

How Much Information Do You Remember? -The Effects of Short-Term Memory on Scientific Visualization Tasks

Wen Qi

Department of Industrial Design,
Eindhoven University of Technology,
P.O.Box 513, 5600MB, Eindhoven,
The Netherlands
w.qi@tue.nl

Abstract. In this paper we describe our experiences and lessons learned from two user studies about the effects of short-term memory on data analysis tasks with graphical user interface and virtual environment. Visualization is a powerful tool for representing scientific data for analysis purpose. Current research progress enables us to create a high fidelity visual representation of scientific data. However, the value of traditional graphical user interface and Virtual Reality (VR) as human computer interface in data analysis via visualization tool is questioned by the domain scientists. We carried out two user studies that asked users to perform data analysis tasks. The first user study compared the user performance on selecting and manipulating transfer function (TF) with different graphical user interfaces in volume visualization to search certain structures within a 3D data set. The second user study investigated the performance difference of immersive VR, fish tank VR and haptic-based fish tank VR systems on four generic visualization tasks. The two user studies indicates that short-term memory plays an important role in 3D user interaction during data analysis task. The pitfall of 3D user interface design for scientific visualization is that too many efforts have been put into the interface itself or technology, rather than designing user interface to match or amplify human capabilities, for instance the limit amount of short term memory and its importance in data analysis process.

1 Introduction

Three-dimensional (3D) interaction with scientific data involves studying visualization methods to faithfully represent data, on the one hand, and designing interfaces that truly assist users in the data analysis process, on the other hand. However, manipulating and viewing data in order to reveal valuable information effectively and efficiently is still not an easy task. Some of the bottlenecks are the huge amount of data and the ability to visualize and understanding these data. Scientific visualization investigates possible methods to translate data into a visible form that highlights important features, including commonalities and

anomalies. The progress in visualization by itself does not guarantee that users will be able to work more efficiently and effectively. In his review paper, “Top Ten Questions in Scientific Visualization”, Chris John pointed out that one of the ten problems in scientific visualization research is still human computer interaction (HCI) [1]. HCI research has become more and more important in the area of scientific data visualization. Since 3D interface is so immature, developing technology is still the main research activity. 3D interface developers argue that interfaces or interaction techniques are extremely complex to design and make, and that the demand for the function that interfaces can provide has always outdriven the demand for better usability. Until recently, the focus of 3D user interfaces and interaction research has shifted from emphasizing design hardware (e.g. development of 6 DOF input devices and 3D output devices) or software toolkits to paying more attention to studying the user with currently available knowledge from cognition, psychology.

In this paper, we specifically study how the human computer interface influences user performance in scientific visualization through two kinds of user interface: graphical user interface for TF specification and VR as interface for generic visualization tasks.

2 User Study I

The TF is a critical component of the direct volume rendering process. It specifies the relation between scalar data (e.g. densities measured by CT or MRI imaging), and possibly also their first- or second-order derivative values, and optical characteristics, such as color and opacity [2]. The most common method is the trial-and-error method. It involves manually editing the TF by modifying a graphical curve and/or by adjusting numerical parameters, and visually inspecting the resulting image.

The trial-and-error method provides only minimal visual feedback in TF specification, and we therefore have devised an experiment by which we explore the usefulness of additional information in this trial-and-error method. More specifically, we aim to assess the effects of the following additional feedbacks: 1) data-dependent information, such as the histogram; 2) data-independent information, such as suggested or standard TFs. We also wanted to investigate the effect of a graphical user interface with a limited number of DOFs.

Our experimental hardware setup consists of a DELL graphics workstation (Pentium IV, 2.4 GHz, 512 MB RAM, ATI *FireGL 4* graphic card). The TF specification is implemented by means of a texture look-up table with the help of hardware-accelerated 3D texture mapping.

The experiment involves five interface conditions. The baseline interface with free-style control, referred to as condition 1, consists of parts 1a and 1b in Figure 1. With this free-style interface, a user has full control over the TF. The part 1a allows the user to manipulate the TF by creating and moving control points of a piecewise linear function along the horizontal and vertical direction within a 2D interaction area. The panel 1b controls the course of the experiment.

