Investigating the Influence of Display Size on Aspects of Spatial Memory

Master Thesis

by

Vidhu Jain

at the

Universität Konstanz

Department of Computer and Information Science

1. Evaluated by Prof. Dr. Harald Reiterer
2. Evaluated by Prof. Dr. Bela Gipp

Konstanz, 2017

Universität Konstanz

Declaration

Plagiarism is the copying or paraphrasing of other people's work or ideas without full acknowledgement. Rules of good scientific practice generally require that all sources and references have to be indicated by quotes. Intentional plagiarism may incur severe penalties, including failure of your degree.

Vidhu Jain	01/862601
(First name/Last name)	(Student ID)

I hereby declare that the attached term paper for the following class

Master's thesis

(title/lecturer)

on the following topic

Investigating the influence of display size on aspects of spatial memory

(Topic/title of the work)

is the result of my own, independent work. I have not used any aids or sources other than those I have referenced in the document.

For contributions and quotations from the works of other people (whether distributed electronically or in hardcopy), I have identified each of them with a reference to the source or the secondary literature. Failure to do so constitutes plagiarism. I will also submit the term paper electronically to the lecturer. Furthermore, I declare that the above-mentioned work has not been otherwise submitted as a term paper.

I am aware that papers that turn out to be plagiarized will be graded "insufficient" (5,0). Every suspected case of plagiarism will be submitted to the examination board, which will decide on further sanctions. For repeated or otherwise serious fraud or plagiarism cases, the Examination Board may decide to deny the candidate permission to re-sit the examination, resulting in a complete loss of right to examination. The legal basis for this procedure can be found in the examination regulations of the Master's programme "Computer and Information Science" § 8 (7). These rules also apply to all students taking classes in our department.

Date:

____

Signature of the student

Abstract

Spatial memory is an up-to-date research topic within the research domain of human cognition and HCI. Researchers have started to investigate how it can be harnessed in interaction design in order to improve users' performance (navigation, search, recall etc.). In the first part of my thesis, I identify different aspects of spatial memory: *Assembling/Disassembling*, *Rotation, Visual Recall* and *Navigation & Recall* based on an extensive literature review and an Affinity Diagram session with three HCI researchers. Furthermore, I have planned and conducted an experiment in order to investigate the influence of display size on the identified aspects of spatial memory. Comparing a 10.6 inches (SMALL) and 55 inches display (LARGE), participants were asked to perform four tasks (each reflecting one of the four identified aspects) in a controlled within design lab setting. Results of the study show that display size indeed has an influence on the *Rotation* and *Assembling/Disassembling* aspect of spatial memory. For the *Rotation* aspect, Participants performed significantly faster (time) on small display with difficult histogram (6-bar histogram and 270-degree rotation) and significantly more accurate on large display with difficult tasks (6-bar histogram). Participants performed significantly faster (time) when working with the small display in task of *Assembling/Disassembling* aspect of spatial memory.

Table of Contents

List of Figures

List of Tables

1 Introduction

Spatial memory is one of the aspects of cognitive science that supports a multitude of cognitive and behavioural activities like, how to pick up an item ('spatio motor' actions), how to move in an environment and how to recognize the complex figures (Allen 2004). Spatial memory gives us the ability to recall spatial information and to navigate into previously encountered environment. An important aspect of our daily activities is to remember where we left recently used objects. For example, if we have placed a paper on a pile in our work place then for a long time we are able to remember the paper's location. Spatial means occurring in space. When we see something surrounding us, a mental map representation of that space is developed in the human mind (Darken and Peterson 2001). It also relates to our brain to know how the brain stores information regarding the location of any object in the physical space that is available around us (Foreman and Gillett 1997).

Humans have been using spatial relationship and its concepts either directly or indirectly for navigation and communication for thousands of years. It acquired us with the ability to locate the sources of food, water and shelter during the evolution time whereas it aided us to perform spatial tasks beyond the real environment (Patrick, Cosgrove et al. 2000). Many studies have been conducted on spatial memory from the 70ies on (Baddeley, Hitch et al. 1974) until nowadays (Mueller, Rädle et al. 2015) .

