Perception-based Evaluation of Emphasis Techniques Used in 3D Medical Visualization

Alexandra Baer, Friederike Adler, Daniel Lenz, Bernhard Preim

Department of Simulation and Graphics, Otto-von-Guericke University Magdeburg
Contact email: Alexandra.Baer@ovgu.de

Abstract

This paper presents an experimental user study of emphasis techniques used in medical visualizations for focus and context illustration. The study serves as a framework for the evaluation of various kinds of techniques in complex medical scenarios. Guidelines and findings from psychophysical experiments are adapted to a common therapeutic question of a treatment planning process. We searched for a trade-off between realistic visualizations used in clinical routine and constraints of psychological approaches. Therefore, we designed a conjunctive search experiment using the example of the enlarged lymph node detection in 3D models derived from CT neck datasets. According to that, we statistically analyzed cutaway, stippling and red coloring as three representative emphasis techniques and validated the subjects' psychological response bias with the signal detection theory.

1 Introduction

In the past decade many novel techniques for visualizing medical volume data and derived segmentation information have been developed and refined. However, it is difficult to decide which techniques should be used for particular applications, how they should be combined and how parameters should be adjusted. Most descriptions of new visualization techniques do not ground their development with arguments from visual perception. Without such guidance, the potential of the techniques remains underutilized in medical education and treatment planning systems.

In order to answer such questions, in principle, two strategies are possible:

- The effects of visualization techniques and their application are investigated with respect to specific examples with medical doctors as users.
- Visualization techniques are systematically compared by means of controlled physical experiments which reveal the perceptual effectiveness of selected techniques.

Both strategies have their advantages and drawbacks. The findings of the first kind of experiment are strictly bound to particular applications; it is difficult to generalize from them. Moreover, it is hard to avoid that feedback from doctors is more than just subjective opinions. On the positive side, the first kind of experiments reveals conventions in specific application areas which should be considered in designing computer-based visualization systems. Controlled perceptual experiments can be performed such that the results are more general and more objective. However, it is difficult to adapt common psychophysical experiments that are usually based on rather simple geometric shapes and tasks to complex 3D medical scenarios. The major challenge is the integration and adaptation of empirical criteria like objectivity, validity and reliability to medical visualizations and still producing reliable and traceable results.

We present an experimental user study in terms of a conjunctive search for focus objects in 3D medical neck visualizations. Guidelines and findings from psychophysical experiments are adapted to a common therapeutic question of a treatment planning process. Hence, the following paper includes the study design, the implementation, the analysis and the achieved results.

2 Related Work

In computer graphics, perceptually driven research has perhaps had the longest tradition in virtual reality, where the users' response to specific interaction and rendering techniques is evaluated using a variety of methods [1]. Findings from psychophysical studies are more and more used to improve virtual and augmented environments and to enhance the presentation and effectiveness of other computer graphics visualizations. Bowman et al. [2] introduced guidelines for evaluation studies in virtual reality. Similar to Swan et al. [3], they concentrate on the evaluation of interaction techniques and virtual reality platforms according to the users' performance. Contrary to Swan et al. [3], who defined user performance as the individual time needed by the users to answer the questions, Bowman et al. [2] analysed the accuracy and completeness of a user task as well.

A user study discussing focus and context perception in an immersive virtual reality was presented by Jackson et al. [4] who tested the effects of visual context on feature search and recognition abilities. However, their context consists of a simple natural background like a brick or a blank wall without distracting objects and had nothing in common with the focus object. Apparently it is easier to detect the focus object.

Well-known user studies in computer graphics are dealing with the efficiency of a visualization by means of the desired information transfer. Laidlaw et al. [5] quantitatively evaluated several 2D vector field visualizations. They presented six visualization techniques, such as arrow icons or wedges for each vector field and proposed three simple but representative tasks for understanding the illustrated flow. For each task, they measured the accuracy and execution time of each participant.

Interrante et al. [6], Kim et al. [7] as well as Bair and House [8] discussed the evaluation of texture features used to improve 3D shape representation. The accuracy of normal vectors drawn by the users turned out to be a good criterion to judge the different techniques.

