than 10 sec. The number of wobbles was compiled across all participants and all objects and revealed that the participants made such movements more around plan views (front, total wobbles = 174; sides, total wobbles = 64; back, total wobbles = 35) than they did around intermediate views (intermediate front views, total wobbles = 23; intermediate back views, total wobbles = 16; see Figure 8).

## DISCUSSION AND CONCLUSIONS

The results of the present study provide the first demonstration that active control of visual input in a VR environment during perceptual learning leads to more efficient object recognition. The participants who, during study, actively rotated 3-D novel objects in the VR CAVE later recognized these objects more rapidly than did the participants who had passively viewed exactly the same sequence of images of these virtual objects during initial study. In addition, while exploring such novel objects, the participants concentrated on particular views.

Although other studies have demonstrated that active exploration can improve scene recognition via the detection of changes in a stimulus array (Christou & Bülthoff, 1999), the present study provides convincing evidence that fundamental mechanisms mediating object recognition can be influenced by active exploration. In other words, active control over the way in which the different views of an object are revealed leads to faster recognition. It is not clear from this result, however, what factor or fac-

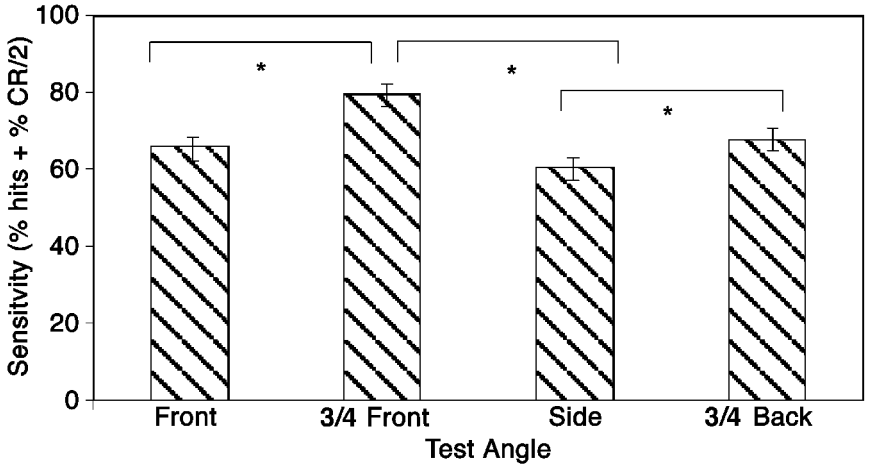


**Figure 5. Main effect of test angle on sensitivity. The intermediate views are recognized with greater sensitivity than the plan views. Asterisks indicate significance at the $p < .05$ level, and error bars indicate the standard error of the mean.**

