

Physical User Interface Design for Visualizations

Rebecca Dreezer
Supervised by Melanie Tory

August 22, 2008

In technology, efficiency is maintained by improving the interaction between the human and the computer. Computer Scientists are concerned with increasing the usability of user interfaces. We ran two empirical research studies, Comparing Multi-dimensional User Interface Controls, related to physical user interface design, and Comparing Graphic Landscape Representations,. The former focused on testing how efficient users are at performing a simple task using various input devices, while the latter determined the users' ability to remember various types of landscapes. These studies were relevant as they had applications in fuel cells, and medical imaging. Immediately following are two papers, each one outlining one of the studies.

Acknowledgements:

I would like to thank the Computer Research Association's Committee on the Status of Women in Computing Research (CRA-W) for providing me with this opportunity to participate in the Canadian Distributed Mentor Program (CDMP). In addition, this work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC). I would like to thank my mentor, Assistant Professor Melanie Tory, and Colin Swindells for their guidance and assistance. I'd also like to thank the whole Visual Interaction Design (VisID) Group at University of Victoria for their cooperation and support.

Comparing Multi-dimensional User Interface Controls

1. Introduction

1.1 Purpose and Introduction

This empirical research study tested participants' interaction for techniques to physically interact with visualizations of high dimensional data (i.e. representing complex data using novel computer graphic and interaction techniques). The purpose of this human computer interaction study is to compare the effectiveness and usability of three user interfaces for manipulating parameters: a graphic input device, a physical input device, and a touch input device. Participants are required to perform a simple colour-matching task by moving four sliders on a computer monitor in order to match the colour on the screen (Fig. 1).

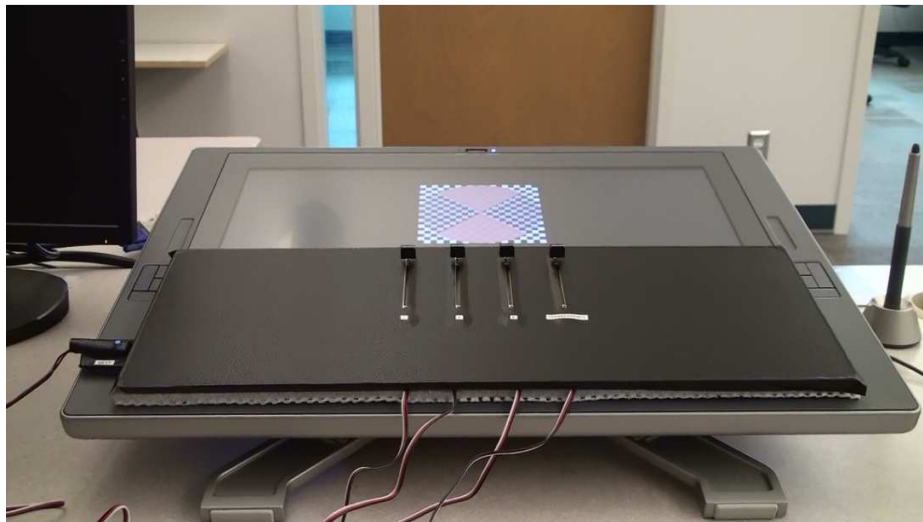


Fig. 1 Manual sliders placed on top of monitor. This was the setup used to run the study when the participant was working with a physical input device.

This research project has broader implications in the sciences. For example, when trying to enhance medical images in order to highlight certain areas of X-rays, it is important to know what type of interface would enable the technician to complete the task effectively, and would require him looking at his hands as little as possible and at the image as much as possible.

1.2 Related Work

When I began working on this study, all of the programming was completed (except for some debugging). As such, and because I was unfamiliar with Visual Basic, with which the program was coded, I did not look at the code. Indeed, for this project my primary role was conducting the experiment, and working with the videos recorded during the experiment. Because I entered into the project quite late in the game, I was unfamiliar with the papers on which this study was based, and the related work.

2. Experimental Design

2.1 Participants

In all, 13 participants were tested for this experiment. Originally, there were 12 participants tested. Unfortunately, one participant was leaning far back in the chair and her video footage is unusable for analysis because of the distorted small appearance of the monitor. Once this video was determined to be unsalvageable, the study was run on a 13th participant. All participants for this study were undergraduate students or masters students from various faculties who were living on University of Victoria campus during the summer 2008. Most of the participants were friends with the person conducting the study. Because of their affiliation to the person conducting it, I felt that they were more likely to work harder at task.