In contrast to these experiments, Ritter et al. [9] presented a perception study using 3D anatomic structures from clinical datasets visualized with four illustrative techniques. The experiment analyzed the depth perception of vascular branches in a blood vessel tree illustration. They concluded that distance-based methods can be used efficiently, while no significant improvement was shown for their specific hatching techniques. A questionnaire-based study that examined medical scenarios was presented by Tietjen et al. [10]. They investigated hybrid visualization using direct volume rendering, surface shading, and line rendering for liver surgery planning scenarios. The results primarily indicate personal preference of the participants. Contrary to Tietjen et al. [10], Salah et al. [11] restricted the number of combined techniques based on findings from perception research and psychophysical experiments to evaluate their illustrative medical visualizations. They compared 24 illustrations generated according to their perceptual guidelines with illustrations that violate one or more of them. According to psychological experiments, more than 24 images are required, especially when using different focus objects as well as various models and viewpoints. Furthermore, they compared more than one technique and desired effect at the same time, and therefore, the results might not be expressive or traceable.

Other perception studies are applied to analyze the visual search in 3D visualizations. Burgert et al. [12], for example, performed eye tracking in order to quantify the effects of 3D medical visualization in the perception process. They used scanpath analysis, area of interest analysis, attentional landscapes, single subject and group comparison analysis to interpret the inspection strategies of experienced and young medical doctors. Eye tracking experiments are very sensitive to errors. Furthermore, these techniques are more used for analyzing users scanpaths or viewing strategies while exploring a scenario or searching for focus objects.

A study that discussed the detection of focus objects by using the sharpness of an object as the visualization technique was presented by Kosara et al. [13]. They evaluated the semantic depth-of-field (SDOF) for guiding the viewers' attention to specific information in an image. The sharpness depends not on the physical position of each object, but on its relevance. Their user study tested the ability of subjects to locate and estimate the number of sharp objects in a field of blurred objects with low-level attention.

3 Visual Search

In psychophysics, user studies are carried out to examine aspects of human perception. There are various user studies and theories to evaluate and explain the results. A general and influential theory for visual perception is the feature integration theory (FIT) introduced by Treismann and Gelade [15]. The theory is an important framework that characterizes the human visual perception and validates the performance in visual search tasks. According to the FIT, visual perception is characterized by two processing stages, whereat the first stage is parallel and preattentive and covers the target objects' detection in case that the target object stands out. The correct location of objects is determined in the second serial stage and is controlled by visual selective attention [15]. Some experiments indicate that the first stage may also depend on expectations and attention. Therefore, it is often referred to as *low level processing*. Later on, Wolfe et al. [16] introduced the guided search theory and showed that information from the first stage could be used to guide deployments of selective attention in the second stage.

Search task experiments are conducted to evaluate visual perception. A visual display comprising a number of elements to be searched, called targets, are shown to the subjects. They had to determine whether a target element is present or absent in a field of background distractor elements that were more or less similar to the target. All trials are shown randomly, and in half of the trials a target is present alongside one or more distractor elements. Such experimental studies focused on feature search and conjunctive search tasks. The feature search is characterized by the detection of a target that differs from the distractors in one feature. The subjects' performance will be unaffected of the display size that is defined as the number of distractors contained in the display. In contrast to that, a conjunctive search experiment requires the combination of features, for example of color and shape, to detect the target. Thus, the performance is dependent on the presented display size. To validate the results, the participants' reaction time and the accuracy were measured. The accuracy was assumed as a function of the display size.

As human subjects are often moody, subjective and differently motivated, Green and Swets [17] introduced the signal detection theory to validate the subjects' reaction tendency as well. They suggested to differentiate between the sensitivity and the psychological bias of the subjects by examining the correct responses and the false alarms. A false alarm occurs once the subject reacts on a signal which was not given. Further on, the signal detection theory can be used to assess the detectability of a target from a background of distractor elements.

