IDI – Usability use case – Depth perception in VR

Pere-Pau Vázquez – Dep. Computer Science, UPC

Contents

1 Introduction

In these notes we will present a usability evaluation use case, more concretely, an experiment tailored to *measure* different performance aspects of depth perception in a virtual reality environment. As a consequence, the experiment was implemented as a **measurement-based usability testing**.

1.1 Context

Volume Rendering is a popular rendering technique that can be applied to many fields and types of data, such as medical visualization or flow rendering. The main feature of volume rendering comes from the fact that the data to visualize consists on a whole volume instead of having the geometry stored as a set of planar faces. The reason behind that is that we want to inspect what is *inside* the models, being able to see the whole volume, as in Figure 1, where a body that has been captured using a Computational Tomography (CT) is inspected using direct volume rendering.

Figure 1: Volume rendering of an anatomic model using the emission-absorption model with Phong Shading.

Volumetric rendering techniques are very important because they facilitate a type of exploration that cannot be done using other techniques. In order to produce such effects, it is necessary to strongly use semi-transparency and this sometimes limits the perception of depth: the ability of a person to classify the different distances at which the objects are placed in the real 3D data.

1.2 Goals of the study

The purpose of the study is to analyse whether the shading technique can influence the depth perception of volumetric models. More concretely, we want to evaluate if any of the advanced rendering techniques that have been developed recently, produces better images in terms of depth perception in Virtual Reality environments [Grosset13].

The perception of depth in our brain depends basically on the fact that, since our eyes are separated by a certain distance, when we have both of them open, we see two slightly different images. In stereoscopic environments, we simulate this effect by generating two different images, one for each eye, and project them using some technique such as immersive systems (e.g. head mounted displays), or semi-immersive systems such as CAVEs, PowerWall, 3D TVs, etc.

2 Experiment setup

As we have already seen, the goal of a user study is to determine the response to certain stimuli. Therefore, we must set up the conditions such that we can make sure we are measuring/calculated what we actually intend to. Thus, we have to dive a bit into the human perception system to ensure that we know how to determine the influence of the shading technique in the ability of perceiving depth from other, confusion variables.

2.1 Shading techniques to compare

Many new advanced shading techniques have been developed rendering for volumetric models [Hernell09, Kniss03, Langer00, Schott09]. Testing all would make the experiment too long, and very difficult to evaluate: we cannot make the users test 20 different shading conditions for many models, since this would create obvious fatigue. Perception experiments should be carried out for a relatively low limited time, because the attention and vigilance capabilities of the human brain are quite limited.

As a result, we decided to test 4 conditions: two basic techniques, no shading and Phong shading, and two advanced shading methods, half-angle slicing and directional occlusion shading. Figure 2 shows an example of synthetic model with the four shading conditions. Note that some of the basic shading techniques seem to produce little or no depth cues, and therefore, one might think that the depth perception will likely be hampered. However, we do not know if the use of two images, one for each eye, will compensate. On the other hand, the advanced techniques presented in the two bottom images yield visual cues that encode the relative positions of the objects. However, we do not know if such shadows will make the users to perceive darker regions as further than they really are. All of this can only be determined by making a formal user study where the answers of the participants are gathered in such a way that there is no ambiguity on what they are answering. This does not mean that the users will all of them perceive all the scenes equally. We know that perception is variable among human beings. However, previous studies have demonstrated that low-level perception tasks, such as color identification or depth perception is very similar among persons with correct or corrected vision.