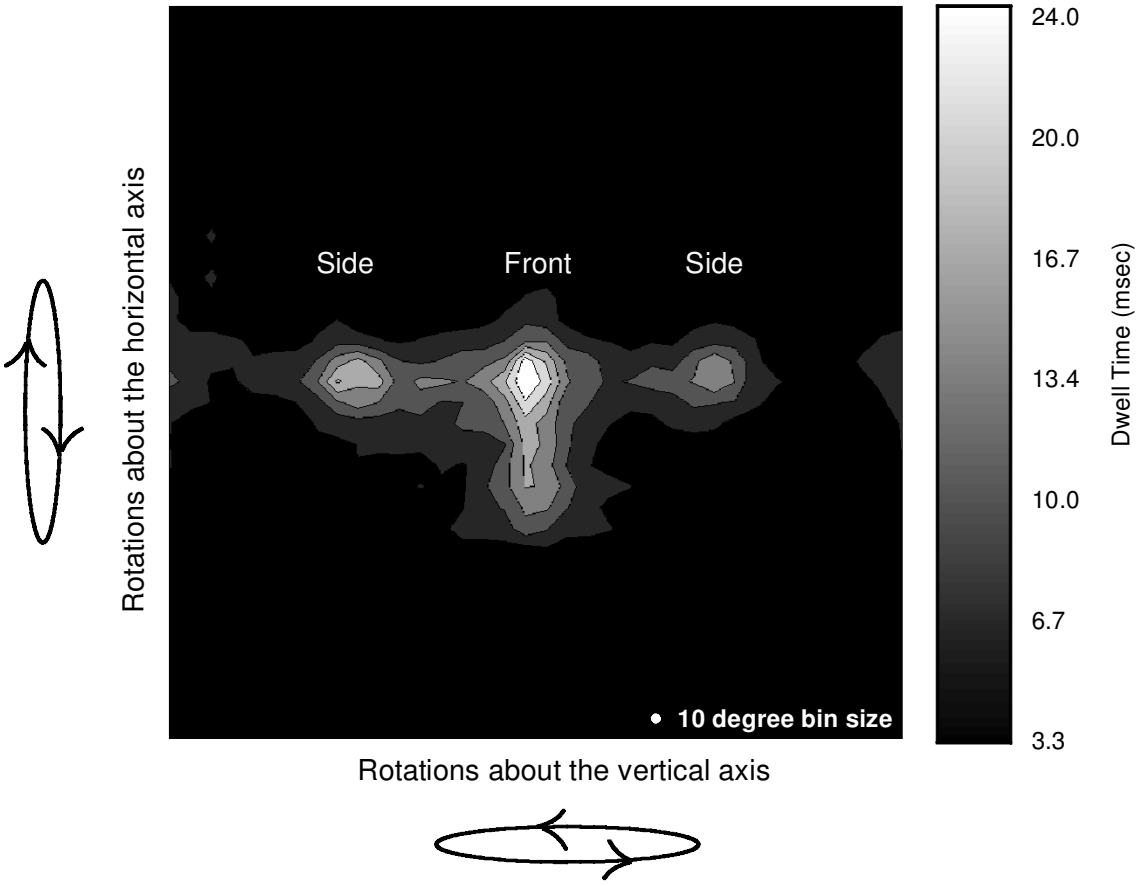


**Figure 6. A contour map depicting dwell times during the exploration (study) phase. This map is a depiction of the viewing space and represents the mean of all the actively explored objects by all the participants. Lighter areas indicate a long dwell time, whereas darker areas indicate a short dwell time. The top half of the map depicts the dwell times of objects in the upright orientation; the bottom half depicts the dwell times of inverted objects. Therefore, the *y*-axis represents rotations about the horizontal axis. The *x*-axis represents rotations about the vertical axis. The spatial resolution of the dwell time calculation was 10°. Therefore, the time scale refers to dwell time within a 10° bin. The total dwell time for all the peaks (lighter regions) was 9.2 sec, close to half the total exploration time.**

tors lead to the more efficient recognition in the active group. For actively viewed objects, both kinesthetic and visual information about object orientation were available, whereas for passively viewed objects, only visual information was available. It seems unlikely, given our earlier results (Harman et al., 1999; James et al., 2001), that kinesthetic information per se could account for the active/passive difference. In the previous research, a track ball was used to control object orientation, and unlike the box in the present experiment, such an input device would not give kinesthetic information about object orientation. Nevertheless, we found the same active/passive difference as in the present report. To be certain, though, that kinesthetic information is not playing a role in the active/passive difference will require further experiments.

To explain the active/passive difference in recognition efficiency, we have hypothesized that direct manual control over the sequence of views provides efference copy and/or proprioceptive information that helps to integrate the different views by allowing the participants to anticipate the upcoming view and relate it to the previous view (Harman et al., 1999). In addition, active exploration could allow the participants to test "predictions" about the expected deformations in the image that would occur when the object is rotated in a particular way (Harman et al., 1999). In an unpublished experiment (James, 2001), we have found that making the relation between track ball movement and movement of the object on the screen unpredictable during object exploration eliminates the active/passive difference we have reported here and in our earlier research. Thus, it seems that the advantage active movement confers arises only if participants can predict the consequences of their actions.

The participants spent more time inspecting certain views of the objects, as compared with others. In the active exploration condition, the participants tended to rotate the objects mainly around the axis that was perpendicular to the main axis of elongation of the objects. As a consequence, the object was rotated so that it moved between a fully elongated view to a completely foreshortened view.