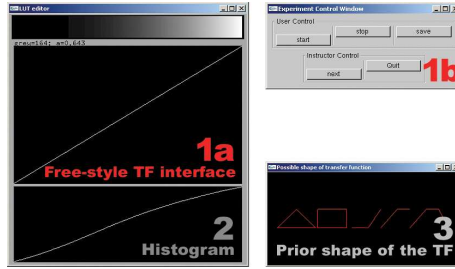


Fig. 1. The user interfaces for experimental conditions 1 (part 1), 2 (part 1+2), 3 (part 1+3) and 4 (part 1+2+3)

Experimental condition 2 includes data-dependent information, and consists of parts 1 and 2 in Figure 1. A cumulative histogram and free-style TF interface are presented at the same time. Experimental condition 3 includes data-independent information, and consists of parts 1 and 3 in Figure 1. In condition 4, both data-dependent and data-independent information are offered, so that all parts in Figure 1 are presented. The interface for our final condition 5 is shown in Figure 2. It is a user interface that allows for a number of piecewise-linear TFs, and that does not provide (data-dependent) histogram information. Each kind of TF is represented by a graphical icon. The available TF graphical representations have only few control points and limited DOFs. This implies that the shape of a graphical representation cannot be altered.

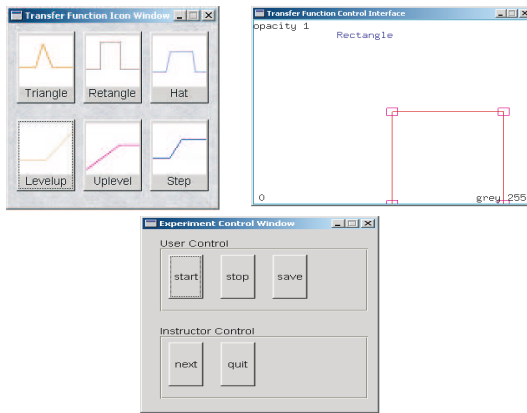


Fig. 2. The user interface for experimental condition 5

There were 13 subjects in the experiment. Each subject participated in all conditions (within-subject design). Each participant performed four tasks in each

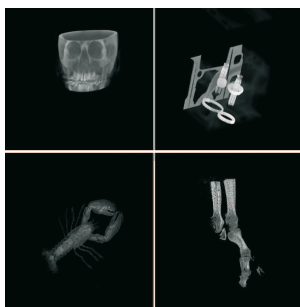


Fig. 3. The required structures being rendered with four data sets

of the five interface conditions. Each task involved a different data set (“head”, “engine”, “lobster”, and “foot”), and required the subjects to visualize a pre-described structure within the data as good as possible (Figure 3). For example, one task is to ask the subjects to find out one TF that can create a 3D image that contains only the bone structure of the skull. The dependent variables recorded during the experiment were the following:

- the total number of the mouse clicks during a task;
- the number of the mouse clicks for each icon in condition 5;
- the time needed to finish an task;
- the rendered image produced at the end of a task.

The subjects were given several questions to answer after the experiment. More specifically, the questionnaires contained the following parts:

- Their agreement or disagreement with general usability statements about the system, such as: “It was easy to use the system”;
- Questions addressing the usability of the individual interfaces.
- In order to assess the task performance, they were asked both to rate the output images individually and to express their preferences for all pairwise combinations of output images;

2.1 Results

There are no significant differences of the time spent among the five conditions for the “lobster”, “head”, and “foot” data set, respectively. For the “engine” data set, the order of the conditions with respect to performance time is one to five from faster to slower. For the “engine” data set, “Analysis of Variance” (*ANOVA*) with repeated measures indicates that there are significant differences among the five interface conditions, $F(4, 52)=9.141$, $p < 0.05$. The post hoc test shows that between condition 1 and 5, condition 2 and 5, and condition 3 and 5, there are significant differences in the response time. Condition 4 and 5 do not demonstrate

a significant difference. *ANOVA* on the number of mouse clicks for all four tasks shows that there are no significant effects in either of the four data sets among the five conditions.

This subjective evaluation is based on the answers from the questionnaire. The results are mainly summarized along several important characteristics, such as, effectiveness, efficiency, satisfaction and overall image quality.