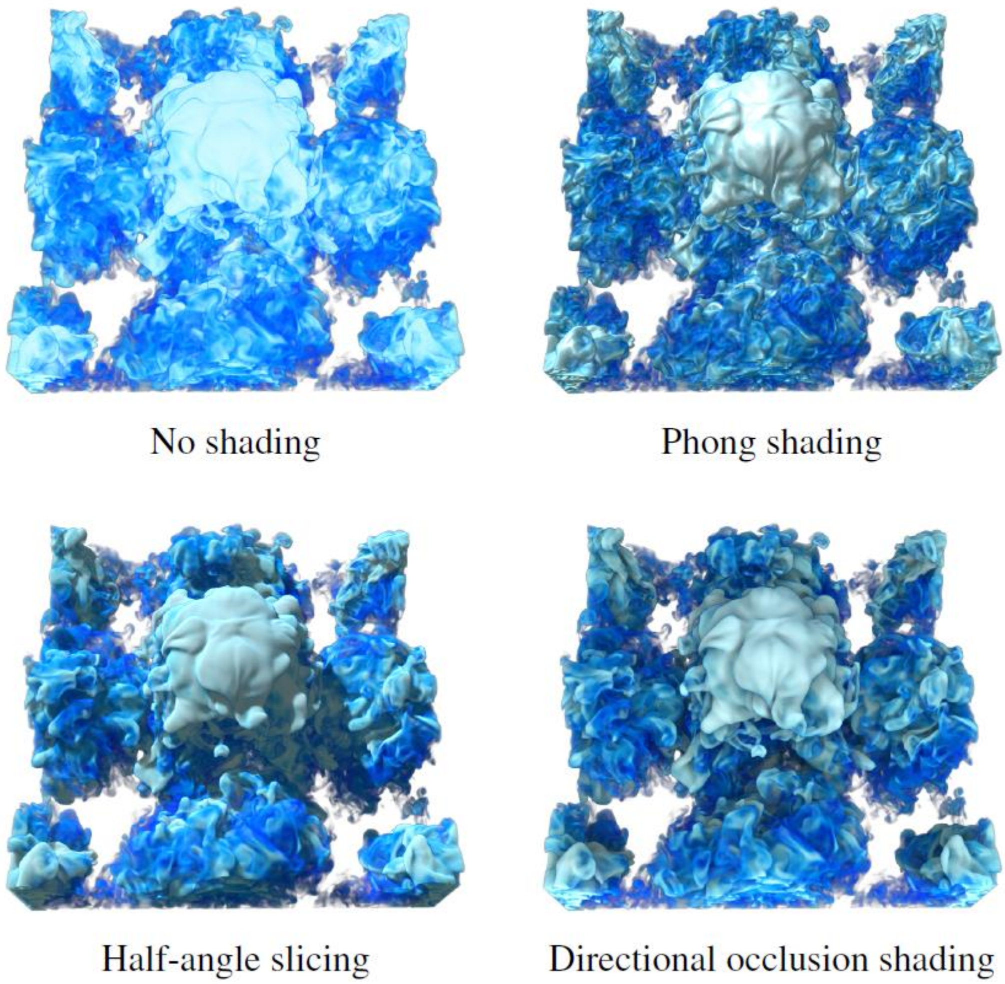

Figure 2: Volume rendering of a synthetic model using the four shading techniques: *no shading*, *Phong shading*, *directional occlusion shading,* and *half-angle slicing*. Some of them, especially the ones on the top, produce more *planar* images where depth seems more difficult to perceive. However, the shadows on the bottom images may sometimes mislead perception.

2.2 Methods and apparatus

There are many aspects that influence 3D perception. For example, the movement of our head, that causes the scene to be seen under different angles, causes parallax effects (small variations of what we see that depend on the distance and position of the point of view, that is, the elements in the background move slower than elements in the foreground). This is a very important factor in the perception of 3D, besides the two different images that we see with each eye.

As a result, if we set up an experiment in which the users can move freely around the scene, it will be very difficult if not impossible, to separate the parallax effect in the 3D perception from the variables we want to test: illumination algorithms. This has to be taken into account when performing a measuring experiment.

As a result, we will consider a stereoscopic environment where the users cannot move freely, so we will use a 3D TV with a fixed position for the participants. They will be seated at a fixed distance and no head tracking will be used to avoid parallax effects influence the results.