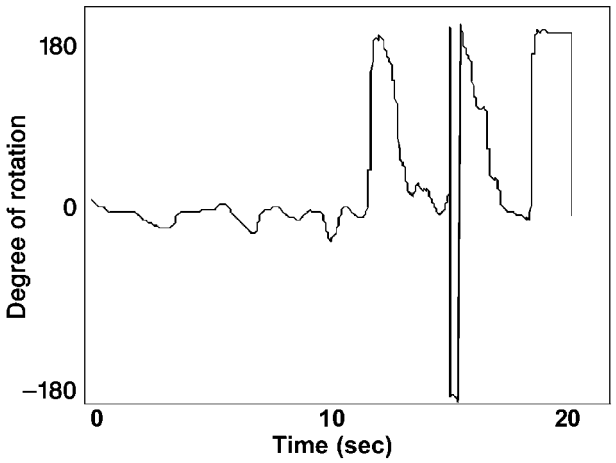


**Figure 7. A depiction of an individual "wobble." A wobble is defined as a back-and-forth motion around a central point. The motion was restricted to ±22.5° around the central point. Wobbles lasted a minimum of 4 sec.**

The participants also treated the flat surface of the object as the *bottom* and generally kept the objects oriented so that this surface was always face down. Of course, the geometry of the manipulandum almost certainly affected the strategy that people used to rotate the object. It is likely, for example, that the participants would not have often turned the input box upside down. Nevertheless, these possible biases in hand posture and grasp cannot be the whole story, since the same tendency to maintain the object in an "upright" position and to rotate it around the vertical axis was evident in our previous experiments, in which participants used a trackball (Harman et al., 1999; James et al., 2001).

Thus, it seems that both the geometry of the object and the convention of a top–bottom rule appear to be driving the inspection strategies. The participants in this experiment, as before, constrained their viewing even more by concentrating on only a few particular views around the *primary* (or chosen) axis of rotation. In particular, the front and the side views received the most looking time. These two views represent ones in which the primary axis of elongation of the object is either perpendicular or parallel to the line of sight. Again, some of this concentration could have been a reflection of the constraints of the input box, particularly the high dwell times on the front view. Because the wires connecting the sensors with the movement recording device emanated from the nearside of the box, the participants might have been reluctant to rotate the box too much away from the *primary* position. But again, it is important to point out that the same concentration on plan views was present in our earlier experiments in which participants used a trackball (Harman et al., 1999; James et al., 2001). Furthermore, Perrett and colleagues, who used a range of manipulanda, also found that when participants explored objects, they concentrated their inspection time on front and side views whether the objects were potatoes (Perrett & Harries, 1988), heads (Harries, Perrett, & Lavender, 1991), or machine-tooled "widgets" (Perrett, Harrries, & Looker, 1992).

Perrett and his colleagues have proposed that observers concentrate on plan views, such as front and side views, because these views are *unstable* and can be thought of as singularities in the viewing space of an object. These are the views in which there is the greatest amount of change in the visibility of the object features as the object is rotated by a small amount. Inspection strategies that concentrate on such views would be important in coding these particular views. We can see now why participants would not dwell on any particular intermediate views. The intermediate views are all perceptually similar: All the major features of the objects are visible over a wide range of image projections. Thus, participants do not need to concentrate on one particular intermediate angle, because of the high similarity among many of the successive images. This may explain why participants deviate a little from side to side when viewing a plan view; larger excursions would not produce much more information than they already have. According to this hypothesis, the back-and-forth movements, or wobbles, around plan views provide the maximum amount of information that is needed for accurately storing an object representation. That is, wobbling around plan views allows the observer to code the unusual plan view, as well as the maximum changes that occur with small deviations from that plan view.

It is interesting, and perhaps somewhat puzzling, that although the participants spent more time on plan views during exploration, intermediate views were recognized best according to our sensitivity measure. This result was

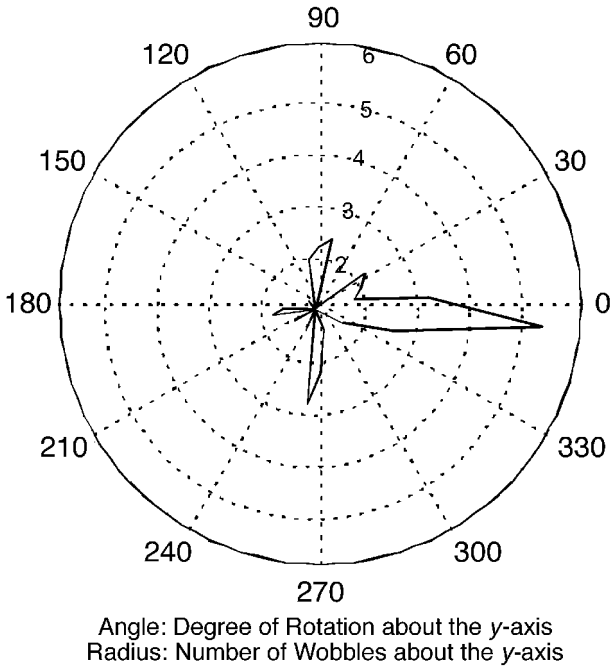


**Figure 8. A polar plot of the number of wobbles that were measured at each orientation of the objects. More wobbles were made around the plan views than around the intermediate views.**