2.2 Apparatus

2.2.1 Computer

The study was conducted on an Intel Core 2 Duo desktop processor running at 2.2 GHz with 2 GB RAM, and an nVIDIA GeForce 8800 GTS video card. The computer was running Windows Vista Ultimate 32-bit operating system. The study was displayed on a Cintiq 21UX Wacom Tablet. The screen size of the tablet was 21.3” on the diagonal, and the screen resolution was 1600 x 1200 (LCD with UXGA resolution).

2.2.2 Input Devices

Participants interacted with the software by using three types of input devices: a graphic device, a touch device, and a physical device. For an equal number of trials, the participant used either a standard mouse (Fig. 2a), or a direct pen input (Fig. 2b), or manual sliders (Fig. 2c). The mouse used was a Microsoft Wheel Mouse Optical. The pen used was a Cintiq 21UX Grip Pen. The board for the manual sliders was custom-made by Colin Swindells. The manual sliders fit perfectly over the sliders on the screen. These sliders were handled using either one hand or both hands, and more than one slider could be controlled at a time.



(a) Mouse



(b) Pen



(c) Sliders

Fig. 2 Input devices used in the study

2.2.3 Software

The software used to run the study was coded in Visual Basic. Two programs were used in order to analyze the video records of the experiments. The first program was Locarna's Eye Gaze Calibration [1]. This program is used to generate calibration files of the video, and to produce a video file with fixations included. The second program was Locarna's Video Annotation. This program was used to annotate the fixations for the video file produced using the Eye Gaze Calibration software.

2.2.4 Eye tracking

The eye tracker used in this study was developed by Locarna. This consisted of a special pair of glasses with two cameras attached. The first camera records where the eye is looking, and the second camera records the movement of the eye. These glasses are attached to a small screen which records the video and saves it. These cameras are then calibrated to produce an accurate representation of where the participant is looking while performing the task.

2.3 Study Environment

Each study lasted approximately 50 minutes. The study was conducted in a small office. The door was closed to ensure there were no noise or movement distractions from outside. The participant was sitting upright in a chair in front of a desk containing the computer monitor. The monitor stand was consistently at a 10 degree angle. In order to get the same view of the screen for all participants, the desk was raised or lowered depending on how tall they were. While the participant was completing the study, the conductor was in the room, switching the input devices when necessary and making sure that the participants eye tracker was working and had not shifted.

For the duration of the study, the blinds were shut tightly, and the lights were turned off. The darkness in the office aided in minimizing the glare on the screen. Having a glare would perhaps distort the appearance of the colours on the screen, and inhibit the participants' ability to see the screen. It would distract them from the task, and it would increase the time it took them to complete the tasks.

3. Task

3.1 Overview

Using three interface devices, participants were asked to move the four sliders on the screen up and down in order to change the colour of the top and bottom quarters of the circle so that they matched the colour of the left and right quarters (Fig. 3).

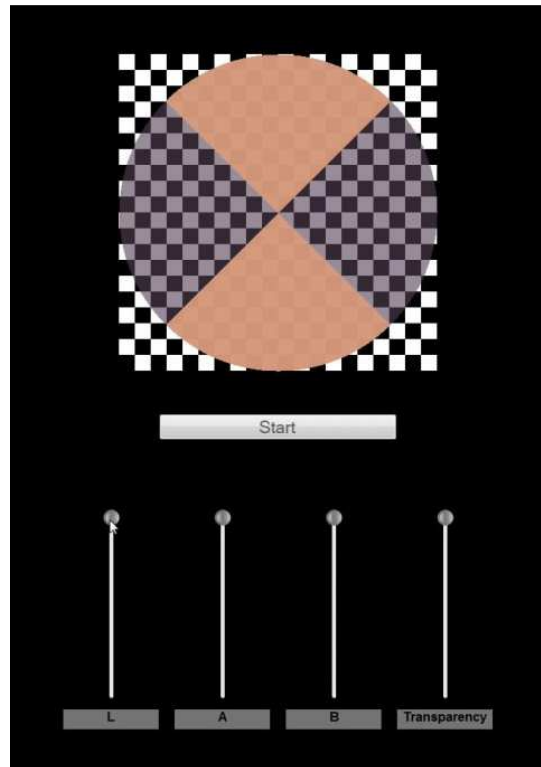


Fig. 3 Screen printout of a typical trial during the practise or real task.

3.2 Colour Space

The four sliders represent a Lab linear colour space [2]. Starting at the left, the first slider controls the luminance. The second slider controls the red to green spectrum. The third slider controls the yellow to blue spectrum. Lastly, the fourth slider controls the transparency. This colour space was chosen because it is a linear colour space. This colour space can be controlled by a slider because all the colours are equally spaced out in intervals of the same size. The Lab colour space is designed to approximate human perception. Indeed, the luminance is an accurate representation of how humans perceive lightness.