Within the following sections we introduce our study design, conduction and results concerning the detection of focus objects of complex 3D medical visualizations. A focus object in our visualization corresponds to a target object used in visual search experiments. The study is based on the search task experiments and the results are validated with the signal detection theory.

4 Perception-Based User Study

We designed and performed a perception-based experimental user study that enables an evaluation of various visualization techniques used in a medical scenario. The primary goal is designing a framework to evaluate various visualization techniques that are used to support the focus object detection in therapy planning scenarios. The presented study is similar to that of Kosara et al. [14]. However, we use clinical datasets like Tietjen et al. [10] and Salah et al. [11]. In contrast, we designed a 2D conjunctive search and validated the subjects' results by using the signal detection theory.

The major challenge using medical visualizations is the arrangement of the structures. Anatomic structures are located very close to each other and there are several similarities between shape and representation of focus and context objects. As already mentioned, we analyzed the accuracy and the reaction time for finding the focus object within a specific therapy planning scenario. Therefore, we showed our subjects rendered images of 3D patient models derived from CT neck datasets. These representative visualizations are used in clinical routine to determine the size of the lymph nodes. The enlarged lymph nodes are the structures of interest that have to be detected very quickly. Krüger et al. [18] presented various techniques that may be applied to neck visualizations to support the detection of pathologic lymph nodes.

In our experiment, the enlarged lymph nodes serve as focus objects that have to be detected by the subjects. Thus, the subjects' task is to search for the conjunction of two features, a specific $structure \times size$. While all other structures are rendered with predefined standard colors and transparency values, we chose four representative illustration techniques

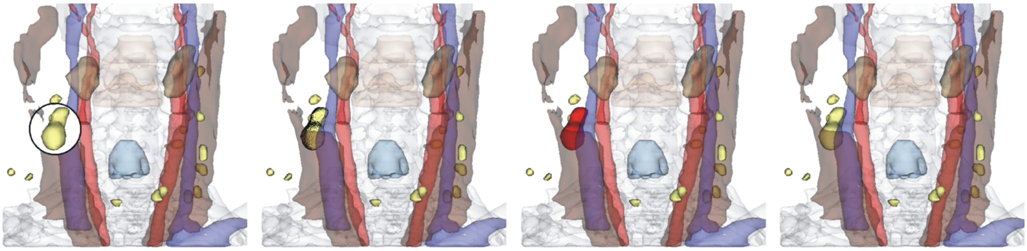

Figure 1: Rendered images of 3D models derived from clinical CT neck datasets are presented. The focus object and therefore the target to be searched is an enlarged lymph node that is either emphasized with cutaway, stippling, red coloring or it is rendered yellow and opaque like the other lymph nodes.

for the enlarged lymph nodes shown in Figure 1. Cutaway as a kind of smart visibility technique, stippling as a pen-and-ink technique, red coloring as a typical emphasizing technique and just the normal lymph node color with no special technique. According to that, the focus lymph nodes are either visualized contrary to or the same as the surrounding structures.

4.1 Hypotheses

Before the experiment is designed hypotheses need to be defined. The more hypotheses are drafted, the more trials are needed to properly test each of them. Because of this, experiments are often restricted to the most important conditions. Our experiment tested the following two conditions:

1. Emphasized enlarged lymph nodes are detected more often and faster than those without emphasis.
2. Cutaway views will be more efficient regarding accuracy and reaction time than stippling and red coloring.

These research hypotheses were postulated in advance. To accept or to falsify these hypotheses, a statistical significance test on their complementary theses was executed. This kind of test and the results regarding our hypotheses will be presented in Section 6.

4.2 Experimental Setup

Since our subjects had to determine the enlarged lymph node that is visualized with one of the four techniques, our study is a one-factorial study with four factor levels. Further on, this study follows a within-subject design. More precisely, we only examined one group of subjects where each subject is given the same kind and amount of stimuli. Thus, just a few subjects are needed and statistical tests are more comfortable in terms of conduction and analysis because of the same subject and stimuli dimensions. Still, it is recommended to test more than 30 subjects to expect a normally distributed and reliable sample size.

