Study of stereoscopic shape perception for virtual design

Laure Leroy
CAOR - Mines Paristech
60 Boulevard Saint Michel 75006 Paris
laure.leroy@ensmp.org

Philippe Fuchs
CAOR - Mines Paristech
60 Boulevard Saint Michel 75006 Paris France
philippe.fuchs@ensmp.org

Alexis Paljic
CAOR - Mines Paristech
60 Boulevard Saint Michel 75006 Paris France
alexis.paljic@ensmp.org

Guillaume Moreau
Ecole Centrale Nantes, CERMA - IRSTV
rue de la Noë - 44300 Nantes France
guillaume.moreau@ec-nantes.fr

Abstract

Virtual environments are increasingly used to replace real prototypes, but are we sure that shape perception is the same in both environments, i.e. can we affirm that the shapes that are designed and validated in virtual environments will be perceived the same once a physical prototype has been built? This study reports on a series of perception tests in which two virtual shapes (a sphere and a shape without symmetry or obvious monoscopic cues) will be compared to their respective actual shapes. We compare the influence of the head tracking, adjusting interpupillar distance, the position of the virtual object in relation to the screen and the orientation of the modified shape. We show that if all these parameters seem to have an influence, head tracking appears as the most important for good perception.

1. Introduction

Virtual reality techniques are used in the design process of manufactured objects and especially in the automotive industry. One of the advantages is that designers can modify the virtual objects as much as they wish. When an object suits their needs, they can then decide to build a real object, using the virtual object as a template. But the important question remains: will the virtual object as seen in an immersive system be perceived as the real object created ? In other words, the issue is whether immersive design (and especially stereoscopic vision) can create a perception of a virtual object which corresponds closely enough to the resulting real object.

The perception of a shape depends, among other things, on the relative size in three dimensions (length, height, depth) of the object, so we may wonder if these proportions are being preserved. There are many studies that focus on depth perception in stereoscopy, but very few on these relative dimensions.

Therefore, in this study we try to define what the parameters that influence the perception of a virtual object are. Is the stereoscopic vision actually more effective than monoscopic vision? To what extent can head tracking improve its performance? Below, we will try to answer all these questions. Does having the virtual object on the screen to minimize disparities influence the judgement of the user?

The rest of the paper is organized as follows: Section 2 will present our hypothesis, section 3 will summarize related work. Then, in section 4, we will present methodological elements about our experiments and the tests themselves. Section 5 will be devoted to results analysis.

2. Hypothesis

2.1. It is preferable to have the object on the screen

We believe it is preferable to have the main virtual object spatially positionned on the screen rather than in front or behind it. Indeed, we know that the more horizontal disparities we have, the greater the eyestrain becomes. It is reasonable to believe that the perception of shapes also deteriorates where disparities increase. But when the subject is tracked, placing the object in front of the screen could allow the user to move around it.

2.2. it is preferable to have a adjusted stereoscopic vision

A well regulated stereoscopic vision should allow a better perception of the objects than monoscopic vision because stereoscopic vision provides additional information. However, to achieve this, we must be very careful about our actual settings, and they must be customized for each

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

subject. Indeed, if the device is not adjusted properly, the subject may have a vision that completely deteriorates.

2.3. It is preferable to track the subject's head

The fact that the subjects head is tracked might be sufficient for a impression of depth that may be considered quite acceptable, even in monoscopic vision. We will check whether perception with head tracking is more effective than without, either for monoscopic and stereoscopic vision. We will try to quantify the performance improvement.

3. Related work

We use a lot of clues to assess shape and depth of objects that surround us. We have many monoscopic and static clues: perspectives, light intensity differences, relative size of objects, occlusions, texture gradients. Kinetic clues are added: when we wove, or when the object moves, the movements of images on the retina is a very powerful way of perceiving relief. Some studies show that the perception of depth through the movement is independent of other depth cues [1, 2].

We also perceive depth by using our two eyes. They are not located in the same place (and separated by an interpupillar distance), so they see slightly different images. These differences in information are used by the brain to infer the depth of the observed objects. Contrary to the perception of relief through the monocular cues, it is a neurobiological phenomenon and not a cognitive one.