Spatial memory is an interesting topic of research in the field of Psychology as well as in the Human Computer Interaction (HCI) domain. Both domains provide their own definitions, area of applications, and work with different measures that they apply in different studies. Researchers analyse spatial memory using a multitude of very different kinds of spatial tasks; for example, Sample Vandenberg task based on spatial orientation (Winner, von Karolyi et al. 2000), Sample card rotation task & Boat test task based on mental rotation in 2D (Winner, von Karolyi et al. 2000), Shepard Metzler & Cube test task based on mental rotation in 3D (Tan, Gergle et al. 2006), Navigation tasks (Spatial learning from direct experience by navigating an environment in real world and in virtual world) (Hegarty, Montello et al. 2006) and many more.

Psychologists have built underlying psychological models of spatial memory (Baddeley, Hitch et al. 1974). Some researchers have shown how spatial memory is responsible for spatial remembering and how it plays an important role for object location in long-term memory model (Baddeley, Kopelman et al. 2003). Baddeley et al. (Baddeley, Hitch et al. 1974) have introduced a working memory model that shows that the visuo-spatial sketchpad (one of the slave component of working memory) is responsible for all the visual and spatial tasks, for example, for visualizing a route to a destination, for supporting mental rotation tasks and many more.

In HCI, spatial memory can be harnessed in order to improve users' interaction with digital information. In the field of graphical user interfaces, the required cognitive and physical efforts for the interaction can be immensely reduced if users are strongly spatial aware about the interface layout and the control locations (Scarr, Cockburn et al. 2013). At the earlier time, HCI researchers were focused on the visuo-spatial metaphors on the screen and to measure how they influence the users (Darken and Peterson 2001). Later, they have performed many studies based on different facets of HCI (interaction & presentation) and by considering different measures to follow in any experiment. For example, some studies focus on the memorization of objects and their re-positioning (Leifert 2011), on the navigation and reconstruction of items (Jetter, Leifert et al. 2012), on the searching and recalling of objects (Mueller, Rädle et al. 2015), and on object location and object identification tasks (Mueller, Rädle et al. 2015). In addition to the different tasks and measurements, these studies have been conducted using different interaction modalities (e.g. interaction techniques like mouse, touch gesture) and output modalities. For example, Leifert et al. (Leifert 2011) have claimed that grids and spatial structure support spatial memory, Klinkhammer et al. (Klinkhammer, Tennie et al. 2013) have investigated the influence of body panning on spatial memory, and Tan and his co-workers show that kinesthetic cues aid spatial memory performance. Currently, they are working on the joint impact of input and output modalities (Zagermann, Pfeil et al. 2017). The common goal of these studies was to find out how to improve the human-computer interaction by designing for spatial memory.

In my thesis, I aim to build on previous studies and investigate further the role of spatial memory in HCI. As discussed above, previous studies only ever focus on a specific definition of spatial memory, making it difficult to generalize the findings across the studies. In my work, I aim to provide a holistic overview over the aspects of spatial memory (Jain 2016) and analyse how the elicited aspects play together. For my project, I analyse these aspects in an experiment (with four tasks) investigating the influence of display size (small: 10.6 inches and large: 55 inches) on spatial memory. By measuring different aspects of spatial memory, results from this experiment can reveal detailed information on how the particular aspects of spatial memory are influenced by the display size.

The complete thesis is divided into 10 chapters. Chapter 2 presents an overview of related work in the field of spatial memory in HCI. Chapter 3 discusses the theoretical background of my work and introduces the research question and research objectives. Chapter 4 focuses on the pilot-study where mainly the experiment with the four tasks was prepared and modified for the final study after overcoming the challenges faced. Chapter 5 shows the implementation of the final four tasks for the experiment. Chapter 6 explains the experimental design (tasks, lab-settings, procedure, participants) of the experiment. Chapter 7 elucidates the analysis procedure of different measures with their related statistical tests in detail. Chapter 8 reports the result of the analysis. Chapter 9 discusses the results and provides a conclusion of the research incorporating suggestions for future work in chapter 10.