The difficulty of the tasks. The tasks with these four data sets presented different levels of difficulty for the subjects. The task with the “engine” data set has been recognized as the easiest one. The most difficult task is the one with the “head” data set.

The task performance by the image quality. The image quality is scaled by how close the produced image is compared with the original goal. With the “engine” and “head” data set, the task performance in condition 5 is the worst. However, the best performance has been achieved in condition 5 with “lobster” data.

Effectiveness. In terms of effectiveness, Condition 4 gets the highest rank, with condition 2 as the close second. Condition 5 is considered as the least effective.

Efficiency. Efficiency was defined as “how fast the user feels that he can finish the task” with each of those interfaces. Condition 4 again scores best, while condition 5 is the worst.

Satisfaction. The subjects are requested to give an evaluation on “the arrangement of the five interfaces”. The results illustrate that condition 4 still gets the highest appreciation. Condition 5 scores higher in this attribute than in the previous attributes.

3 User Study II

The second user study compares three kinds of VR systems: HMD based VR, fish tank VR and fish tank VR with haptic feedback. Relative performance of these systems is compared over four generic tasks involving the visualization of volumetric data. The rendering paradigms are only tested in their most common configurations: inside-out for HMD and outside-in for fish tank.

3.1 Apparatus

All three systems display the volumetric data using the Visualization Toolkit (VTK). To enable real-time interaction, we chose Marching Cubes as the primary algorithm and render isosurfaces of the volumetric data.

Immersive HMD VR system. The immersive VR system uses a V8 HMD from Virtual Research System. Head tracking is performed via a 3rdTech HiBall tracking system. The outward-looking HiBall sensor is mounted on the back of the HMD (Figure 4b). We used a DELL Precision 530 (dual 2.8-GHz Xeon with



Fig. 4. Head-mounted display system: head-mounted display with head sensor

2GB RDRAM) and an NVidia *Quadro FX 1000* graphics card. The working space for a user in this VR system is about 4.5 meters wide by 7 meters long by 4 meters tall (15 feet \times 23 feet \times 13 feet) as shown in Figure 4.

Fish-tank VR. The second VR system is based on the concept of fish tank VR introduced by Colin Ware [4]. The central computing platform of this VR system is identical to the HMD system with the following additional components:

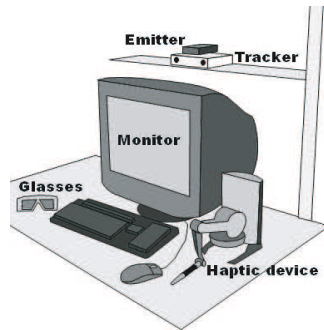


Fig. 5. A diagram of the fish tank VR system

- A 17-inch CRT monitor with a refresh rate of 100 Hz to support stereo display, an infrared emitter and shutter stereo glasses from StereoGraphics.
- A PHANTOM DesktopTM haptic device. In fish tank VR mode, the PHANTOM was used to rotate the volume around its center only.
- A DynaSight 3D optical tracker for measuring the 3D position of a target (reflective disc) attached to the front of the stereo glasses.

The hardware components are organized to enable accurate and easy calibration (Figure 5).

Fish-tank with Haptics. The fish tank VR with haptics prototype uses the same hardware setup as the fish tank VR system, except that the PHANTOM also provides force feedback. An axis-aligned on-screen icon follows the stylus's motion in 3D, producing an effect similar to using a mouse to control the on-screen cursor. The haptic representation of volumetric data employed different force models for different objects within the volume: viewers felt the outside surface of spheres and ellipsoids, but the inside of long curved tubes and cylinders.

3.2 Data and Tasks

Simulated volumetric data are generated to act as trials during our studies. A random number of two to four types of differently-shaped objects (sphere, ellipsoid, cylinder, and curved tube) are inserted at random positions (details are found in [5]). The objects' properties (such as size, shape) and the density of each volume form experimental conditions that vary between trials. The bounding box of the volume is subdivided into eight equally-sized subregions within which object density may differ. Subregions are labeled with unique numbers (1 through 8).