The tasks we will ask the users will consist on classifying two points, selecting the one that appears to be closer to the observer. These points will be randomly selected in the scene, but all the points will be the same for different users, so that we have data to compare with. Moreover, we will ensure that the distribution of the real depths of points will be also relatively uniform along the depth values of the scene: some point pairs will have similar distances, some of them will have medium distances, and some of them will have great depth separation. Also, for the points, some of the closer points will be at the left part of the image or the right. Moreover, the horizontal separation of the points will also vary quite uniformly, to test whether the users distinguish better points that are closer from points that are far away from each other. Generating such a set of points is quite time consuming. Testing absolutely random points per each user makes the statistical analysis more difficult, or even impossible unless we have thousands of samples, which is very complicated, due to the fact that the test must be carried out in a controlled environment.

Concerning the datasets, we need to provide examples where the user cannot get an idea of the relative position of the points just because the 3D models are known to the user. For example, we cannot use the model in Figure 3 and ask the users to classify points denoted by the blue circle and square because any participant will know that a point in our lip will be closer than a point in the neck.

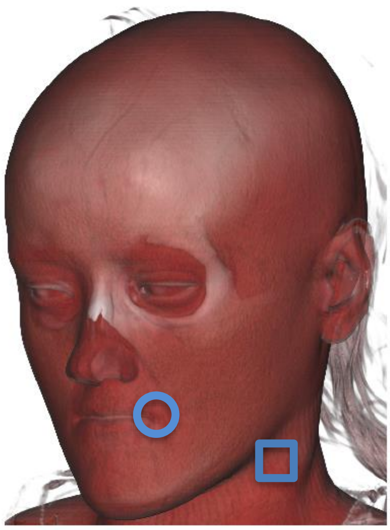

Figure 3: Volume model of a geometry the user will know and that would yield misleading results: the user can answer properly independently of the shading function used.

As a result, we will use either models that are not known by the vast majority of the potential participants, and in case of doubt, such as the *backpack* model, that is used in volume rendering research, we will generate view positions that make classification of points according to aspects that are not just the shading, more difficult (see Figure 4).

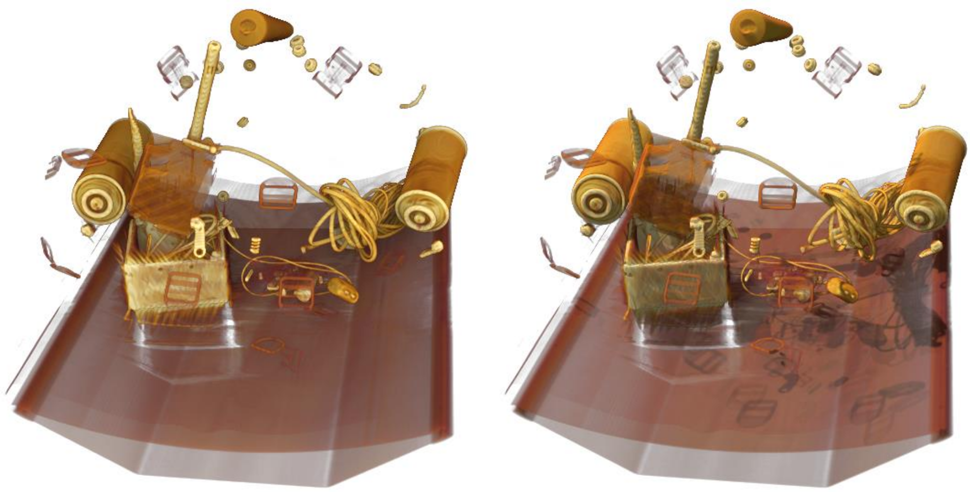

Figure 4: Volume rendering of the *backpack* model using two different shading techniques: *directional occlusion shading* (left) and *half-angle slicing* (right), at viewpoints that make very difficult to relatively place the objects with respect to the viewer.

2.3 Test preparation

Once we have decided the number of shading methods (4), and that we are going to use models that are unknown, we need to select the number of participants. This will depend on the nature of the problem we deal with. Since low-level perception is not very variable among humans, we need a sample group that is larger than 10 to have significant results.