2 Related Work

In this chapter, I review related work that covers different facets of this thesis. The sub-chapters focus on different research areas in the field of spatial memory & HCI, namely *Spatial memory in HCI*, *Display size in HCI*, and *the influence of display size on spatial memory*.

2.1 Spatial memory in HCI

Spatial memory is an essential cognitive process that humans use to encode the space by which they are surrounded. HCI researchers have always found spatial memory an interesting topic of research. They are mainly focused on finding different ways of how interaction devices can leverage spatial memory. In this sub-chapter, I am going to summarize the influence of different interaction patterns and visualization techniques (influence of output modalities) that are related to spatial memory in HCI.

2.1.1 Influence of interaction patterns on spatial memory

In this section, I focus on the related research based on different input modalities that can harness and make use of spatial memory to improve user's interaction with the system.

Tan et al. (Tan, Pausch et al. 2002) have performed a direct manipulation task where 30 objects appeared on a 18.1´´screen and 28 participants were instructed to drag those objects on a distinct location on the screen. Mouse interaction and touch interaction were used as input modalities to perform the task. Results show a 19% increment in spatial memory performance with touch interaction.

Building on this work, Jetter and his co-workers (Jetter, Leifert et al. 2012) investigated the impact of panning and zoomable user-interfaces on spatial memory and navigational performance. They conducted two experiments with two types of tasks: a navigation task and a spatial memory/reconstruction task. The goal of the experiments was to know if multi-touch instead of mouse input supports users' spatial memory and improves navigation performance for such UIs. Results show that panning UIs raise the spatial memory performance by 37% as well as navigational performance by 29% for touch input. This finding was not found for zoomable UIs though user preferred to use touch and had 17 % improvement in task completion time.

Klinkhammer et al. (Klinkhammer, Tennie et al. 2013) have investigated the body position as an additional input modality. The main idea of body panning is to use a person's movement as an input modality for the interaction with a system. Klinkhammer and his co-workers have conducted a 'locate & drag' task to measure navigational and spatial memory performance by comparing the impact of touch panning and body panning on a horizontal panning user interface. Results for navigation performance indicate that the body panning condition does not lead to a significantly better performance than touch panning for the navigation task whereas it was shown that body panning results in better spatial memory performance in a spatial memory task.

2.1.2 Influence of visualization techniques on spatial memory

In this section, I present the research that investigates the influence of output modalities or visualization techniques on spatial memory.

Leifert (Leifert 2011) has performed a study with 24 participants to measure the influence of grid lines and spatial arrangements on spatial memory by conducting an experiment in two phases: a memorization phase and a reconstruction phase. Results of the experiment show that grids are useful but spatial arrangement is more important than grid lines for supporting spatial memory. Spatial arrangements and grids have positive influence on recalling locations.

Rädle and his co-workers (Rädle, Jetter et al. 2014) have performed an experiment of peephole navigation to understand the effect of a peephole's size on users' map navigation behavior, navigation performance, and task load. They have used four different sizes of handheld displays as a peephole: tablet, mobile phone, tangible display and a handheld projector. Results show that larger peepholes significantly improve learning speed, navigation speed, and reduce task load.

Similarly, Müller and his co-workers (Muller, Radle et al. 2015) have performed a study investigating the influence of display orientation (horizontal vs. vertical) in dynamic peephole navigation on spatial memory. Participants were asked to perform a *search* and *reconstruction* task on a horizontal and vertical display. Results show that the canvas orientation has no significant effect on navigation performance and spatial memory.