In immersive systems, two slightly different pictures are seen at the same time but separated by a horizontal distance (ideally the same between the two eyes) and the images are seen by the appropriate eye. This is obviously what is done in virtual environments where two virtual cameras take two virtual images that are presented to each eye (through artifacts such as auto-stereoscopic screens, stereo goggles...). Stereoscopic perception is however not identical in real and virtual environments for two major reasons: first in stereo-projected images, there is a discrepancy between convergence (on the interest point which might be behind or before the screen) and accommodation (still on the screen); second, most of real-time computer-generated images have no depth-of-field effect which proves to be disturbing [3].

To perceive correct depth and dimension of objects, binocular vision (neurophysiological phenomenon) must be consistent with the monocular vision, especially with object perspective (cognitive phenomenon). For this, the vision must be returned isomorphic to the vision of the real world, if it is technically possible.

Luo showed that background plays a strong role in the perception of the shape of an object, it also showed that the distance between the subject and the object was important [4]. Other studies have confirmed that the distance was important for shape perception [5, 6, 7], but none of them addresses the issue of the distance between the object and the projection screen (the distance between the object and the subject staying the same).

Wann have shown that interpupillar distance is important for a good depth perception [8]. But, he does not focus on shape perception.

We know thanks to Scarfe, that the human being overestimates the dimension in depth at near distances (under one meter) and underestimates at far distances (over one meter) [9]. So we will try to work around this threshold.

4. Methods and experiments

4.1. Subjects

The subjects were all healthy people. If they had vision problems (myopia, astigmatism...) they wore their eyeglasses during the test. They had to be able to move around the room without trouble.

We also verified that the subjects had good binocular vision. A simple test was carried out before the perception test itself: the subject was shown two adjacent blocks of different sizes, one being much closer than the other; the person then indicated which one was closer; we repeated this test three times to avoid random answers.

We also measured the interpupillar distance for each subject and integrated it into the virtual images generation.

We have 18 subjects (12 men and 6 women) for the test with the sphere, and 18 subjects (16 men an 2 women) for the unknown shape. 9 people have done both tests. The subjects are between 23 and 60 years old.

4.2. Global method

We will be presenting the subjects a real object and a virtual object which is similar but slightly different from the actual object. The modification between the real object and the virtual one is its relative dimension in X, Y or Z axis. The X axis is horizontal and parallel to the screen (the width of the object), the Y axis is vertical and parallel to the screen (the height of the object) and the Z axis is horizontal, perpendicular to the screen (the depth of the object).

4.3. Constant stimuli method

The stimuli are presented on a random basis. The subject must express stimuli difference between actual and virtual stimuli. This method has the best guarantees of validity and accuracy for measuring absolute JND or differentials, especially because it allows unbiased estimates of the psychometric function. However, stability estimates require a large number of tests and a stability of the subject condition [2, 10]. There are a number of rules which must be enforced to apply this technique [10]:

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

- A threshold level should be chosen. It may be determined by such a test using the method limits
- During pretests, the lowest and highest stimulus values where determined. They where chosen in order that almost each answer for the lowest level is ”smaller”, and almost each answer for the highest level which is ”bigger”. The determined interval is then divided into 5 to 7 parts to choose the stimuli levels.
- A rule of progression of these levels must be chosen. We can choose a rule of arithmetic or geometric progression. This choice is based on prior knowledge we have on the perception of the studied dimension.
- A constant number of trials (N) must be chosen and it must be identical for all n selected levels. A higher N will be the best estimate of the psychometric function.
- We chose the levels to show randomly without a second chance. So to see the first level again, we have to show all the other levels before.

We seek a level of differentiation (”Do you perceive the virtual object and the physical object to be the same?”), therefore we will use as a starting shape the virtual object that corresponds to the real object. We will determine a set of 7 variations in each dimension centered around the dimensions of the original virtual object. By original object we mean the virtual object that is theoretically equal in size to the presented real one.