Since we have 4 shading conditions, and we need enough data for the statistical analysis, we used 6 different modes in the study. These were a mixture of CT models and synthetic data sets. In total, users were exposed to six different models, half of them were CTs and the other half were synthetic. The generated stereo images are representative of volume visualization, since they present cluttered regions with complex shapes and combine semi-transparent with opaque layers. Note that, as commented previously, we used specific data sets and added arbitrary rotations to them to reduce the possibility that a previous knowledge of the data or too evident shapes facilitate the recognition of depths aside from the proper perception. We also avoided models that represent well-known parts of the human body, for example.

The participants were exposed to the different models where the images were systematically sorted based on the shading technique using Latin squares: We have 4 shading conditions and 6 models, so we can use a 4x4 Latin squares model. The 18 images of each technique (variations of models and views) can then be presented randomly, to avoid learning and fatigue effects. We need therefore a number of participants that can be divided by 4, to repeat all the Latin squares conditions the same number of times. That is why we used 16 participants for each task. In our case, we changed the participants for each task, to avoid learning effects. But due to the nature of the tasks, this would not have been necessary.

To be able to perform a proper data analysis, we will capture the correctness to the answers, as well as the time devoted to answer. This was necessary because it will be interesting to analyze whether a higher amount of correct answers can be due to a larger amount of time devoted to solve the questions. Note that in this case this implies to modify the software for the experiment, since accurate data of time answering can only be obtained by measuring the time between clicks.

2.4 Physical setup

For the experiment, we used a passive stereo system consisting on a 46” JVC 3D TV (GD-463D10 model) with polarized glasses. Moreover, we added a lamp placed top-left and some objects around the TV (two balls, a plastic glass and a couple of books) so that the lamp generates shadows of different sizes inside the participants’ Field of Regard (at least the peripheral vision was aware of such shadows, see Figure 5). The lamp position was chosen to be coherent with the virtual light source that is used in half-angle slicing and Phong shading (the position of the virtual light source is changed in task 3 to make it non-coherent with the lamp position). Users sat during the study (2 m from the screen) to avoid movements and thus, limiting the impact of the perspective distortion.

Figure 5: Physical setup.

The application shows static stereoscopic images to avoid depth inferring via other elements such as the model motion. We created a small application that shows two markers in each image, and lets the users select the one which is placed closer to the observer. The markers are designed as small windows with two shapes: circular and square. They indicate the points of the model to be classified by the users. The markers are placed at the same distance from the observer, and users were instructed to classify the point that the markers show, not the markers themselves. These markers pop up three times, and then disappear. If necessary, the users may request the application to show the markers again. In Figure 6 we show one of the used images.

The selection falls into the category of two-alternative forced choice, which means that the users have only two options and cannot answer "do not know". This eases the statistical analysis, but may have some limitations if our goal is also to measure how sure or unsure a participant is of the given answer.

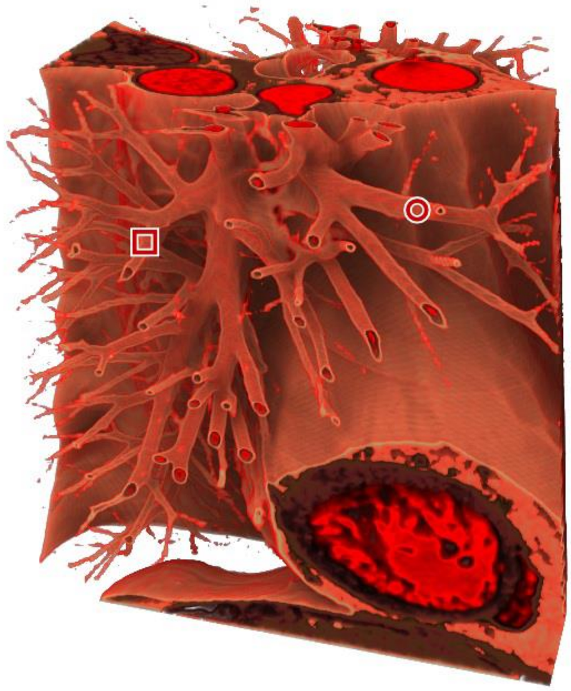

Figure 6: Users were shown an image with two markers that identify two points in the model that must be classified. The markers disappear after popping up for three times and the users must determine which of the points is closer.