2.2 Display size in HCI

HCI always aims to improve the interaction between users and the system by improving the usability in the form of input and output modalities. The contributed research of input and output in relation to spatial memory is summarized in section 2.1. Display size is also one of the important aspect that can influence user performance with a system.

Some researchers (Grabe, Lombard et al. 1999) have investigated the influence of display size on viewers' responses to media content in the field of media communication. They analysed the influence of display size on the perception of reality and presence, user's preference and level of entertainment, arousal as well as memory and attention. Grabe et al. (Grabe, Lombard et al. 1999) have summarized their finding and conclude that large display induce more intense responses from viewers and the reason for this is that users feel more present in the scene when they watch it on a large display.

Some studies investigated the impact of display size on group work and showed that performance increases for groups working on large displays (Dudfield, Macklin et al. 2001). Simmons and his co-workers (Simmons 2001) have compared the results of some productivity tasks (Microsoft word task, Excel spreadsheet task, Web browsing task, power point task, Multitasking, Web encyclopaedia task) in front of a 21" display (as a large display) and small displays (15", 17", and 19"). Results showed that users performed better on the large display.

Some studies have been performed in virtual environments to investigate the effect of different display sizes (Patrick, Cosgrove et al. 2000) on learned spatial knowledge. An empirical study was conducted to investigate differences in spatial knowledge learned for a virtual environment presented in three viewing conditions: a head-mounted display, a large 3.35 m wide x 2.30 m tall projection screen, and a desk-top monitor with 21´´. Results show no significant differences between the head-mounted display and the projection screen condition and between the head-mounted display and the desktop monitor conditions. A significant difference was however measured between the projection screen and the desktop monitor. Results show that the large projection screen was significantly better than desktop monitor and may be an effective, inexpensive substitute for a head mounted display and might be as effective as head mounted display for educational purposes in the field of spatial cognition.

2.3 Influence of display size on spatial memory

The physical display size is an important factor to consider when designing display systems. Researchers have realized that large displays may afford users a greater sense of presence, which might benefit the performance in certain tasks (Tan, Gergle et al. 2006). For example, Tan et al. (2001) (Tan, Stefanucci et al. 2001) claim that the greater the sense of presence invoked in the user by a large display improves the user's memory for learned information. In previous studies, Tan et al. (Tan, Gergle et al. 2003) have performed a set of experiments by holding visual angles and other display characteristics constant and showed that there was no observable difference in a reading task between a large and a small display. In contrast, results show that users' performance increased by 26% in a spatial orientation task (Guilford-Zimmerman Spatial Orientation Task (Guilford and Zimmerman 1948)) that was performed on a large display as compared to a small display. Similarly, researchers (Tan, Gergle et al. 2004) suggest that physically large displays, even at identical visual angles as small displays, increase performance by 17% on spatial tasks such as 3D navigation. Results from previous research (Tan, Gergle et al. 2006) show that participants perform better in mental map formation and memory tasks (learning & recalling) in virtual environment when they work with large displays because a large screen provides a more immersive feeling than a small screen. These types of studies show that display size matters especially for spatial tasks (like spatial orientation task, mental tasks, recall tasks) and therefore it is a crucial aspect when investigating the spatial memory. That is the reason why I decided to investigate the influence of display sizes on spatial memory in detail.

One of the reasons to support this phenomena is that although a given image has the same theoretical information content on larger and smaller screens and is watched at the same visual angle by the participants, the increment in the size of display surface can fundamentally change the user perception and interaction with the information (Tan 2004). The display size may also induce different cognitive and social reactions that influences task performance.

Carpenter et al. (Carpenter and Proffitt 2001) propose two strategies that might be used to perform spatial orientation tasks: an egocentric strategy (rotating themselves (participants) in the environment), and an exocentric one (rotating the environment around themselves (participants)). Tan and his co-workers (Tan, Gergle et al. 2006) have performed a series of experiments and the results indicate that users chose an exo-centric strategy while working on a small display and when they were working on large display, they chose egocentric strategy. It can be concluded from these studies that the display size also influences the type of strategies (ego-centric or exo-centric) that users take on to perform a spatial task. In my thesis, I focus on spatial memory tasks that are based on exo-centric strategies only.