3.3 Task Specifications

When the participants first entered the testing room, they were asked to sit down at the desk, and they were outfitted with the eye tracking glasses. They began the study by performing a short colour vision test. This test showed the participants from the Ishihara Test [3] in order to ensure that they were not colourblind. If the participant happened to be colourblind, he would not be able to see the correct colour that needs to be matched, nor would he be able to create a colour

that matches it. Or, if he was able to perform the task, his completion times would be larger than the others and skew the results.

At the end of the colour vision test was a screen with nine black squares spread out over it. The conductor would call out a number (in increasing order), and the participant would move his eyes to focus on the corresponding black square, while making sure his head was stationary. These fixations were used as calibration points because the participant looked at every area of the screen quite clearly. If the conductor was unsure as to how well the participant focused on each point, or if the participant moved his head during the task, additional calibrations were performed at the end of the task, or at the end of a set. Before the screen with points was integrated into the task, a few participants were asked to stand next to desk in order to focus on calibration points which were written onto a whiteboard.

Next, the testing program was opened. The conductor explained to the participant that this was a colour matching test. The movement of the four sliders was demonstrated, and any questions were answered. When the participant understood the task, he was told to press the start button and begin a practise test. In the practise test, he matched two colours using each of the input devices. After he had matched two using an input device, the conductor would set up the next input device for him. This practise task was given in order to minimize the effects of a learning curve.

The participants were then ready to begin the actual task. They were given three sets of ten colour-matching trials each to complete. Each set was completed using a different input device. The participants were told that they were allowed to take small breaks between the sets of trials if they needed. In each trial, the participant was given a different colour to match. When the participant's colour came within certain boundaries of the stationary colour, the computer outputted a message indicating that the colour had been matched. The order of the ten colours was randomized for each set. However, the order remained constant for all users. After matching a colour, a message would appear on the screen indicating that the colour had been matched. The participant would then have to slide the sliders up to the top before he could press 'next' and begin the next trial. However, when the manual sliders were being used, the user pressed a blue button on the side of the monitor frame labelled 'next' in order to proceed.

While the participants completed each trial, their results were being recorded into a log. Each log contained the number of the slider that was being moved, the start time and end time it was in use (in ms), and the start and end value of the colour. The results from all the trials were automatically compiled into a summary file. This log file contained the trial number, the total time of completion (in ms), the target values for each of the four sliders, and the values that the participants got for each of the sliders.

Since there are three input devices, there are 3! (or 6) possible orders in which to use all input devices. Thus, twelve participants were tested. The order in which they used the input devices for the task was the same order in which they used the input devices to complete the practice task. The 13th participant completed the task in the same order as the person whose video recording she was replacing.

4. Analyzing Video Footage

4.1 Calibrating the File

Firstly, the Locarna "Eye Gaze Calibration" program was loaded. Then, the raw video from the recording machine (NAME.AVI) was loaded into the program as the input file. On the 'Calibrate' tab, the button 'create new calibration' was selected. The next step was to begin playing the video. Under the 'Settings' tab, the editor could change the darkness threshold. This was necessary if the program was selecting too much of the eye as the pupil, or was not selecting enough. Whenever the eye was focused on one of the nine points, it was necessary to select the point as a calibration point by left clicking and then right clicking on it (Fig. 4). A minimum of five calibration points had to be used. Points were only chosen if the eye was clearly focused on the point, and if the dot in the pupil was accurate. After all the calibration points were selected, the editor would hit the 'calibrate' button and let the rest of the video play through until the end.



Fig. 4 Locarna Eye Gaze Calibration. The eye is focused on the first point, which is selected.

Immediately after the 'calibrate' button is pushed, it produces a crosshair on the rest of the video while it is playing. Calibrating the file creates a NAME.cal file. It also generates a NAME.csv file containing frame numbers, time (estimated), and the x and y coordinates of the point. In addition, it generates a NAMEout.AVI file with a crosshair where the eye is looking throughout the video.

4.2 Calculating Fixations

The third tab in the Eye Gaze software is called 'Process.' It has an option 'calculate fixations.' The next step is to select the NAME.csv file, and it will output a message that says, "Fixation File Created." This has just created a file named NAME.Fix.

The button next to that reads 'Make Video with Fixations.' It prompts the editor to open an AVI file: NAME.AVI. Then, it asks the editor to choose to save the file as: NAMEout.AVI. Once the whole video plays out until the end, it creates a file NAMEout.AVI that has a circle on the crosshair whenever the participant is fixating.

4.3 Video Annotation

The first step is to open Locarna's Video Annotation. The file NAMEout.AVI should be loaded, and the tags that are needed (other, slider, slider 1, slider 2, slider 3, slider 4, colour wheel) are added to the repository. One at a time, the fixations are labelled (Fig. 5). When they are all complete, the software says "You have labelled all fixations." Running this software has created a CSV file named NAMEoutAnnotation.csv. The CSV file was then edited into an excel file that would be accepted into SPSS for analysis.