In order to facilitate the tasks, we customized a keyboard by painting all the keys in black and putting some stickers to mark the necessary keys, as shown in Figure 7.

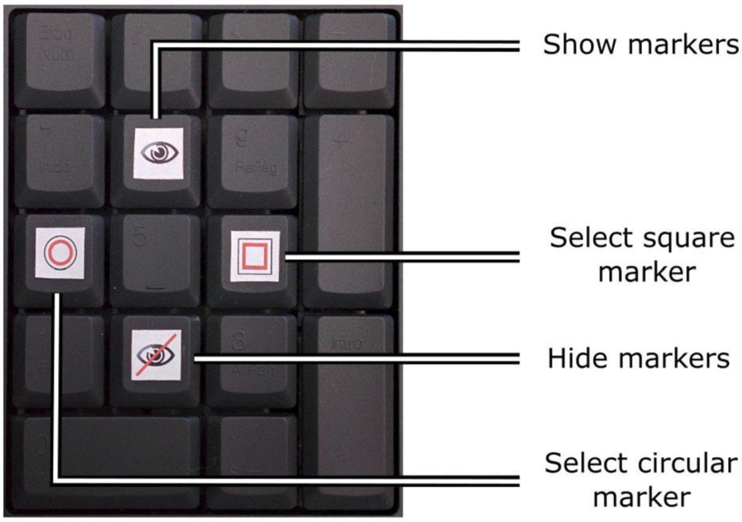

Figure 7: The modified keyboard to facilitate the selection. By painting the keys in black, we eliminate distractions, and the white labels facilitate the identification of the proper keys during the experiment, since the light is highly dimmed.

2.5 Design and procedure

The study comprises three different tasks. The first one evaluates the perception of depth using different shading techniques. Then, the second task evaluates the same methods in the presence of a controlled external light source (we use a lamp that casts obvious shadows inside the field of regard of the user). This task is tailored to determine if real illumination has any influence on the responses. Apart from the presence of the lamp, the experiment and the setup in both tasks are the same. Finally, the third task evaluates the influence on depth perception of having coherent real and virtual light directions. This task was meant to be done for those shading models that allow to change the virtual light position, only if the analysis of the previous tasks showed a real influence of external illumination on depth perception.

3. Implementation

3.1 Test performance

Participants were shown 72 images covering three different points of view, four different techniques, and six different models. Since the variable to analyze was the shading model, the images were systematically sorted based on the shading technique using Latin squares, and the 18 images of each technique (variations of models and views) were then presented randomly. The users were introduced the task with a first example that had to be solved before the experiment effectively started. Everybody showed a proper understanding of the procedure and all the users completed their tasks to their satisfaction (as proved by post hoc questionnaires). After each choice, the users had a neutral screen to let them recover from fatigue if necessary. Participants were instructed to determine the closer point from a pair indicated by two markers and asked to take as much time as necessary. Markers popped up three times before the actual selection might start. In case of necessity, the users could make the markers visible again. The test was previously checked with two extra users whose results were not included in the data analysis. Throughout the experiment, we measured the correctness of the answers and the time spent in each choice.

They had to fill in two different questionnaires: one before the task with personal information (age, gender, quality of eyesight, experience with VREs, etc.) and another after the task, asking about a subjective evaluation of the performed activity. After analyzing the post hoc questionnaires, we found that all of them understood their task properly and found it easy to achieve. Therefore, there was no need of discarding any user.

3.2 Data analysis

The mean correctness of the answers for each task was analyzed by using a one-way analysis of variance (ANOVA) with a significance level of $\alpha = 0.05$. When significant differences between the means were found, we used a post hoc Bonferroni's pairwise test with the same significance level ($\alpha = 0.05$). The same analysis was performed with the means of the time spent by the users to complete each task.