also found in an earlier study (Harman et al., 1999). One possibility is that the *intermediate view* is actually seen for a substantial period of time during study. That is, although there is no *peak* in the dwell time map for a particular intermediate view, if one were to aggregate all of the time spent on intermediate views, it would be substantial. It may be that many of these intermediate views are qualitatively similar enough in terms of visible components and other features that they are essentially, from the point of view of the recognition system, the *same* view (for further discussion, see Harman et al., 1999; James et al., 2001; see Tarr & Kreigman, 2001, for a lucid discussion of what defines a view). This possible explanation for the greater sensitivity to intermediate views is, of course, speculative and needs to be examined empirically.

Although the finding that the decision latency was faster in the active condition than in the passive condition replicates our earlier research (Harman et al., 1999), there does appear to be a difference in the absolute latencies between the two experiments. The overall decision latencies in the present study were substantially longer than those in the earlier study. One factor that could explain this difference is that the test objects were much larger in the present study than in the study by Harman et al. The visual angles subtended by the objects in the present study were almost four times the visual angle of the objects in the previous study. To view larger objects would presumably require more eye movements, and this increase in the number of eye movements may increase the decision latency.

The replication of our previous studies (Harman et al., 1999; James et al., 2001) underlines the robustness of the effect of active exploration on object recognition. Thus, even with well-rendered, large objects presented in full stereo, the results still demonstrate that active exploration facilitates object recognition. In other words, the rich visual information that was available did not wash out the difference between active and passive performance. The absence of any differences between the original study (Harman et al., 1999) and the present study begs the question, of course, as to whether or not anything is gained by conducting these types of experiments in a virtual environment, as compared with using a simple computer display. The advantage of the VR immersion environment, however, is that the kinds of displays and the richness of those displays are essentially limitless. Thus, one could imagine having people walk around large displays, using a head-mounted system or even using devices attached to the hands that give tactile and force feedback about graphically rendered objects. In short, immersion technology offers an opportunity for more realistic interfacing with the virtual environment, which in turn leads to a more accurate understanding of the observer's interaction with their real environment. Moreover, the present study has implications for the increasing use of VR displays for training and education. With the development of highly rendered displays of such structures as organic molecules, anatomical structures, and architectural models, for example, it might be useful to allow the student to actively control the views of the objects that are seen.

## REFERENCES


Christou, C. G., & Bülthoff, H. H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition*, **27**, 996-1007.

Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, **9**, 1315-1318.

Harries, M. H., Perrett, D. I., & Lavender, A. (1991). Preferential inspection of views of 3-D model heads. *Perception*, **20**, 669-680.

Held, R. (1965, November). Plasticity in sensory-motor systems. *Scientific American*, **213**, 84-94.

Humphrey, G. K., & Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Canadian Journal of Psychology*, **46**, 170-190.

James, K. H. (2001). *Controlling what we see: The effects of active exploration on object recognition.* Unpublished doctoral dissertation, University of Western Ontario.

James, K. H., Humphrey, G. K., & Goodale, M. A. (2001). Manipulating and recognizing virtual objects: Where the action is. *Canadian Journal of Experimental Psychology*, **55**, 111-120.

Perrett, D. I., & Harries, M. H. (1988). Characteristic views and the visual inspection of simple faceted and smooth objects: "Tetrahedra and potatoes." *Perception*, **17**, 703-720.

Perrett, D. I., Harries, M. H., & Looker, S. (1992). Use of preferential inspection to define the viewing sphere and characteristic views of an arbitrary machined tool part. *Perception*, **21**, 497-515.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34-50.

Tarr, M. J., & Kreigman, D. J. (2001). What defines a view? *Vision Research*, **41**, 1981-2004.

Tong, F. H., Marlin, S. G., & Frost, B. J. (1995, June). *Cognitive map formation in a 3D visual virtual world.* Poster presented at the IRIS/PRECARN workshop, Vancouver, BC.


(Manuscript received May 20, 2001;
revision accepted for publication May 5, 2002.)