We will look for the point of subjective equality (PSE) which is the virtual size perceived as equal to the real one (the dimension of our real shape) and the just noticeable difference (JND), the smallest perceived variation of the virtual stimuli.[11]

4.4. Objects

4.4.1 The sphere

The real sphere The real sphere is actually a smooth plastic ball with a 20cm diameter, the virtual model matches the color of the green paint applied on it. It is presented in front of a black cloth, and lit the same way as the virtual model.

The virtual sphere The original virtual sphere is the exact reproduction of the ball. It is green and also arranged on a black background. Dimensions vary by steps of 0.6 centimeters centered on the original diameter. When a dimension is different from the size of the original sphere, other diameters retain the original value, to change only one dimension at a time.

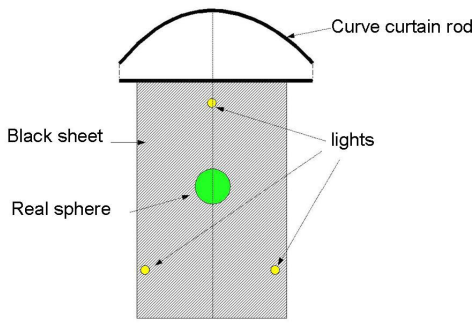

Figure 1. real sphere disposition

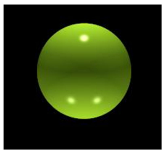

Figure 2. Virtual sphere

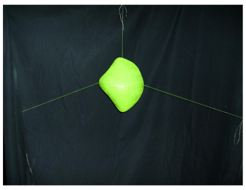

Figure 3. real unknown shape

4.5. The unknown shape

The real shape This unknown shape is a *deformed* sphere. It has the same average dimension as the sphere of the previous paragraph, but presents no symmetry.

It was imagined and designed in CATIA software (CAD software) and built by a rapid prototyping CharlyRobot numerical milling machine in a rigid foam polystyrene. This shape is painted in green with the same paint as the one that was used for the real sphere and is maintained in the same position as the virtual shape by three cables to avoid rotations.

It is also positioned before a black cloth (the same as in the sphere) and uses the same lighting (adjusted the same way). The whole system can be seen in Figure 3.

The virtual shape The virtual shape is derived from the CATIA model used for the real shape, it is green and placed before a black background.

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

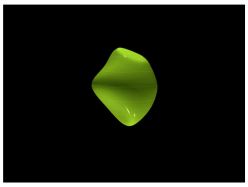

Figure 4. virtual unknown shape

test #	stereo	tracking	position
0	yes	yes	on
1	yes	no	on
2	yes	yes	behind
3	yes	no	behind
4	yes	yes	before
5	yes	no	before
6	mono	yes	on
7	mono	no	on
8	distorted	yes	on
9	distorted	yes	behind
10	distorted	yes	before

Table 1. Test sequences

4.6. Experiments

4.6.1 Sequences

Several test sequences are performed. They are presented in Table 1.

A "yes" in the "stereo" column means that the interpupillar distance is properly used (it was primarly measured on the subject). "Mono" means this distance is set to zero, the projection being therefore monoscopic and "distorted" means that stereoscopic vision is active, but the interpupillar distance (IPD) is set to the double of the actual subject's IPD.

In the "tracking" column a "yes" means that the head of the subject is tracked, and that this information is used by the program to adjust position of the virtual camera. Otherwise the person is not tracked. But, when the observer is fixed, we want the subject to always be at the same distance from virtual and real objects (1.2m). So, there are crosses on the ground to show which position he/she should be standing in. These crosses are showed to the subject before the beginning of the test.

The "position" column is the position of the object with respect to the screen plane. The offset of the object relatively to the screen is -50cm and +50cm.

4.6.2 Per sequence procedures

At the beginning of each sequence, instructions are given on the top left of the screen. The guidelines were :

"You can move around the room"

Or

"stay on the mark on the floor"

In the middle of the screen, a sentence invites the subject to press "A" on the Wiimote as soon as he has finished reading the instructions.