Fig. 5 The first fixation is shown on the screen. The editor is given the option to select a tag, at which point it will advance to the next fixation.

While the fixation points were being tagged (annotated), the editor created a log file named NAME notes.txt. This file contained the frame numbers of when each set (including the practise set) began and ended, as well as when each trial within each set began and ended.

5. Preliminary Results

The relevant information from the logs (the fixation video logs and the initial logs) was compiled into excel charts, which were exported into SPSS. The results were analyzed statistically using repeated measures analysis of variance (ANOVA) and boxplots.

While using the manual sliders, most of the participants did not look down at their hands at all. They could focus on the colour wheel. While using the mouse, the participants would look down at their hands to select a slider, and then look at the colour wheel while sliding it up and down. While using the touch pen, the participants would look down at their hands when changing sliders, and also sometimes while sliding a slider up and down. The results showed that even though the participants performed the worse while using the touch pen (spent a much more significant amount of time looking down at their hands), they still managed to complete the touch pen set in about the same amount of time as they completed the other sets.

6. Conclusion and Future Work

One of the reasons that the participants performed worse while using the touch pen was that it was an input device that they were unfamiliar with. The participants performed well while using the mouse because they were familiar with the mouse. They performed well with the manual sliders because they could keep both their hands on the sliders, and they could feel the location of the different sliders, and thus they did not need to look down.

One of the limitations on this study was that the Locarna software was still in somewhat of a development stage, and was therefore not bug-free. After using the annotation software to tag a few hundred fixations, the program would crash and the person editing the video (editor) would need to restart it and locate the fixation at which they left off in order to resume tagging.

In addition, the video analyzing process was very tedious. If the calibration was bad, then the video was useless as the fixations would be focusing on areas where they were not actually looking. In order to prevent this, sometimes it was necessary to chop a video in order to extract a good set of calibration points from later on. Then, the chopped video would be loaded into the Eye Gaze software as a file and a CAL file would be generated. The original video file could then be loaded into the Eye Gaze software, and instead of choosing to create a new calibration, the editor could load the previous chopped CAL file as the calibration file. In addition, it took a very long time in order to tag all the fixations; in fact, it was quite an inefficient process. More waiting occurred whenever it was necessary for the program to run through the whole video.

Occasionally, when the participants were asked to use the sliders, the sliders would not work. Sometimes the software would need to be reloaded, or the computer would need to be restarted. I am unsure as to why this error occurred. Having to restart a computer in the middle of a study loses the flow of the participant, and changes the atmosphere, which could ultimately skew the results on trials following it.

From talking to participants after they had completed the study, it was clear that some participants thought the test was too long, and were very frustrated with it. As a result, they lost focus on the last few trials, and the trials did not represent their actual capabilities. Perhaps in future, a similar study could be run where it is mandatory to take breaks, or where the task is shorter so that the participant will have the equal amount of focus and the same attitude throughout.

References

[1] Locarna. Found at <http://www.locarna.com/>

[2] Wikipedia. Lab Color Space. Found at http://en.wikipedia.org/wiki/Lab_color_space

[3] S. Ishihara. Ishihara's test for colour blindness. Found at <http://www.toledo-bend.com/colorblind/Ishihara.html>, 2007.

Comparing Graphic Landscape Representations

1. Introduction

1.1 Purpose and Introduction

As is discussed in Tory et al. [2], there are certain tasks in which using landscapes to represent multidimensional datasets is beneficial. Tory et al. believes that since landscapes may show a continuous distribution of values, as opposed to discrete data points, using landscapes makes it easier to understand the ‘shape’ of a data space, and to remember this shape. Thus, in this study, in order to evaluate which type of landscape is better understood, participants are asked to complete a memory task.

In the modern day, it is commonplace to find user interfaces developed using a 3D perspective. However, it is unknown whether this perspective is useful, or whether it is detrimental to the ease and efficiency of using the software. This research experiment aims to determine whether or not it is easier for users to remember (and thus, absorb) multidimensional datasets when they are graphed in a 3D form, a 2D form, or in the form of a 2D dataset without a surface behind it.

For the purpose of this report, a landscape refers to a computer-generated dataset. Fig. 1 shows an example of each type of landscape used in this study. There are three types of landscapes: a two-dimensional (2D) layout of points with a surface behind it (Fig. 1a), a three-dimensional (3D) layout of points with a surface behind it (Fig. 1b), and a flat layout of points which is a 2D layout that does not have a surface behind it (Fig. 1c). As can be seen, the 2D and flat layout are both shown from a Birdseye viewpoint, whereas the 3D layout is shown from an oblique viewpoint.