Subjects are asked to complete four tasks within each trial. Each task involves judging the properties of a specific object or of the overall volume, specifically:

- *Shape task*: Subjects identify the number of differently-shaped objects within the volume and name the objects.
- *Size task*: Subjects report how many differently-sized spheres exist.
- *Density task*: Subjects identify the densest subregion in the volume.
- *Connectivity task*: Subjects report how many curved tubes exist in the volume, and then determine which subregion(s) the longest curved tube passes through.

3.3 Results

Forty subjects volunteered for our experiment, 33 males and 7 females. The subjects were randomly assigned into one of the three display system groups. Two measures of performance were derived for each trial a subject completed: response time rt and error rate \widehat{P}_e . A single rt value representing the total time in seconds needed to complete all four tasks was captured for each trial. Four separate \widehat{P}_e values for the four tasks subjects completed were also generated.

For rt statistics, trials were divided by display system (HMD, fish tank, or fish tank with haptics). For \widehat{P}_e statistics, trials were divided by display system (HMD, fish tank, or fish tank with haptics) and task (shape, size, density, or connectivity). The logarithm of rt in average and \widehat{P}_e for different conditions were then compared (using *ANOVA* with a 95% confidence level for $\lg(rt)$, Chi-Square test for \widehat{P}_e).

In summary, the following significant differences in performance were identified:

1. The HMD group had significantly longer rt , compared to both the fish tank and the haptic groups. Using haptics in the fish tank also resulted in longer rt .

2. For most tasks (counting number of different shapes and number of curved tubes and finding the densest subregion), the HMD group had higher \widehat{P}_e , compared to both the fish tank and haptic groups.
3. In counting number of different sizes about sphere objects, none of these three groups is accurate. The HMD group made significantly more errors than the fish tank and the haptics group in case only one size of sphere was present. In case more than one size was present, subjects from all three groups mainly underestimated the number of sizes on average.
4. For identifying the subregions that the longest tube passes through during the connectivity task, the HMD group had higher chances in both missing the right subregions and misjudging the wrong subregions, compared to both the fish tank and the haptic groups.

Subjective measurements were obtained through analysis of the post-experiment questionnaires. The answers indicated that overall, subjects preferred the haptic and HMD VR systems due to perceived ease of use, presence, and immersion. We summarize our findings over the following categories of questions we asked.

Overall perception of the VR systems. This category asked about the perception properties and characteristics of a VR system, including the immersion, presence, depth, and spatial relationships. For the question: “the extent that you felt you were within a virtual environment”, the HMD system ranked significantly higher than the fish tank and haptics systems, $F(2, 37) = 5.481, p = 0.008$, with a post-hoc comparison between HMD and haptic of $p = 0.006$. There was also a significant difference on the question: “the extent you had a sense of acting in the virtual space, rather than operating something from outside”. The HMD system ranked significantly higher than the other two systems, $F(2, 37) = 15.666, p = 0.001$.

Usability of a VR system. The ease of learning and using a VR system is the main focus of this category. Although subjects from all three groups felt their system was easy to use, the HMD group ranked highest for the perceived difficulty in carrying out their tasks. Moreover, HMD subjects reported a significantly higher demand for memorizing than the other two groups, $F(2, 37) = 5.534, p = 0.008$. Finally, HMD subjects were less confident about the accuracy of their answers, $F(2, 37) = 5.521, p = 0.008$.

4 Discussion

A further analysis through the knowledge of engineering psychology and perception explains the reasons of such performances for those tasks. Human beings have two different storage systems with different durations: working memory and long term memory [6]. Working memory is the temporary, attention-demanding store that a human user uses to retain new information until he or she uses it [7].

A human user uses working memory as a kind of "workbench" of consciousness where he or she examines, evaluates, transforms, and compares different mental representations. A human user might use working memory, for example, to carry his or her mental arithmetic or predict what will happen if he or she set a TF one way instead of another. Working memory is used to hold new information (for instance the resulting image with one TF setting) until a human user gives it a more permanent status in memory, that is, encodes it into long term memory.

Several experiments have demonstrated the transient character of working memory [8]. Estimates generally suggest that without continuous rehearsal, little information is retained beyond 10 to 15 seconds. This transient character of working memory presents a serious problem for those work domains/tasks when information can not be rehearsed because of time constraint [9].