To test the linear correlation between the measured variables, we used the Pearson's r statistic and assessed the linear model testing the regression coefficient β_1 with $\alpha = 0.05$. Finally, the Chi-square test of association with a significance level of $\alpha = 0.05$ was used to analyze categorical variables.

The most relevant elements we analyzed with the gathered data were:

- Effect of marker position: is there any difference in depth perception if the points are close or far from each other?
- Effect of left and right eye: is there any difference if the closer point appears in the left eye or the right one?
- Effect of the luminance: has the luminance of the pixels around a point to classify any effect on the correctness of the answer?
- Had time any effect on the correctness?

3.3 Results

In this experiment we found that the shading effect had effective influence on the correctness of the points classification. Directional Occlusion Shading performed significantly better than no shading, and marginally better than the other techniques, though no statistically significantly better. Half-angle slicing was also better than Phong Shading and no shading, but with no statistical significant level. The results can be seen in Figure 8.

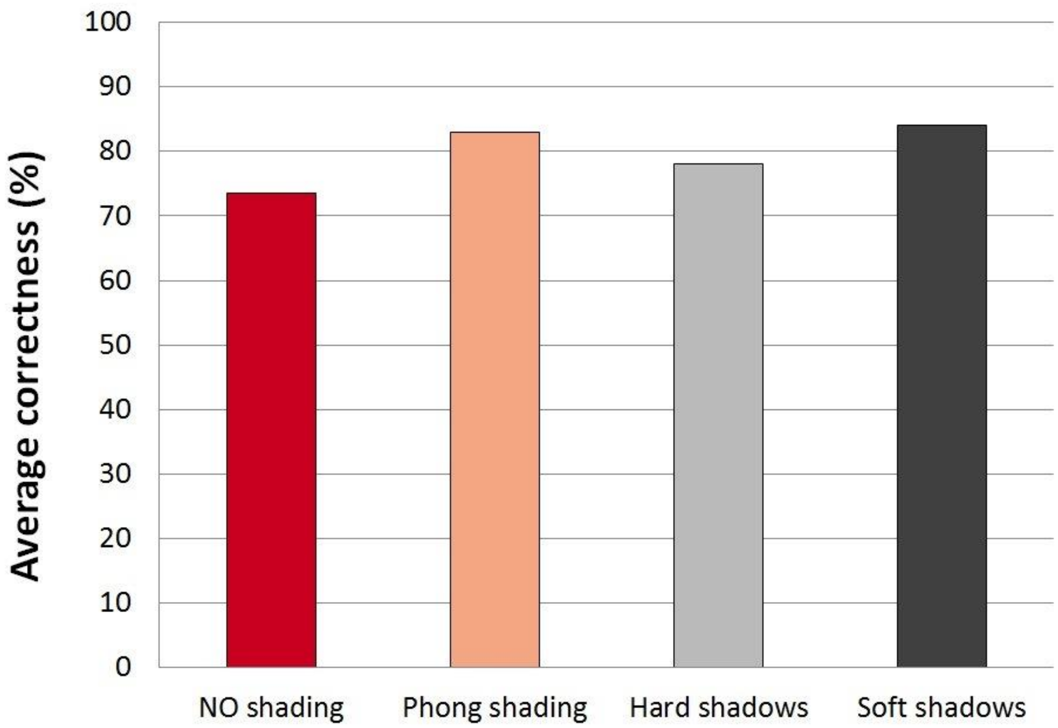

Figure 8: Average correctness for each shading technique.

The average correctness can also be analyzed per tasks. For task 1 and task 2, the results are summarized in Figure 9.