"Press A to start"

When a sequence starts, the width of the virtual sphere starts varying. The subject is asked the following question throughout the test:

"The sphere width varies. Is the virtual sphere less (-) or more (+) wide? "

The subject is then invited to press "-" or "+" of the wimote. When a button is pressed, a confirmation appears on the screen:

"You have pressed on" - / + "press" A "to confirm"

Once he has validated his reply, the sphere disappears half a second to avoid comparisons between virtual spheres. If the subject presses "A" without having given an answer, the program does not validate and shows once again the same sphere. There are 2 sets of 7 random variations without the same size being shown twice. Once these 14 questions have been answered, the second parameter variation is started (height) and finally the third one (Z dimension). For tests with the unknown shape, the process is the same except for the name of the shape.

4.7. Experimental conditions

The immersive display consists in a projection screen of 3.1m by 1.8m. The projector is a Christie Mirage HD3 of a 1920 by 1080 resolution. Our pixels are thus 1.6mm wide. The projection frequency is 60 frames per second. The tracking system is an ART2 operating at 60Hz. We notice that the differences that can be detected will be necessarily higher than 1.6mm, as pixels are 1.6mm wide.

5. Data Analysis

5.1. Collection and pre-data

At the end of each sequence, the program saves the subject's answers in a text file for the entire sequence. The files are then sorted and concatenated. Each large file combines all subjects' answers to a sequence.

From these files we compute the probability of "virtual sphere is bigger than the real one" for each variation. The probability is represented on a graph, and follows a cumulative distribution function. Figure 5 shows the graph for the

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

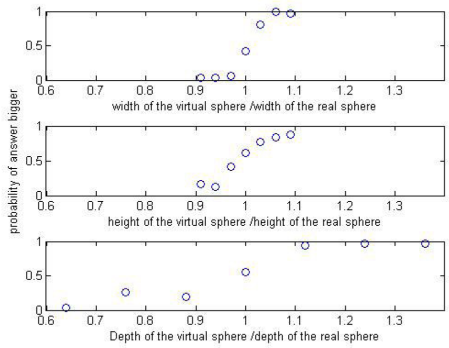

Figure 5. probability of response "bigger" of the subject

first sequence (active stereovision with adjusted interpupillar distance, tracked head) achieved by the sphere.

From these data, thanks to the least-square algorithm in the normal distribution domain, we compute the best fitting line.

+++++++++++ A CHANGER ++++++++++++ From this best fitting line we deduce the PSE (the point where the straight line cuts the abscissa) and the JND. And we can trace the curve that approximates the points in the probability space to visually ensure that it approximates our baselines (see in figure 6). Note that in this experiment we used a standard definition of the JND, as the stimulus value for which the probability of "bigger" response is 0.8143.

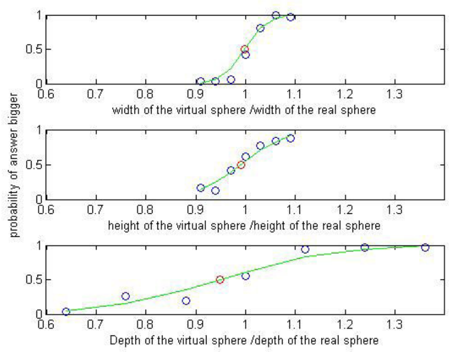

Figure 6. approximation of probability of response "bigger" of the subject

5.2. Treatment and data analysis for the sphere

For all the following graphs, we represent the PSE and the JND on both sides (box diagram). If the JND is large, it means that the subject dithers.

5.2.1 Position of the virtual sphere

The PSE (point of subjective equality) and JND are an averages of PSE and JND points of several sequences:

- sequences 2 and 3 are used for the analysis with the sphere behind the screen
- sequences 0 and 1 are used for the analysis with the sphere on the screen
- sequences 4 and 5 are used for the analysis with the sphere in front of the screen

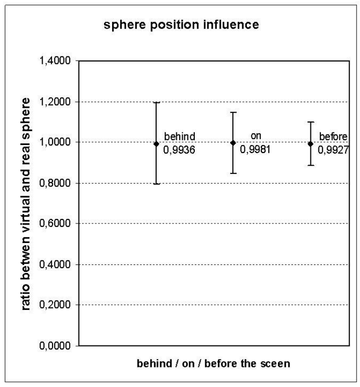

Figure 7. sphere position influence influence

As we can see in figure 7, the JND is much larger when the virtual sphere lies behind the screen. This is quite logical since the sphere appears further, it has a smaller retinal image, thus less accurate. Moreover, as it is displayed smaller, a sub-sampling effect might appear, which also can adversely affect the accuracy of sight.