Beyond that, reaction time and accuracy are the two parameters that have to be measured due to our hypotheses. Accuracy is defined as the hit rate which indicates that the subject correctly detected the focus object. Since this type of evaluation is very time-consuming, it has to be certain that the experiment will give a valid and satisfying answer. Hence, to achieve a reliable result, many subjects are needed, and regarding their individual personal motivation it is recommended to use an experimental design based on the signal detection theory by Green et al. [17] described in Section 3. To validate the subjects' reaction tendency, we measure the false alarm rate that indicates whether the results are usable as well. Based on these parameters derived from the signal detection approach, target and noise stimuli are required which will be outlined in the next section.

4.3 Design of Stimuli

The stimuli are clinical neck visualizations (see Figure 1 and Figure 2), where an enlarged lymph node is the focus object and serves as the target and the other anatomic structures are the context and serve as distractors corresponding to the FIT [15]. The task of the subject is to detect the enlarged lymph node. We tried to gain a trade-off between the visualizations used in the clinical routine and the psychological conditions concerning the stimuli and the performed conjunctive search. Common psychological user studies present simple scenarios as stimuli containing basic shapes that differ strongly and that are equally distributed on the display. A major condition of such experiments is that stimuli are similar to reduce errors caused by the presentation of different scenarios. Contrary to that, our stimuli are neck visualizations from different patient-specific neck datasets. We used 16 datasets to provide a representative sample of the anatomic variety. Thus, we avoid that the results are strongly influenced by the peculiarities of one specific patient. Since our stimuli are patient-individual, we had to restrict the 3D scene to a representative number of structures and a specific viewport. These conditions are necessary to generate stimuli that are similar as possible but still representative images from clinical routine. Furthermore, the restrictions enable results that are traceable to our techniques and not caused by the stimuli appearance.

Though, each neck model contains bones, muscles, glands, pharynx, trachea, veins, arteries and a few lymph nodes as illustrated in Figure 2. Further on, almost the same amount of lymph nodes are visible to provide a search between various nodes. More similarities are obtained by presenting same sized images with a viewport defined by the muscles extensions and generated using a coronal orientation.

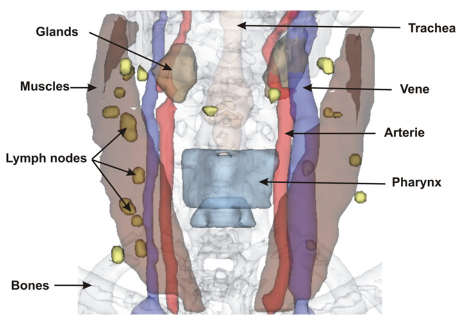

Figure 2: A neck visualization with all structures. The viewport is defined by the muscles extension.

Two kinds of stimuli are required based on the setup mentioned in Section 4.2. A permanent stimulus called noise and a second stimulus called target.

4.3.1 Target Stimulus

Target stimuli are rendered images of 3D neck models that contain one enlarged lymph node which is visualized with one of our techniques (see Figure 1). The images are rendered with orthogonal projection and we define an enlarged lymph node as a node that has a displayed size of > 30 pixels. This value represents an appropriate minimum size of an enlarged lymph node compared to the illustrated healthy lymph nodes' size.

4.3.2 Noise Stimulus

In contrast to that, the noise stimuli are images without any enlarged lymph nodes. They only contain healthy lymph nodes, as shown in Figure 2. Additionally, a noise stimulus for the techniques is included. That means, also a healthy lymph node may be emphasized with cutaway, stippling or red coloring to provoke false alarm and to evaluate the target detection capability of the techniques.

5 Conduction

Due to the effort associated with running an experiment, it is valuable to conduct a pilot study with a few viewers. We performed a pilot study with seven subjects, which allowed us to test and refine the experimental design before starting a full-fledged study with many more participants. We recruited 30 subjects from various parts of the university like psychology and engineering students, designers and a few medical experts who participated in the experiment.