Tyndiuk et al. (Tyndiuk, Thomas et al. 2005) have been done to investigate the impact of display size on the performance of different kinds of 3D interaction tasks (travel and manipulation). There was a little modification on the study of Tan and his co-workers (Tan, Gergle et al. 2006) for physical large display size in terms of types of tasks and dimension of display sizes. They (Tyndiuk, Thomas et al. 2005) showed that not all users benefit similarly from the use of large displays. The impact of the large display depends both on the nature of 3D interaction task, for example egocentric and exocentric task, and on some parameters of the users' cognitive profile like the selective visual attention ability, and the ability to select visual information. Users who have low visual selective attention abilities benefit more from large displays than others. In other words, large displays can be considered as a cognitive help for them. Tyndiuk et al. (Tyndiuk, Thomas et al. 2005) have contradicting result with Tan et al. (Tan, Gergle et al. 2006). Tan et al. (Tan, Gergle et al. 2006) summarize that large displays lead to better feeling of presence and have a positive impact on interactive exocentric task performance. In my experiment, I include one task where participants have to follow the visual patterns on the screen. In that sense, it will be interesting to investigate the finding on large and small display to see whether the results support to outcomes of previous studies or generate some new results.

Zagermann et al. (Zagermann, Pfeil et al. 2017) have investigated the combined influence of input modalities and display size on spatial memory, efficiency of task completion and user satisfaction. They performed an experiment consisting of a navigation phase and an object location memory phase in a task with 28 participants. They have chosen three input modalities; trackpad, direct touch & gesture based motion controller and two display size; 10.6 ´´Microsoft surface 2 pro & 55´´ Microsoft perspective pixel. Their findings show that participants perform significantly better with TOUCH and MOVE input modalities in front of large display while measuring navigation performance. Navigation speed increased for TOUCH and MOVE on the small display. In terms of user satisfaction, there was no impact of display size but PAD and TOUCH were rated statistically significant more novel on the large display. TOUCH and PAD are rated equal with respect to subjective task load and user experience. Further, their combined work of input and output modalities provide the motivation for some further research in the field of display size and influence of identified aspects (Jain 2016) of spatial memory.

This chapter was all about, how the spatial memory harness the HCI settings in multiple directions. In the next chapter, I focus on the preliminary work that is also a part of my seminar work (Jain 2016) and the motivation of my research.

3 Preliminary Work

Spatial memory is a vivid research domain among HCI researchers and Psychologists. Depending on their field of interest, researchers have analysed spatial memory from different perspective, provided their own definitions and applications. Psychologists claim that spatial memory is essential for way finding and in navigational tasks (Hegarty, Montello et al. 2006) whereas HCI researchers focus more on the support of navigation and reconstruction tasks (Jetter, Leifert et al. 2012). As already discussed in chapter 2 that, according to their research interest, they focus to different aspects. Up to now, most studies have only ever focused on one or two aspects of spatial memory. For example, Muller et al. (Muller, Radle et al. 2015) has investigated the influence of canvas orientation on spatial memory and Tan et al. (Tan, Pausch et al. 2002) has investigated the influence of kinaesthetic cues on the spatial memory performance.

There has been a lot of research in the area of spatial memory up to now but some concrete issues are still awaiting investigation as mentioned in chapter 2. For example, there is no clear consent about the definition of spatial memory, its aspects, its measures and many more. In my seminar work (Jain 2016), I address this problem and tackle it by conducting a large literature review about spatial memory. I have investigated literature from Psychology and HCI and assembled the work done in the field of spatial abilities. In total, I have gone through 11 research articles and 5 online tests, resulting 38 spatial tasks at the end. These all the tasks address one or more aspects of spatial memory. Some tasks are performed in real-world settings and some are performed in a virtual environment or in mixed reality. In addition, some tasks can be solved by touching them (items) or navigating them physically and others are solved based on good imagination power and sharp memory.