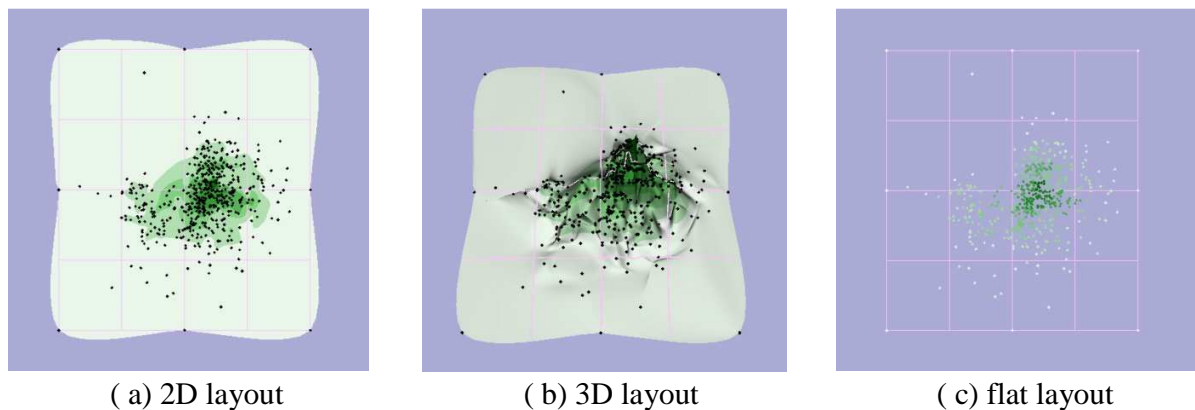


Fig. 1 Landscape displays used in our experiment. All displays show the same data.

1.2 Related Work

As this report is focused on the method and on designing the task, all related work is not mentioned. As is discussed in the conclusion, the basis for this study is *Spatialization Design: Comparing Points and Landscapes* [2].

The idea of using a landscape where the height is derived from the density of points in an area is from the paper: *Visualizing the non-visual: spatial analysis and interaction with information from text documents* [4]. 3D landscapes are referred to as Themescapes, where the themes of the landscape are the areas of points.

2. Experimental Design

2.1 Participants

Of the first 20 participants tested, 12 were males and 8 were females. The average age for those 20 participants was 21. In all, 30 participants were tested for this experiment. All participants for this study were undergraduate students or masters students from various faculties who were living on University of Victoria campus during the summer 2008. Most of the participants were friends with the person conducting the study. Because of their affiliation to the person conducting it, I felt that they were more likely to work harder at task.

2.2 Apparatus

2.2.1 Computer and Software

The study was conducted on an Intel Core 2 Duo desktop processor running at 2.2 GHz with 2 GB RAM, and an nVIDIA GeForce 8800 GTS video card. The computer was running Windows Vista Ultimate 32-bit operating system. Due to a malfunctioning error with the license for DirectRT beta version software, the last ten participants were run using an AMD Athlon 64 bit dual core PC running at 1.99 GHz with 1GB of RAM and a Radeon x1600 series video card. This computer was running Windows XP Professional.

The study was displayed on a Cintiq 21UX Wacom Tablet. The screen size of the tablet was 21.3” on the diagonal, and the screen resolution was 1600 x 1200 (LCD with UXGA resolution). Participants interacted with the software by striking keys on a standard Microsoft keyboard. In the position of the ‘z’ key was a white label with ‘YES’ printed on it, and in the position of the ‘/’ key was a white label with ‘NO’ printed on it.

The study was run using DirectRT v2008 software. [5] DirectRT is software that generates a basic testing interface using a CSV (comma-separated values) excel input file.

2.2.2 Stimuli

There were two types of stimuli that appeared in the test. The first were images for the practice test, and the second were images for the testing. Images of fruit from *Go For 2&5* [6] were used for the learning set. These images were stylized photos of identifiable fruit mounted on a white background. The size of the fruit images ranged from 300 x 159 pixels to 300 x 336 pixels (depending on the image). There were 12 fruit stimuli in total. The second were images of landscapes generated by using the Visualization Toolkit [1] through Java-based software. The

size of the landscape images ranged from having a length of 494 to 500 pixels, and a width of 489 to 501 pixels.

Each participant completed six sets of experimental trials (not including the first practice set). The six sets of experimental trials were: (1) 3D layout with 500 points, (2) 2D layout with 500 points, (3) flat layout with 500 points, (4) 3D layout with 1000 points, (5) 2D layout with 1000 points, and (6) flat layout with 1000 points. Each set presented the participant with eight stimuli in a learning phase. Then it presented them with eight stimuli in the testing phase, four of which were present in the learning phase and four of which were new. Each of the twelve possible stimuli presented in a trial consisted of a landscape generated from a different dataset. However, all 12 stimuli had the same number of points, and were the same type of landscape. There were 12 distinct datasets, and six images for each, yielding a total of 72 generated landscape images.