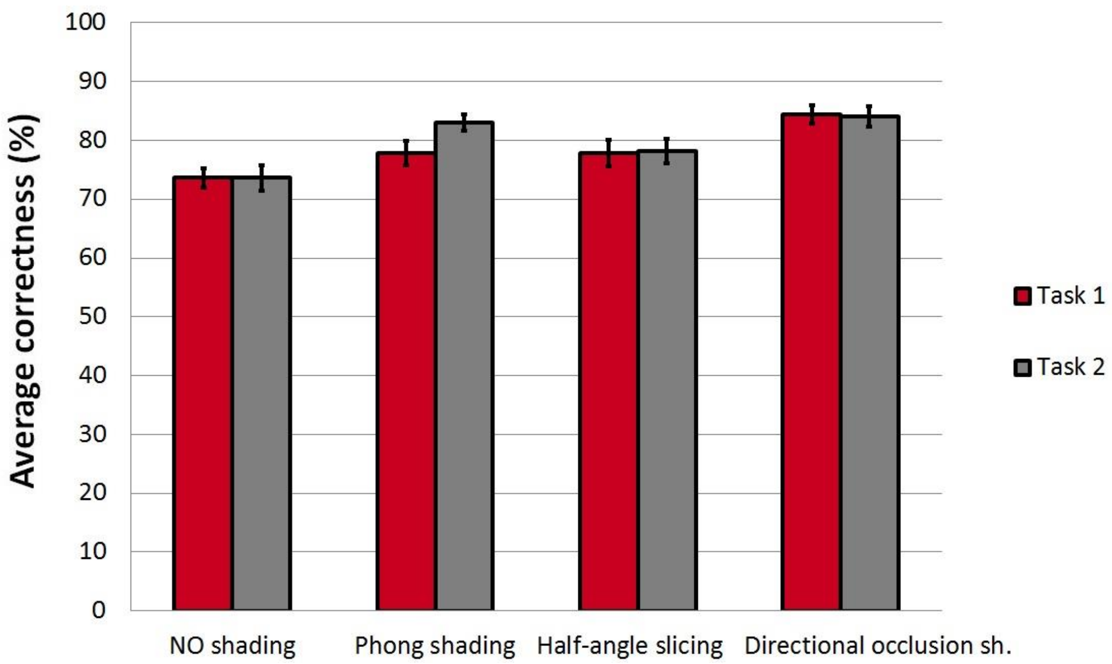

Figure 8: Average correctness for each shading technique, separated per tasks.

No significant differences seem to be related with the time elapsed to answer the questions and the different shading techniques. There seems to be a weak linear relationship between correctness and time spent. However, no significant enough.

We can conclude, after the experiment, that advanced volumetric shading techniques provide a better advantage at perceiving depth in virtual reality environments. There was also no detected correlation between the X/Y position of the markers and the correctness in the answers. Furthermore, real illumination around the scenario was not influencing in the answers.

Further details on the experiment and the results can be found in [Díaz17].

References

- [Díaz17] Díaz, J., Ropinski, T., Navazo, I., Gobetti, E., Vázquez, P. *An experimental study on the effects of shading in 3D perception of volumetric models*, The Visual Computer 33(1), 47-61 (2017)
- [Grosset13] Grosset, A., Schott, M., Bonneau, G.P., Hansen, C.: Evaluation of depth of field for depth perception in dvr. In: Proceedings of the 2013 IEEE Pacific Visualization Symposium (PacificVis), pp. 81–88 (2013)
- [Hernell09] Hernell, F., Ljung, P., Ynnerman, A.: Local ambient occlusion in direct volume rendering. IEEE Trans.Vis. Comput.Gr. 15(2), 548–559 (2009)
- [Kniss03] Kniss, J., Premoze, S., Hansen, C., Shirley, P., McPherson, A.: A model for volume lighting and modeling. IEEE Transactions on Visualization and Computer Graphics 9(2), 150–162 (2003)
- [Langer00] Langer, M.S., Bülthoff, H.H.: Depth discrimination from shading under diffuse lighting. *Perception* 29(6), 649–660 (2000)
- [Schott09] Schott, M., Pegoraro, V., Hansen, C., Boulanger, K., Bouatouch, K.: A directional occlusion shading model for interactive direct volume rendering. In: Proceedings of EuroVis'09, pp. 855–862 (2009)