After having a collection of huge varieties of visual and spatial tasks, I conducted an Affinity Diagram (Holtzblatt and Beyer 1993) session with three HCI researchers in order to arrange and categorize the identified tasks. Based on this Affinity Diagram session, I formulized aspects of spatial memory. The result of the Affinity Diagram session is shown in Figure 1 where the concluded aspects and sub-aspects with their respective figures are specified.

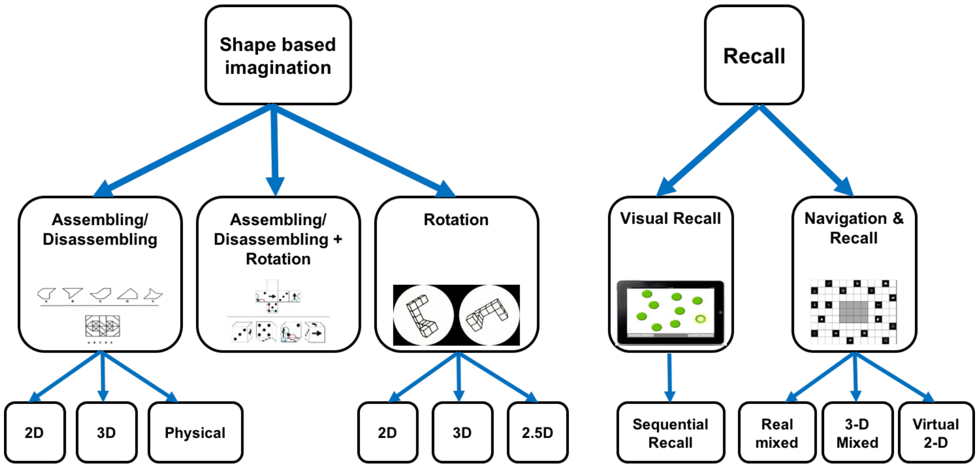


Figure 1:Result of Affinity Diagram in the form different aspects of spatial memory.

Figure 1 shows that there are mainly two aspects of spatial memory; *Shape based imagination* and *Recall.* These two aspects of spatial memory also have their respective sub-aspects and their sub-categories. The brief description of both the aspects are mentioned here.

Shape based imagination

In the *shape based imagination* category, tasks for which task completion is based on the imagination of shapes are included. For example figuring out simple shapes in a big complex shape (Winner, von Karolyi et al. 2000), pointing out overlapped shapes into separate shapes[1] or disassembling the bigger block into smaller sub-blocks[2] are included in this category. This category is divided into the following three sub-categories:

a. *Assembling/Disassembling*
All tasks belonging to this sub-category involve assembling multiple pieces into one piece or breaking down one shape into multiple shapes, for example, solving a Pyramid Puzzle (Winner, von Karolyi et al. 2000) . This sub-category of spatial memory is further divided into three sub-categories depending on the presentation of the shapes, either digital 2-dimention or 3-dimention or physical.

b. *Assembling/Disassembling + Rotation*
All tasks that involve assembling or disassembling shapes by rotating them mentally are summarized in this category, for example, the task Sample Card Rotation (Winner, von Karolyi et al. 2000). This sub-category of spatial memory has no further sub-categories.