All subjects were tested under the same conditions to produce meaningful results and to avoid discriminations. This means that all of them performed the experiment alone by daylight on a $26''$ monitor. No other processes were run on the computer during the experimental session. Moreover, the subjects were instructed in written form to provide the same initial requirements for every participant. A first practice session followed to ensure that the subjects understand the experimental task. We presented 1163 images that were generated of 16 different datasets, to obtain a representative sample of patient-specific neck visualizations. These

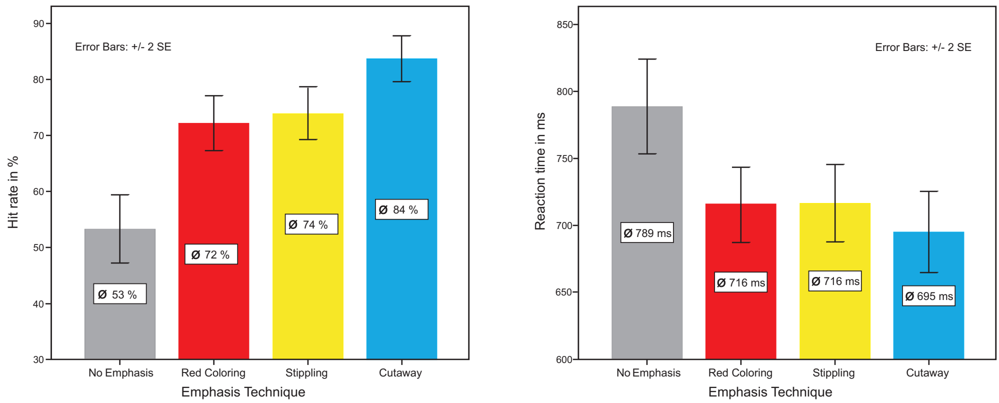

Figure 3: Left: The average results for hit rates in percent and Right: reaction times in ms for 25 subjects are illustrated in these bar charts. Each bar represents a technique and its height defines the averaged achieved results. The black bars on the top are standard error bars that indicate potential significance.

images were arranged into eight trials. Individual rests between the trials were integrated to avoid becoming fatigued or tired during the trials. Besides that, we randomized the presentation order of target and noise images but presented them equally often as the shown techniques. That means, the subjects saw $\approx 50\%$ target and $\approx 50\%$ noise stimuli. Each stimulus was presented $1.1s$ and had a displayed size of 512×512 pixels. A fixation cross followed each stimulus for a varying time of $0.75 - 1.25s$. This varying display time of the cross is a common psychological method, to avoid expectations and to promote the subjects attention.

The subjects interacted with the computer using an external mouse. Every time a stimulus was presented, the subjects had two choices:

1. press the left mouse button if an enlarged lymph node was present or
2. press the right mouse button if no enlarged lymph node was found.

Each experiment with a subject took one hour. After the experiment the subject had to fill out a short questionnaire asking for some personal details and subjective opinions regarding their preferred technique for target detection.

6 Results and Analysis

We recorded the false alarm rate, the average hit rate and reaction time for each subject and each presented technique. Since five subjects had a bad physiological sensitivity, we had to neglect them. With a hit rate below 50% they varied widely from the hit rate results of the other subjects and would have distorted the analysis. The measured data of the remaining 25 subjects was used to evaluate the techniques and to examine whether there is a significant difference between them. Their achieved false alarm rate of 8% represents a very good response bias and indicates that all results are valid.

Figure 3 illustrates the average values for all subjects. Each bar represents a technique and the height in the left figure illustrates the average achieved hit rate, and the required reaction time in the right figure. The black bars on top of each technique bar are standard error bars that may indicate a significant difference. If two error bars do not overlap, there will be a significant difference between those bars and the represented techniques. If they overlap, it does not mean that there is no significant difference. The bar charts show that the subjects achieved better results with each technique concerning the target detection than without emphasis of the enlarged lymph nodes. In detail, the subjects' hit rate was better especially with cutaway and they detected the emphasized enlarged lymph nodes faster. Cutaway seems to be the most suitable technique regarding accuracy ($\varnothing 84\%$) and detection time ($\varnothing 695ms$). In contrast to that, there is no apparent difference between stippling and red coloring, neither for hit rate nor for reaction time. However, since the appearance of scaled bar presentation can be deceiving, a