These landscape stimuli were created from multidimensional datasets which contained approximately 5000 rows, and at least 3 columns. The dataset was laid out in 2D space using MDSteer system for multidimensional scaling [3], and then two copies of the dataset were saved: one containing the 500 lowest stress points, and one containing the 1000 lowest stress points. When generating landscapes using more than 1000 points, the points tended to overlap creating areas of solid black. In addition, these areas would block out the coloured surface behind it and since there were dense areas of points everywhere, and not just in small clusters, the landscapes appeared to be less distinct. When generating landscapes using less than 500 points, the lack of points resulted in fewer hills and valleys being formed. As a result, the landscapes also seemed less distinct.

With the 2D space as an input file, a java class created by Rebecca Dreezer called heightCalculator was used to generate a z coordinate for each (x, y) point coordinate. The landscape for the 2D space is divided into 48 divisions along the x-axis, and 54 divisions along the y-axis, yielding 2592 areas of small intervals. There is a single array (initialized to have all values of 0) where each element corresponds to a different interval. The heightCalculator class reads each (x,y) point from the input file, and determines which of the 2592 intervals it is in. Then, the value of the corresponding array element is incremented. In order to make a smooth contour for the landscape, the average of the point densities in the area around the point was used, instead of just the individual point density. For each point, the average is taken of the number of points in the interval that the point lies in, as well as of the 8 intervals surrounding it. If the interval is one that does not have a full 8 points surrounding it, then the original interval value is maintained. These values are then divided by the maximum number of points found in any interval in order to scale all the numbers between 0 and 1. These new values were manually added into a new column in the input file under the heading of ZCoordinate. The java program was then run with the updated input file, and produced a representation of the landscape where the z coordinate, or height, corresponds to the point density in that area.

The landscapes were generated from 12 real sets of data. They were selected because of the large number of rows and multidimensionality (columns) that they had. The points on the

landscape were chosen to be black in colour to ensure that they would be consistently visible. The points were chosen to be quite large so that they could be easily seen in the 2D and 3D landscapes. This gives participants the ability to use the scatter of the points as a memory aid, in addition to using the shape of the landscape. Also, having larger point sizes makes it easier to distinguish which shade of green they represent.

The landscape has a 4 x 4 grid laid over it. The grid helps participants to visually break up the image so that they can spot the differences among the landscapes' shapes more easily. The horizontal bands across the landscape surface help distinguish in the levels of height (which is equivalent to the point density). The lightest shade starts at a point density of 0, and the darkest shade ends at a point density of 1 (scale from 0 to 1).

As Figure 2 demonstrates, there are seven different bands of colour. Using more than seven bands decreases the variation between the colours and ms with similar values." Initially, the landscapes had five bands, but that made the images seem too simplistic. Different bands of surface colour are used because, as is mentioned in Tory et al. [2], the colours aid participants in "judging the size of a spatial area containing points with similar values."



Fig. 2 Colour of contour bands found on the landscapes

The green was chosen because it was visually aesthetic. When the landscapes are full size, the green pops against the purple background and all seven green shades are easy to distinguish. The green shades were generated using an online colour conversion tool created by Peter Forret [7]. The colour scale chosen was a gradient from a very light green to a dark green. This type of colour scheme was chosen because it helped the participant distinguish between the different height values, without overwhelming the participant by introducing too many fixation points for them to look at. These landscapes were coloured and not done in greyscale because it allowed the black points to stand out. Since the landscape type was being compared, all other variables were as controlled as possible. Thus, all of the landscapes were generated in the same style, using the same colour scheme.

2.3 Study Environment

Each study lasted approximately 25 minutes. The study was conducted in a small office. If it took place during the morning or early afternoon, the blinds were shut tightly, and the lights were turned off. In addition, the door was closed to ensure there were no noise or movement distractions from outside. If the study was conducted in the late afternoon or early evening, the door was left ajar in order to maintain a consistent amount of light. However, this was only done if the office outside was completely empty and silent.

3 Method

3.1 Procedure

After signing a consent form, the participant was taken into the room where the study was held and the DirectRT program was started up. All instructions were presented to the participant on the computer monitor at the start of the study. The conductor was present in the room while the participant completed the practice test in order to be on hand to answer any questions that the participant might have about the instructions. At the end of the study, the participant was asked about his or her strategy for memorizing the images and their impressions of the experiment, but this was not recorded as formal data.

