Realism in 3D Object Recognition

The influence of object realism on human 3D recognition

F. Subklew (s0095133)

f.subklew@student.utwente.nl

Supervisor:

Drs. F. Meijer

Dr. E. L. van den Broek

#### Abstract

The role of realism in recognition of 3D objects in virtual reality was examined. It was hypothesized that users of VR would benefit from realistic objects and active exploration of objects in general. Furthermore was tested, whether visuo-spatial ability can work as a moderator for the kind of exploration that provides the best recognition. Through an experiment, we found that object realism has much influence on human 3D recognition. People are faster and more accurate in remembering realistic objects in contrast to abstract ones. Furthermore were the participants the most accurate for unstudied objects and didn't benefited significantly from active exploration, when studying objects. This process was shown to be moderated by the object condition (realistic vs. abstract). For people, studying the realistic objects, the form of exploration was of less importance.

## Introduction

Our lives become more and more digital through digital cameras, mp3 music and communications over the Internet. With our computers, we manage our photos, home videos, music, e-mails, our bank accounts and entertain our selfs with games that take us to different worlds, different virtual realities (VR), be it a realistic or fantasy one, where we can interact with the computer-simulated environment. Well known examples are the online game 'World of Warcraft' and the online world 'Second Life' (See Figure 1a). But VR can be found in more areas than just entertainment. More and more working fields are taking advantage of the new possibilities that result from VR. The medicine, gas- and oil-industry, military and engineering (Burdea et al., 2003) have found the technology very useful. Whereas especially the last field, product design, is very interesting. The use of VR in the design process of products speeds up the development cycle, by replacing the time- and money-consuming task of building and testing physical prototypes. It allows even to look at architectural concepts and the future neighborhood, as can be seen in the 'i-CONE' of the Fraunhofer Institute for media communication (See Figure 1b). The i-CONE is already more than a prototype and used to virtually test whether gas- and oil-pipes are leak-proof.



Figure 1a and 1b. A typical screenshot of the virtual world 'Second Life' (left) and a presentation in the 'i-CONE' of the Fraunhofer Institute for media communication (right).

A complex and virtuosic VR is certainly very expensive through uses of special screens and much processor-power to simulate entire virtual worlds. Computing only a limited set of features makes Virtual Reality more time and cost efficient and thus even accessible to smaller companies to benefit from this advantage. The question is, what features can be disregarded when building a cheap virtual reality. The first points that come in mind to cut down on are the graphics, screens and effects (for example sound). So, using smaller screens, renounce spectacular effects and using simple graphics, should economize the VR. But how about the most important feature of interactivity? How necessary is active interaction with objects in the virtual world and how much difference does the kind of objects make? To answer these questions it is necessary to take a look into the fundamental processes of human object recognition, that explain how people perceive virtual reality.

Irving Biederman (1978) has studied human image understanding and developed in 1978 the recognition-by-components theory (RBC), that states that perceptual recognition of objects is basically a process of segmenting the objects objects at regions of deep concavity into simple geometric components, so-called geons (for 'geometrical ions'), like cylinders, blocks and cones. Geons are view invariant, discriminable, noise resistant and parsimonious, in other words: they look the same from most angles, are difficult to confuse one for another, can be identified under less-than-ideal conditions and can create complex representations. The determined components are further matched to object representations and thereby identified as a specific object.

In 1999, Harman et al. investigated the involvement of actions in form of active and passive exploration in VR on human object perception. They created three-dimensional (3D) objects for their experiment, based on Biederman's findings, by attaching 'geon-like' parts onto a bigger central body. Halve of their participants could rotate the objects free for 20 seconds about any axis, whereas the other participants where watching a 20 sec recording of the previous subject's active

## Realism in 3D Object Recognition

exploration. The recognition was tested by the use of four images of each three-dimensional object from four different angles whereupon the subjects had to decide whether they had studied the objects earlier in the experiment as a 3D model. Harman et al. discovered that observers who actively rotated the 3D objects on a computer screen later showed more efficient visual recognition than observers who passively viewed the same objects.

These findings were confirmed by Meijer et al. (in preparation) in an analogue experiment where the participants made their decision on a mental rotation test. In addition, they showed, that the participant's ability to form mental representations, the so called visuo-spatial ability (VSA), modulates the effect of active exploration. Subjects with low VSA benefit from an interactive exploration of objects opposed to subjects with a medium or high VSA, who are not sensitive for the type of exploration.

For an experiment in the line of Harman et al. (1999) and Meijer et al (in preparation) in which participants study objects both actively and passively, we expect people to benefit from active exploration dependent on their visuo-spatial ability. However in contrast to Harman et al. and Meijer et al., who used only abstract, so called 'novel' objects for their experiments, we think that people are acquainted with most products in a general form and thus quite unlikely to come across very novel objects in a virtual design process. From the stages in object perception that were described by Biederman (1987), we expect that realistic objects are recognized more accurate and faster, because of their better match with mental representations, than abstract objects.