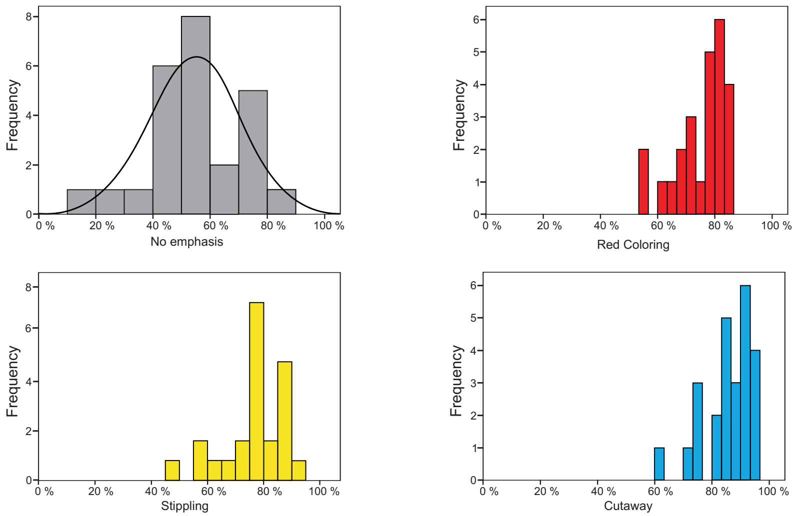

Figure 4: The distributions of the hit rates for each technique. As illustrated, the distribution is left-skewed for the techniques and normally distributed for the target stimuli without an emphasis technique.

significance test is necessary.

6.1 Significance Analysis

The major condition for choosing an appropriate significance test and consequently achieving correct and valid results are normally distributed data. Even though psychologists will assume a normal distribution if at least 30 subjects attended the experiment, it is recommended to verify the measured values. As shown in Table 1, we applied the Shapiro-Wilk test to the hit rate and reaction time results.

The test confirmed that the hit rate results are not normally distributed for the emphasis techniques. The bar charts in Figure 4 show that the distribution is left-skewed. That means, contrary to the normally distributed hit rates of the target stimuli without an emphasis, the subjects detected more lymph nodes in the emphasized visualizations. Especially, for the smaller ones better hit rates were observed. Since, the hit rate is not normally distributed for red coloring, stippling and cutaway we applied the non-parametric statistical Friedman test. In contrast to that the reaction time results are normally distributed, because just the times for the hits are considered. Thus, the one-factorial ANOVA test was applied. Both significance tests confirmed the existence of a significant difference with $p \leq 0.001$ between the applied techniques. Thereupon we compared the techniques with each other by using paired tests, Wilcoxon and the T-test with Bonferroni correction, as shown in Table 1. Both tests revealed significant differences between various pairs of techniques in accuracy and reaction time, except for red coloring compared to stippling ($p > 0.05$).

	Hit Rate	Reaction Time
Normal Distribution	Shapiro-Wilk ✗	Shapiro-Wilk ✓
Significance Test	Friedman p≤0.001	ANOVA p≤0.001
Paired Test	Wilcoxon	T-Test

Table 1: This table covers all tests that were applied to our measured data. Non-parametric significance tests were applied to the hit rate and parametric tests to the reaction time based on the results of the Shapiro-Wilk test for normal distribution.

Table 2 shows the paired test results and the corresponding effect sizes according to Cohen [19]. In

Pair		Significance	Effect Size	
No emphasis	Stippling	$p \leq 0.01$	$\delta_{acc} = 1.5$	$\delta_{rt} = 0.9$
No emphasis	Red coloring	$p \leq 0.01$	$\delta_{acc} = 1.4$	$\delta_{rt} = 0.9$
No emphasis	Cutaway	$p \leq 0.01$	$\delta_{acc} = 2.4$	$\delta_{rt} = 1.13$
Red coloring	Stippling	$p > 0.05$	$\delta_{acc} = 0.14$	$\delta_{rt} = 0.01$
Red coloring	Cutaway	$p \leq 0.01$	$\delta_{acc} = 1.0$	$\delta_{rt} = 0.28$
Cutaway	Stippling	$p \leq 0.01$	$\delta_{acc} = 0.9$	$\delta_{rt} = 0.3$