Working memory is also limited in its capacity (the amount of information it can hold) [10]. And this limit interacts with time. Experiments show that faster decay is observed when more items are held in working memory, mainly because rehearsal itself is not instantaneous [11]. The limiting case occurs when a number of items can not successfully be recalled even immediately after their presentation and with full attention allocated to their rehearsal. The limiting number is sometimes referred to as the memory span.

Task analysis can suggest that TF specification is a task that puts high demands on working memory from a user. When a user uses a TF interface to search for required results or structures, he or she continuously inputs different parameters for the TF through the interface and judges whether or not the corresponding rendering results are the ones he or she needs. Often he or she needs to retrieve previous settings that are better after comparison. A user needs to perform so many interactions (modifying the TF parameters and observing the corresponding visual feedbacks) and has to hold mapping information between TF setting and visual feedback with respect to each data set in working memory, which introduces more possibilities for error. The graphical user interface of trial-and-error method proposed by many researchers did not present any mechanism to relief the workload of memory. Clearly, the limited capability of working memory has a major impact on the effectiveness and efficiency of an TF interface. On the other hand, it also indicates that a user interface will be more efficient and effective if it can relieve the workload of a user's working memory during the specification process.

The second user study showed that immersive virtual environment highly increased the immersion and presence experiences of the subjects, but asked subjects to memorize the data structure heavily so that they could perform the tasks. In the absence of an overview capability, subjects were forced to make an internal representation of the total volume; the dense nature of the data removed visible landmarks that can normally provide such a frame of reference. The time needed to complete tasks in the HMD system was significantly longer, compared to both the fish tank and the fish tank with haptics systems. It is due to the reported inability of HMD subjects to remember where they had previously seen target items within the volume because of the high density of data sets. They

would often have to re-search the volume for objects they had previously located, but had “lost” as they walked into a different region.

5 Conclusion and Future Work

In this paper, based on the results of two user studies, we emphasize the importance of short-term memory on the data analysis tasks and propose the factor of short-term memory should be taken into account while designing 3D user interface for data visualization applications. The perception and cognition skills of a human user play an important role in determining the success or failure of a 3D interface or interaction technique.

Acknowledgments. This project is funded by SenterNovem as part of the Innovation Oriented Research Program (IOP-MMI) supported by the Dutch Ministry of Economic Affairs. We would like to thank the NIH National Research Resource in Molecular Graphics and Microscopy at the University of North Carolina at Chapel Hill, which is supported by the NIH National Center for Research Resources and NIH National Institute of Biomedical Imaging and Bioengineering grant 5-P41-RR02170-21.

References

1. Johnson, C.R.: Top scientific visualization research problems. *IEEE Computer Graphics and Applications* 20(6), 2–6 (2004)
2. Lichtenbelt, B., Crane, R., Naqvi, S.: *Introduction to Volume Rendering*. Prentice-Hall, New Jersey (1998)
3. Qi, W., Martens, J.-B.: Transfer Function Specification with Trial-and-error Interface for Volume Visualization: A User Study. In: *The IASTED International Conference on Human Computer Interaction 2005*, November 14–16, Phoenix, USA pp. 50–57 (2005)
4. Ware, C., Arthur, K., Booth, K.S.: Fish tank virtual reality. In: *Proceedings of CHI 93*, pp. 37–42. ACM and IFIP (1993)
5. Qi, W., Taylor, R., Healey, C., Martens, J.B.O.S.: A comparison of immersive HMD, fish tank VR and fish tank with haptics displays for volume visualization, In: *Proceedings of the 3rd symposium on Applied perception in graphics and visualization 2006*, pp. 51–58, ACM (2006)
6. Wickens, D.C., Hollands, G.J.: *Engineering Psychology and Human Performance*. Prentice-Hall, New Jersey (1999)
7. Baddeley, A.D.: *Working memory*. Clarendon, Oxford (1986)
8. Brown, J.: Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology* 10, 12–21 (1959)
9. Moray, N.: *Handbook of perception and human performance*. Wiley, New York (1986)
10. Baddeley, A.D.: *Human memory*. Allyn and Bacon, Boston (1990)
11. Melton, A.W.: Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and verbal Behavior* 2, 1–21 (1963)
12. Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I.: *3D User Interfaces: Theory and Practice*. Addison-Wesley, New Jersey (2004)