# Method

#### *Participants*

Sixteen students of the Lephana Universität Lüneburg participated in the experiment (9 women and 7 men). All students participated on a voluntary basis. Their age ranged from 20 to 26 years

(average: 22). All participants had normal or corrected to normal visual acuity and no known neurological or visual disorders.

# Materials and apparatus

A set of 48 different three-dimensional objects was created in the 3D modelling solution Cheetah3D for Mac OSX. Each object consisted of three to five 'geon-like' components (Biederman, 1987). Half of the objects were built to look realistic, familiar to the participants everyday life. The other 24 were produced by deconstructing each realistic object into its parts and putting them back together to an object that arouses no association with anything real. The objects were colored gray and were equal in illumination and luminance, presented on a black background (see Figure 2) at a viewing distance of ca 60 cm.



Figure 2. A realistic object (left) and an abstract object (right), as constructed for the experiment, using the same (in this case) five geons. For a list of all objects, see appendices.

These objects formed the basis for both phases of the experiment. In the study phase they were used i their three-dimensional form and for the test phase a picture was taken of each object from four different angles ( $0^{\circ}$ ,  $90^{\circ}$ ,  $180^{\circ}$  and  $270^{\circ}$ ).

The experiment ran on a MacBook C2D 2GHz in a Windows XP environment, installed via Boot Camp Beta 1.3 (Apple, Inc.). The Computer had a 13" display with a native resolution of 1280 by 800 pixels, but the experiment ran on 1024 by 768 pixels, so only a part of the screen was used (see Figure 3).



Figure 3. The computer screen, as used in the experiment. Because the experiment ran on 1024 by 768 pixels, only a part (in red markings) of the whole screen (blue) was used.

Authorware 7.01 (Macromedia, Inc.) was used for the presentation of the objects in the study phase, which was enabled by the Cortona VRML Client 5.1 (Parallel Graphics, Inc.) plug-in, using the ActiveX option in Authorware. The interaction with the objects in the (inter)active condition was enabled through a one button Mouse, connected to the computer. E-Prime 1.1 (Psychology Software Tools, Inc.) was used to collect the response data from the participants through a standard keyboard.

# General procedure and design

Before the experiment started, participants were tested on their visual-spatial ability (VSA), using a mental rotations test (Vandenberg & Kruse, 1978). Next, they perceived an example object, which they studied both interactively and passively. Following, the experiment was run with study phase and test phase, which are described further in the following sections.

Subsequent to the experiment on the computer, the participants were asked to look through a list of 48 pictures (on picture of every object) and decide if they recognize a real object in the

images, and which one. This was done to verify whether the objects were designed well in terms of realism.

## Study phase procedure

The experiment started with a study phase in Authorware, split in four blocks in which participants studied the objects in two exploration conditions. In passive exploration conditions, participants did not interact with the object but observed it while rotating 360° over the x-axis, y-axis, and z-axis. In (inter)active exploration conditions, participants were free to control the rotation of every object in any direction with the mouse. The block order was counterbalanced (See Figure 4).

Subjects	Condition Block Order
1, 5, 9, 13	ΑΡΑΡ
2, 6, 10, 14	ΡΑΑΡ
3, 7, 11, 15	ΑΡΡΑ
4, 8, 12, 16	ΡΑΡΑ

Figure 4. The counterbalancing of the conditions (inter)Active (A) and Passive (P) in the block order for the study phase.

Each block consisted of six objects, 3 realistic and 3 abstract ones, randomly selected from the set of 48 objects. Each object was studied for 30 seconds with a 1 second interval between them.

## Test phase procedure

After the study phase, the computer automatically switched to E-Prime for the test phase. The test phase was also split in four blocks. Per block 48 2D images of the objects were shown. The participants were required to determine as quickly and accurately as possible for every picture whether they had seen the shown object earlier in 3D, or not. The pictures remained visible until a key-response was given: M for "not seen earlier in 3D", Y for "seen earlier in 3D". The response

latency and accuracy were recorded. After each block the participants received feedback about their accuracy and reaction time averages and a notification to improve their scores, when needed. In total, 192 responses were given in the whole experiment.

#### Results

To test the hypotheses, two repeated-measures 3x2 (active, passive, and no-study condition versus abstract and realistic objects) ANOVAs were run. One for the accuracy data and one for the reaction times, each times with VSA group as between-subjects factor. Two percent of the whole data was excluded from the analyses by discarding two times all trials (per block and condition) that exceed tree times the standard deviation. The data distribution over the study conditions active, passive and no-study was 25% to 25% to 50%.