Table 2: This table covers the results of the pairwise technique comparison. Any significant difference is existent if $p \leq 0.05$. The effect size specifies the size of the significant difference between the techniques whereby values of $\delta < 0.2$ represent low and $\delta > 0.8$ high differences.

statistics, effect size is a measure of the strength of the relationship between two variables. In our example, the effect size specifies the size of the significant differences. The δ-value considers the standardized mean difference of each technique pair. Values of $\delta < 0.2$ represent low and $\delta > 0.8$ represent high differences.

Stippling compared to red coloring exhibits a very low effect size for accuracy $\delta_{acc} = 0.14$ and reaction time $\delta_{rt} = 0.01$ which indicates a very low difference as already expected from the bar chart illustration in Figure 3. Since the p-values of pairs with no emphasis confirm a significant difference and the δ-values show a high effective size we can confirm our first hypothesis introduced in Section 4.1. Every emphasis technique is better than no emphasis regarding the detection of the enlarged lymph node. The results show a preference for cutaway. With this technique, the subjects achieved the best results in accuracy and reaction time. We are therefore able to confirm our second hypothesis as well (see Section 4.1).

6.2 Questionnaire Analysis

As mentioned in Section 5, the subjects had to fill out a questionnaire which we analyzed as well. All subjects preferred any technique instead of no emphasis technique, which additionally confirms the first hypothesis. Their subjective ranking concerning the detection capability of the presented techniques, starting from the most supportive technique resulted in:

1. Cutaway
2. Stippling
3. Red coloring
4. No emphasis technique

Their overall technique rating corresponds to the computed results, except for the order of stippling and red coloring. Moreover, we compared the subjective opinions with the individually achieved results. One subject preferred stippling and achieved the best results with this technique. Although 24 of 25 subjects had the best performance with cutaway, eight of them favored another technique. That is a variance of 32% of their quantitatively determined most qualified technique.

7 Conclusion and Future Work

The key point of our paper concerns applying principles derived from psychophysics to visualization techniques for their quantitative evaluation. We therefore conducted an experimental user study using neck visualizations from clinical routine, a common therapeutic question and three simple representative visualization techniques. In detail, we validated the capability of the techniques by analyzing the accuracy and the required reaction time of the subjects to detect an enlarged lymph node rendered with each of those techniques. We defined capability via the performance of our subjects on a search task. This evaluation gives an insight into the experimental design, implementation and how to evaluate the observed data to gain significant results. We illustrated that psychological user studies can be applied to complex scenarios and application-oriented tasks considering a few conditions. The most important aspect is a setup that produces traceable results to the tested visualization techniques. Therefore, it is necessary to minimize potential error parameters like varying models, too small sample sizes or a comparison of several different visualizations. Minimization of varying parameters and a proper design enables objective, valid and reliable results for a statistical analysis. An appropriate sample size and study including adequate stimuli as well as suitable significance tests are essential. Although it is challenging to design a good perception-based experiment that will give robust answers to the question of interest, a well conducted study is usually worth the effort.

Our study serves as a framework for evaluating the capability of techniques concerning specific purposes in 3D therapy planning scenarios. Such user studies can improve the quality of visualization technique research, since we normally strive for effective illustrations. Moreover, the presented experimental setup is applicable to other medical domains, such as treatment planning and other techniques. Quantitative user studies that objectively assess the capability of the techniques and their combination regarding the desired visualization purpose, lead to more efficiently illustrations.

So far, we designed a framework for a user study with 3D patient-specific visualizations. Hence, our next experimental study will be an evaluation with emphasis techniques for focus and for context structures applied in one visualization. Such complex search tasks require appropriate study design that can be build upon our setup. Furthermore, the evaluation of techniques and technique combinations may lead to a classification of focus, focus-relevant and context techniques according to the used structure categorization.