Most surprising is that the JND is greater when the object is located on the screen, which can lead to the hypothesis that eyestrain due to high disparities (when the object is in front of the screen) does not disturb perception. Moreover, almost all subjects said they preferred to look at the objet when it was located before the screen (between the physical screen and the user) and be able to turn around. We assume that the ability to see the sides of the object helps for a more global comprehension of the object.

5.2.2 Influence of head tracking on perception

The PSE and JND are averages of PSE and JND of several sequences:

- sequences 0, 2, 4, 6, 8, 9 and 10 are used for the analysis with head tracking
- sequences 1, 3, 5 and 7 are used for the analysis without head tracking

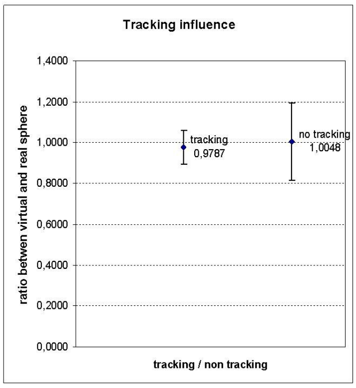

Figure 8. head tracking influence

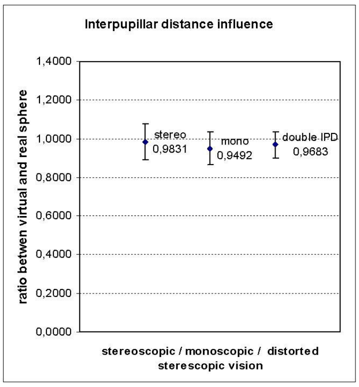

Figure 9. stereoscopic vision influence

We see in figure 8 that the JND is very different for the case with or without tracking. It appears that the adjustment of the virtual camera in relation to the position of the subject is essential for a good perception.

This difference is especially marked for depth perception, indeed, the detection threshold of change in Z axis is 16.5% of the diameter with tracked head against 47.97% with fixed position. Subjects had a very strong tendency to move in the room, and around the object, to look at its depth from different angles.

5.2.3 Influence of stereovision

In order to analyze the influence of stereovision, 3 conditions are studied : correct IPD, null IPD and double IPD, this results in 7 sequences as summed up in table 1.

- sequences 0, 2 and 4 for are used for the analysis with adjusted stereovision (correct IPD)
- sequence 6 is used for the analysis with monoscopic vision (null IPD)
- sequences 8, 9 and 10 are used for the analysis with stereo with a double IPD with respect to the one of the subject.

Note in figure 9 that the JND does not seem very different depending on the used IPD (adjusted or doubled or null). This would tend to show that having a tracked head is more interesting than stereoscopy for shape perception. But it is true that when issues related to width or height, some subjects (however a small part) measured with their fingers while closing one eye, which of course distorts the difference between monoscopic and stereoscopic vision. But we do not believe that the number of these cases is significant compared to the number of people.

5.2.4 Analysis by axes

The PSE and JND are averages of PSE and JND calculated for the variation in width (X), height (Y) and depth (Z) on all sequences.

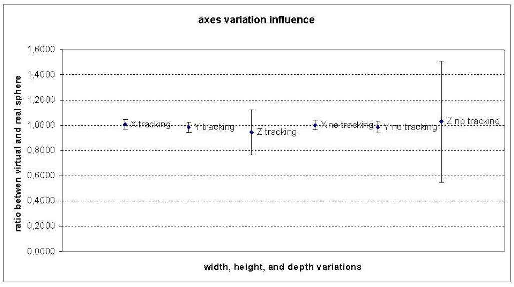

Figure 10. variation data influence

As we see in the figure 10, the thresholds in width and height are quite similar. Depth changes are less easily perceived.