The analyses of the accuracy revealed a main effect of both the study condition (F(2,15) = 30.96, p < .001) and the object condition (F(1,15) = 51.07, p < .001). Participants were more accurate in the no-study condition (M = .97) than when they had studied the objects actively (M = .84) or passively (M = .76). The response of the participants was more accurate, when the objects were realistic (M = .93) rather than abstract (M = .84). The ANOVA of the reaction times confirms the main effect of the object condition (F(1,15) = 33.75, p < .001), but the study condition showed no significant effect. Participants made their decisions faster for realistic objects (M = 1.015 ms) than abstract objects (M = 1.214 ms).

Furthermore both ANOVAs (accuracy and reaction time) revealed a significant interaction effect of study and object condition ( $F_{accuracy}(2,15) = 26.29$ , p < .001), ( $F_{reaction time}(2,15) = 7.566$ , p < .006) (See Figure 5). The object condition (abstract vs. realistic) had more influence on the reaction time in the explore conditions (active vs. passive) than with no study.

A regression analysis revealed no significant correlation of reaction time and accuracy (F(1,15) = .004, p > .94), so differences in response latencies cannot be attributed to a speed-accuracy trade-off (See Figure 6)



Figure 5. The interaction effect of the study and object condition, as found in the ANOVA of the reaction times (RT) in milliseconds (left) and in the ANOVA for the proportion correct responses (accuracy/ACC) in percent (right).



Figure 6. The relation of reaction time (RT) and the proportion correct responses (accuracy/ACC). Each cross marks the mean RT and mean ACC for one of the participants. No significant regression, that would point to a speed-accuracy trade-off, could be found.

#### Discussion

The presented study was interested in the difference in recognition for abstract versus realistic objects under active and passive study conditions. As we expected, were the participants decisions for realistic objects faster and more accurate than for abstract objects. The advantage that we are familiar with realistic objects like eyeglasses, cups, hammers or a wine glass, makes it easier to decide whether we saw one of them recently or not. While we recognize both the abstract and the realistic objects by their components (Biederman, 1987), the matching to the mental object representations works faster for realistic objects. The more congruent an object is with a representation, the faster the identification and thus the decision.

Although Harman et al. (1999) and Meijer et al. (in preparation) constantly report that active exploration leads to more efficient object recognition than passive or no exploration, we found an opposite effect. The participants in our study were the most accurate in the no-study condition, followed by active and passive exploration. The advantage of active exploration over passive exploration could not be replicated in our study. The fact that our participants were the most accurate in their decision for objects, they had not studied before, results most likely from the different testing procedure we used in contrast to Meijer et al. (in preparation). In their experiments the effect of the study condition was tested with a comparison task, where the participants had to decide whether two presented pictures showed the same object. Whereas our study used a memory task, based on the question, whether an object was explored earlier in the experiment or not. Certainly people are quite good at recognizing objects they do not know, but have problems making a decision on a comparison task, based on mental rotation, when an objects appears, they had not studied before.

For VR, the no-study condition has no relevance, for the interest lies in the question whether active exploration is needed to provide a better experience for the user in contrast to a passive view of a product. As stated earlier in this study, we think that people are acquainted with most products in a basic form and thus quite unlikely to come across very novel objects in a virtual design process. We found as expected the lifelike objects to be recognized faster and more reliable compared to abstract items.

Very interesting is the interaction of the forms of exploration and the kinds of objects. The data reveals that the varieties between the three study conditions are less in evidence when people were studying realistic objects than with abstract ones. In a virtual environment for product design, people are not developing completely new and, to this point, unknown products, but modify already existing product categories; hence, the form of exploration and interaction is of minor importance. A car, for example, is easy detected as such, even if it has a different design through new shaped parts around the wheels, the wheel are still left. As long as the observer is familiar with the item, there is no need of interaction. This effect will be the same for an engineer, observing just a very specific part of a car. Opposite to a layperson, that has never seen an object like this, the engineer has a coherent mental representation, in other words: he knows, what he's seeing and does well without direct interaction.

The rare findings for the no-study condition need further investigation to make sure whether flaws in the experiment procedure (like inconsistency) were the reason or if there exist a cognitive process, that explains, why decisions are easier for unstudied objects than for studied ones. Further research for the interaction effect of object- and study condition to prove the findings of this study is necessary. If confirmation can be found, interaction could be another feature to ignore in a cheap VR. For now we can concluded that as long as VR works with familiar forms and objects, no extra capital investment in favour of interaction is needed.

## References

Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, 94, 115-147.

Burdea, G. C., Coiffet, P. (2003). Virtual Reality Technology, Second Edition. Wiley-IEEE Press

Fraunhofer IAIS. (n.d). Retrieved on November 11, 2007, from

http://www.iais.fraunhofer.de/643.html?&L=1

- Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, *9*, 1315-1318.
- Meijer, F., & van den Broek, E. L. (in preparation). The impact of interactive exploration on human 3D object recognition.

# Appendices