References

[1] D. Bartz, D. Cunningham, J. Fischer, C. Wallraven. The role of perception for computer graphics. In *Eurographics State-of-the-Art Report 4*, 2008.

[2] D. A. Bowman, J. L. Gabbard, D. Hix. A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence: Teleoperators and Virtual Environments*, 11(4):404–424, 2003.

[3] J. E. Swan II, J. L. Gabbard, D. Hix, R.S. Schulman, K. P. Kim. A comparative study of user performance in a map-based virtual environment. In *Proceedings of the IEEE Virtual Reality 2003 (VR'03)*, 259–266, 2003.

[4] C. D. Jackson, D. B. Karelitz, S. A. Cannella, D. H. Laidlaw. The great potato search: the effects of visual context on users. In *Poster Proceedings of IEEE Visualization*, 2002.

[5] D. H. Laidlaw, M. Kirby, C. Jackson, J. S. Davidson, T. Miller, M. DaSilva, W. Warren, M. Tarr. Comparing 2D vector field visualization methods: a user study. *IEEE Transactions on Visualization and Computer Graphics*, 11(1):59–70, 2005.

[6] V. Interrante, S. Kim, H. Hagh-Shenas. Conveying 3D shape with texture: recent advances and experimental findings. In *Proceedings of Human Vision and Electronic Imaging VII (HVEI'02)*, 197–206, 2002.

[7] S. Kim, H. Hagh-Shenas, V. Interrante. Conveying shape with texture: experimental investigations of texture's effects on shape categorization judgments. *IEEE Transactions on Visualization and Computer Graphics*, 10(4):471–483, 2004.

[8] A. Bair, D. House. Grid with a view: optimal texturing for perception of layered surface shape. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1656-1663, 2007.

[9] F. Ritter, C. Hansen, V. Dicken, O. Konrad, B. Preim, H.-O. Peitgen. Real-time illustration of vascular structures. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):877–884, 2006.

[10] C. Tietjen, T. Isenberg, B. Preim. Combining silhouettes, surface, and volume rendering for surgery education planning. In *Proceedings of Eurographics–IEEE VGTC Symposium on Visualization*, 303–310, 2005.

[11] Z. Salah, D. W. Cunningham, W. Straßer, D. Bartz. Perceptually emphasized illustrative visualization for multiple objects. *Technical Report*,Wilhelm-Schickard-Institut, University Tübingen, 2008.

[12] O. Burgert, V. Örn, M. Gessat, M. Joos, G. Strauß, C. Tietjen, B. Preim, I. Hertel. Evaluation of perception performance in neck dissection planning using eye tracking and attention landscapes. *Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment (Proceed-*

ings of SPIE),6515:65150–65159, 2007.

[13] R. Kosara, S. Miksch, H. Hauser, J. Schrammel,V. Giller, and M. Tscheligi. Useful properties of semantic depth of field for better f+c visualization. In *Proceedings of the Joint EurographicsIEEE TCVG Symposium on Visualization (VisSym 2002)*, 205–210, 2002.

[14] R. Kosara, S. Miksch, H. Hauser. Focus + context taken literally. *IEEE Computer Graphics and Applications*, 22(1):22–29, 2002.

[15] A. M. Treismann, G. Gelade. A feature-integration theory of attention. *Cognitive Psychology,* 10:97–136, 1980.

[16] J. M. Wolfe, K. R. Cave, S. L. Franzel. Guided search: an alternative to the modified feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance,* 15:419–433, 1989.

[17] D. M. Green, J. A. Swets. *Signal detection theory and psychophysics.* John Wiley & Sons, Inc., New York, 1966.

[18] A. Krüger, C. Tietjen, J. Hintze, B. Preim, I. Hertel, G. Strauß. Interactive visualization for neck dissection planning. In *IEEE/Eurographics Symposium on Visualization (EuroVis)*, 295–302, 2005.

[19] J. Cohen. A power prime. *Psychological Bulletin*, 112(1):155–159, 1992.