3.2 Input Files

3.2.1 File Organization

There are six landscape test programs (six DirectRT input files), excluding the practice set. Each program is split into six sets and has a distinct and constant order for the sets; in addition, each program begins with a different landscape number of points and type. After the sixth test is run on a participant, the order repeats and the next participant is tested with the first test. The stimuli that are presented to the participant during the learning phase of both the practice test and the testing section are shown in the sequence in which they appear in the stim input file. However, the stimuli that are shown during the testing section are presented in a randomized order that is only determined at run-time. For each test, the first set is always shown in the same sequence, as are the second etc. For example, the first set of images shown in landscape1 will have the same order of images as the first set in landscape2, even though their number of points and height will be different.

3.2.2 File Specifications

All six of the landscape input files are included in a folder with the name landscapeNUMBER in the landscape directory of the DirectRT folder in program files. The CSV excel file landscapeNUMBER is the file that is run through DirectRT as it contains the instructions for the contents. Inside the landscape folder, there is a folder called stim which contains all the text files and image files (must be in BMP form) that are reference in CSV file. The text files contain lists of images (with the file extensions omitted), or paragraphs of texts that are displayed as-is during the execution of the program. The styles (e.g. font, colour) used to display different screens (e.g., instruction screens, or screens with visual stimuli) are controlled by a DRT (DirectRT style) file named landscape.

3.3 Output Files

When an input file is run, DirectRT automatically records two logs. The path DirectRT>landscapeNUMBER>Data>landscapeNUMBER will contain a CSV excel file titled participantNumber and a folder called Log, which contains a second CSV excel file titled participantNumber_log. The file participantNumber is a short data file that records the stim present before the reaction time was captured, the reaction time, and the name of the key pressed

by the subject. The file participantNumber_log is a long data file that is much more detailed. In addition to recording the information from the short data file, it also records the name of the input file, and the data and time when the session was run, etc.

3.4 Running the Program

It is quite simple to run an input file. After launching DirectRT, go to File>Select and Run Input File. Open any folder and select the CSV excel file inside it with the same name as its folder. When the file is chosen, DirectRT will prompt you for a participant number. If this number is not unique compared to the other participant numbers for the CSV file, running the program will overwrite the existing files.

4. Task

4.1 Task Overview

Participants were shown a series of landscapes to memorize. Immediately after, they were shown a second set of landscapes and were asked to identify how many from the second set were also present in the first set. The figures were arranged into groups depending on their layout, whether they were 3D, 2D or flat, and also by the amount of data points, whether there were 500 points (Fig. 3a) or 1000 (Fig. 3b). Their response times, as well as their percentage of correct answers, were being measured.



Fig. 3 Comparison of number of points used. Both landscapes are graphing the same dataset.

4.2 Task Run-through

4.2.1 Introduction

When running the program, the participant is first prompted to enter their age and gender. This is recorded in the log files. The participant was then shown the following overview:

This is a memory test. You will be presented with a set of images, and then you will be required to identify whether images in a second set were present in the first set. The image will be a dataset that is mapped in a 3D graph, a 2D graph, or

is just a set of points. The following three images will demonstrate what it may look like.

The next three slides display the 3D image with 1000 points from a dataset entitled pen (Fig. 1b), the 2D pen image with 1000 points (Fig. 1a), and the flat pen image with 1000 points (Fig. 1c). These landscapes were generated in the same way and using the same style as the landscapes used in the study, however they were not selected to be one of the twelve sets used because their points were not scattered very well, and many points were overlapping one another. All instructions were given on the screen so as to ensure consistency. Participants were welcome to ask any questions they had before beginning the practise test.

4.2.2 Practise

Before the task can begin, the participant performs a practise task in order to get comfortable with the format of the experiment. The times that the images appear on the screen, and the time in between the images are consistent with that of the actual test. In addition, the participants are familiarized with using the format of responding while keeping their fingers over the 'yes' and 'no' keys in order to maintain a consistent, and negligibly short, period of time between choosing an answer, and pressing a button to submit the answer.

The participant is shown the following set of instructions for the learning task:

Before beginning, let us do a practise task. When you proceed to the next slide, you will be shown a sequence of 8 images of fruit. Each fruit will appear for 12 seconds.

They are then shown 8 images of fruit. Each image appears for 12 seconds, and is followed by a black screen for 2.5 seconds. The participant is then prompted to place their hands over the yes ('Z') and no ('/') keys. The next 8 screens each show a piece of fruit (4 that the participant has seen and 4 that the participant has not) and they are asked to respond 'yes' if they have seen it before, and 'no' if they have not.

4.2.3 Testing

After the practise task is complete, the participants are ready to begin the real task. The participant is prompted with the following instructions:

In just a few moments, you will be shown a sequence of 8 images. You are given 12 seconds to look at each image. Try your best to remember them. Press the spacebar when you are ready to begin viewing the images.