But this seems logical, the depth is the only variable that is not in front of the subject when he is at the center of the room. It is therefore considerably influenced by the presence or absence of tracking. But even when the head is tracked, the difference between the JND in width (or height) and the depth is remarkable: a ratio between the depth and width JND is about 5. When the head is not tracked, this ratio is 12.

This difference was already seen during the pre-tests, to adapt our tests to this phenomenon, we had to set the changes presented in the depths as 4 times larger than the width and height.

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

5.3. Treatment and data analysis for the unknown shape and comparison

The sequences used to make the following charts are exactly the same as those used for charts on the sphere perception tests.

5.3.1 Position of the virtual shape

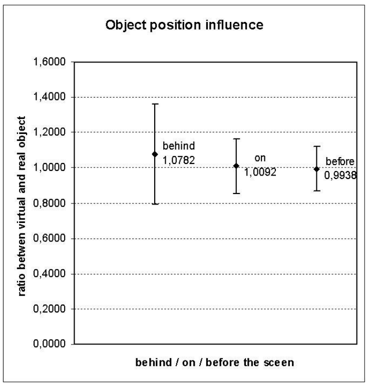

Figure 11. Shape position influence

We note that the trend observed during the distortion of the sphere is more pronounced. The subject perceives fewer deformations on the farthest object from him. We see that JND are generally larger than that for the sphere. This increase is linked to the increase in JND values without head tracking (see 5.3.2).

The threshold of detection for the unknown shape behind the screen was 20%, it is now 28%, so there is a ratio of 1.4 between these two JNDs. As for the shape in front of the screen, the threshold for the sphere was 10.7% it is now 12.4% of the report and is therefore 1.15. Hence we see therefore that even if there is a general increase in the threshold, the difference between a shape in front of and behind the screen reinforces the lack of symmetry.

5.3.2 Influence of head tracking on perception

Again, the difference is more marked when the shape is less symmetrical. The threshold ratio between tracking and no tracking for the sphere was 2.28, it is 3.06 for the unknown shape.

It should be also noted that there is a very strong difference in the behaviour of subjects during these tests, not apparent on those charts. During the sequences without head tracking with the ball, most subjects used the lighting cues (reflections of lights on the sphere) to detect changes in depth. For the non symmetrical shape, the position of reflection is not symmetrical either, and it appears that the subjects do not need them any more. Some have told us to use the deformation of the shadows on the left side of the

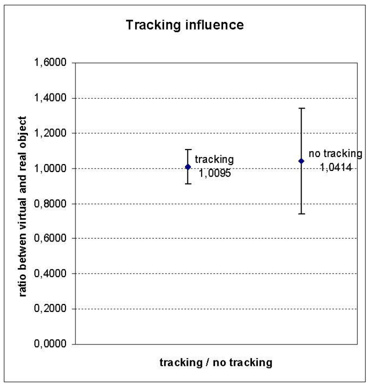

Figure 12. Head tracking influence

shape. Some have noticed moving reflections, but were unable to make the association between distortion and visual cues displacement, others have noticed and made the relationship but they were not so many. Many of them have said they had seen a ”flat” shape all the time when the head is not tracked (either with or without stereoscopic vision).

5.3.3 Analysis of the influence of stereovision

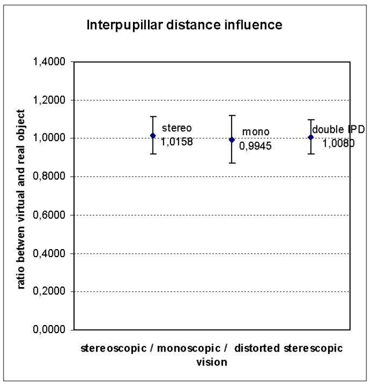

Figure 13. Stereoscopic vision influence

We can note on graph 13 that unlike the sphere, the distorted object is not perceived in a better way in the two stereoscopic modes compared to monoscopic mode.

But it is possible that the sequence order has a significance in the outcome of the distorted stereoscopy. Indeed, tests on distorted stereoscopy occur at the end, the subject has thus had ample time to get used to the shape he has in mind and will detect more easily. It is possible that the results would not have been as much in its favour if these sequences were taking place at the beginning of the test.

Compared to the previous observations on head tracking influence, we can nevertheless note that the influence of the head tracking is much more important than the influence of stereoscopy.

ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

5.3.4 Analysis by axes

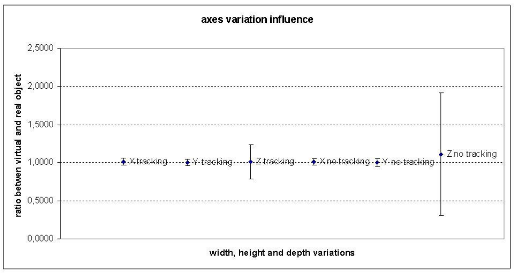

Figure 14. variation data influence

We see again that several JND are much more pronounced in the shape without symmetry than for the sphere (the threshold depth is 42.5%).

6. Conclusion and discussion

We have conducted a series of tests that have enabled us to compare the influence of parameters on the perception of shapes. Among these parameters, we emphasized that head tracking has a more significant influence than other parameters. It is quite possible that the sequences order play a role in the outcome of the influence of interpupillar distance on perception, but we see that this parameter is still not the dominant one against head tracking. Changes in shape depth difficulty remain detectable, whatever the conditions of projection.

So, the designer who wants to see the virtual prototype in the same way as the real one, has to ensure first he will benefit from a head tracking system. He can also put the virtual object just before the screen to let the subject turn around the object. And he has to adjust the distance between the virtual eye on the distance between subject eyes.

In the immediate future, we intend to study the perception of local modifications. This could be done by keeping the real non symmetric shape, and proposing a locally deformed virtual object, that the user will have to remodify to look like the real shape again. The mistakes on different points will be measured and analyzed. Performance issues in different rendering modes will also be assessed on the visual fatigue point of view. We will try to show whether stereoscopic comfort enhancement strategies are of use or not.

References

[1] Steven E. Palmer, *Vision Science: photons to phenomenology*, Bradford book, The MIT presss, 1999. 2

[2] E. B. Goldstein, *Sensation and Perception*, Wadsworth Publishing Company, 2001. 2

[3] David R. W. Butts and D. F. McAllister, "Implementation of true 3-d cursors in computer graphics", in *Three-Dimensional Imaging and Remote Sensing Imaging*. SPIE, 1988, vol. 902, pp. 74–83. 2

[4] X. Luo, R. Kenyon, D. Kamper, D. Sandin, and T. DeFanti, "The effect of scene complexity, stereovision, and motion parallax on size constancy in a virtual environment", in *IEEE Virtual Reality*, William Sherman, Ming Lin, and Anthony Steed, Eds., Charlotte, NC, march 2007, IEEE. 2

[5] Robert G. Eggleston, William P. Janson, and Kenneth A. Aldrich, "Virtual reality system effects on size-distance judgments in a virtual environment", in *Proceedings of the IEEE Virtual Reality Annual International Symposium*, 1996. 2

[6] Alfred H. Holway and Edwin G. Boring, "Determinants of apparent visual size with distance variant", *American Journal of Psychology*, vol. 54, no. 1, pp. 21–37, 1941. 2

[7] Alexis Paljic, Sabine Coquillart, and Jean-Marie Burkhardt, "A study of distance of manipulation on the responsive workbench", in *Immersive Projection Technology Symposium*, 2002. 2

[8] John P. Wann, Simon Rushton, and Mark Mon Williams, "Natural problems for stereoscopic depth perception in virtual environments", *Vision research*, vol. 35, no. 6, pp. 2731–2736, 1995. 2

[9] P. Scarfe and P.B. Hibbard, "Disparity-defined objects moving in depth do not elicit three-dimensional shape constancy", *Vision research*, vol. 46, no. 10, pp. 1599–1610, 2006. 2

[10] C. Bonnet, *Manuel pratique de psychophysique*, Armand Collin, 1986. 2

[11] Bernard D. Adelstein, Thomas G. Lee, and Stephen R. Ellis, "Head tracking latency in virtual environments: psychophysics and a model", in *Proceedings of the IEEE Virtual Reality Annual International Symposium*, 2003. 3