Before each set of landscapes, the participant is told to proceed only when ready. This 'proceed' screen between sets to give them a second to rest. In addition, this allows the participant a few moments to clear their mind and focus on forgetting what they saw in the previous set.

As a means of learning, the participant is then shown 8 images of landscapes in the same format as the practise task. Once again, each image remains on the screen for 12 seconds, and is separated with 2.5 seconds of a black screen.

When the learning phase is complete, the participant is prompted to place their hands over the yes and no keys once more. The next 8 screens each show a landscape (4 that the participant has seen and 4 that the participant has not) and they are asked to respond 'yes' if they have seen it before, and 'no' if they have not. The participants were not told that their response times were being recorded. They were simply told to answer the questions as quickly as they could while ensuring that their answers were correct. If the participants had known that they were being timed, they may have gotten flustered and they would be more likely to submit answers they were unsure of in order to answer quicker. There were no breaks given in between answering the questions – as soon as the participant responded, the next image would appear on the screen.

After all six sets of landscapes were shown, the task was complete. An informal interview was conducted with the participant once they had completed the task. The participants were questioned as to their thoughts about the method and the stimuli. The findings from the interviews were not formally recorded.

5. Preliminary Results

At the time of writing this report, the results for all 30 participants were not completely analyzed. However, the general results seemed to be that there was only a very small difference between using a 3D layout for landscapes, and using 2D or flat layouts for landscapes. Participants were negligibly more efficient when remembering 3D landscapes.

The relevant information from the logs was compiled into excel charts, which were exported into SPSS. The results were analyzed statistically using repeated measures analysis of variance (ANOVA) and boxplots.

6. Conclusions and Future Work

There were some limitations to conducting this study using DirectRT. Since the copy of DirectRT for Vista that was used was a beta version, there were some errors when running the program. For example, there was an audio error screen that pops up at the beginning right before a trial is run. However, this may be caused by the fact that there was no external sound device plugged into the computer in use. Also, sometimes when an input file was selected, the program would crash and would need to be reloaded up to three times. While these errors did not really effect the quality of the software, having to debug right before testing was beginning created a less-professional atmosphere than was intended and may have influenced the quality of the user responses.

Many users commented on the increase in the level of difficulty between the practise task (images of fruit) and the real task (images of landscapes). Many users felt that they were caught off guard by the sharp increase in the level of complexity of the images, and that it took a few sets to adjust. Because an equal amount of participants began the real test with each of the six possible sets, the error level of the first sets should technically cancel itself out. Regardless, it is

perhaps better to rerun the study using another set of landscapes for the practise task, in order to give participants an accurate representation of what they should expect in the rest of the sets.

This study was conducted as a follow-up to a previous study entitled *Spatialization Design: Comparing Points and Landscapes* [2]. The previous study was conducted using various types of spatialization designs – three dimensional, two dimensional and flat landscapes. The purpose of this study was to confirm or deny that there is very little difference with regards to remembering datasets between the different types. Since this study was conducted as follow-up, this avenue does not need to be further explored. However, some participants mentioned that they had a hard time trying to recall whether an image they saw was in the current set or the previous one. A future study could entail repeating the same study but changing the colour schemes between each set - so that the participants can easily distinguish between images for each set. Another possibility would be to maintain the same colour scheme, but to vary the colour of the points. It would perhaps be interesting to explore whether it is easier for users to recall datasets that are graphed using different colour schemes. A comparison could be done between greyscale datasets, colour gradient datasets, and rainbow datasets (similar to that of a topographical map).

References

- [1] W. Schroeder, K. Martin, and W. Lorensen. *The Visualization Toolkit*, Prentice Hall PTR, second edition, 1998.
- [2] Melanie Tory, David W. Sprague, Fuqu Wu, Wing Yan So, and Tamara Munzner, "Spatialization Design: Comparing Points and Landscapes", *IEEE Transactions on Visualization and Computer Graphics (Proc. Information Visualization 2007)*, vol. 13, no. 6, Nov./Dec. 2007, pp. 1262-1269.
- [3] M. Williams and T. Munzner. "Steerable, progressive, multidimensional scaling." In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis'04)*, pages 57-64, October 2004.
- [4] J. A. Wise, J.J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. "Visualizing the non-visual: spatial analysis and interaction with information from text documents." In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis'95)*, pages 51-58, October 1995.
- [5] Empirisoft Corporation. DirectRT. <http://www.empirisoft.com/directrt.aspx>.
- [6] Go For 2 & 5. <http://www.gofor2and5.com.au/>.
- [7] Forret, Peter. Color Conversion. <http://web.forret.com/tools/color.asp>.