[1] http://www.cambridgebrainsciences.com/browse/planning/test/spatial-self-ordered-search

[2] http://www.queendom.com/queendom_tests/transfer

c. *Rotation*
The tasks that involve identifying the angle of orientation and the direction of movement are kept together in this category, for example, the Shepard Metzler task (Shepard and Metzler 1988). This sub-aspect is divided into three further sub-categories based on presentation of shapes either in 2-dimension, in 3-dimension or in 2.5-dimension (the dimension that neither belongs to exact 2-D nor 3-D)

Recall

The second category formulized based on the Affinity Diagram session is *Recall*. In this category, mainly all the types of navigational and way finding tasks are included. The main goal of these types of task is to first become familiar with the environment and then to recall whatever has been learnt earlier either in physical or in virtual environment. ‘Update location while blindfolded’ (Hegarty, Montello et al. 2006) is one of the examples of the task that are sorted in this category. This category is further divided into following sub-categories:

a. *Visual Recall*
Tasks that are based on ‘recall because you have seen it beforehand’ are categorized here, for example, Memorization and Reconstruction Object Cards (Leifert 2011) . Some tasks also included recalling information but in a sequential way (e.g. recalling the order in which blocks flashed (Brunetti, Del Gatto et al. 2014)) and these tasks are summarized in the sub-category named *Sequential Recall*.

b. *Navigation & Recall*
In this category, all tasks are included in which recall is based on ‘navigation beforehand’, for example, Navigation and Reproduction task (Jetter, Leifert et al. 2012). This sub-category is further categorized into three sub-categories depending on the environment in which the tasks are performed, like the real world, a virtual environment, or in mixed-reality settings.

As a result of the Affinity diagram, I concluded with two categories and five sub-categories describing aspects of spatial memory. As the second sub-category *Assembling/Disassembling + Rotation* is reflecting the other two sub-categories *Assembling/Disassembling* and *Rotation*, I am not considering the *Assembling/Disassembling + Rotation* as a separate aspect. So finally, I concluded to sort spatial memory tasks according to the following four aspects; *Assembling/Disassembling*, *Rotation*, *Visual Recall* and *Navigation & Recall*.

3.1 Research Question and Objectives

After eliciting the different aspects of spatial memory based on extensive research and Affinity Diagram session, the aim of my thesis is to find out how these aspects of spatial memory can be harnessed in an HCI setting.

One of the main goals of HCI is to improve the interaction between user and the system. There is huge research to make this interaction smooth by developing user interfaces that allow for an easy flow of information from the system to the user and from the user to the system. The flow of information from user to system is known as interaction whereas the flow of information from system to the user is known as presentation (Tripathi 2011). Basically, this interaction from the user to the system and the presentation from system to the user are the two facets of HCI.

In my thesis, I am interested to analyze the influence of these two directions of HCI on the different aspects of spatial memory. For doing so, I have gone through many studies based on user's interaction with systems and presentation of information from system to the user and came up with different ideas: The first idea was to investigate the influence of input modality (mouse input and touch input) on the identified aspects of spatial memory. Jetter et al. (Jetter, Leifert et al. 2012) has performed experiments to investigate the influence of input modalities on spatial memory performance and their results show that touch input is superior to mouse input for panning interfaces. Similarly, I considered to investigate the influence of touch and mouse input on the classified aspects of spatial memory. This idea of analyzing different input modalities is suitable for the aspects *Assembling/Disassembling*, *Navigation & Recall* and *Visual Recall* but does not fit well with the *Rotation* aspect where people have to decide the order of orientation and direction mentally without requiring interaction with the system. Thus, I discarded this idea because I aimed to select a factor that can be investigated with all identified aspects. So, at last, I dropped the idea of input modalities that supports the interaction facet of HCI.

Next, I considered the other facet of HCI: presentation. To support the idea of presentation, I considered to investigate the influence of output modalities on the identified aspects of spatial memory, for example display size. Previous research has identified the influence of display size on the spatial memory already (Tan, Gergle et al. 2006) (Tyndiuk, Thomas et al. 2005) (Tan, Gergle et al. 2003). They have also found out different characteristics of the tasks that leads to different outcomes. Their findings claim that user performed better on large display while solving spatial orientation tasks. They like to choose ego-centric strategies when they perform on large display and exo-centric strategies while performing on small display (Tan, Gergle et al. 2006). These finding provides a sound basis to investigate the influence of display size on different aspects of spatial memory. Display size is a factor that can be implemented in all the identified aspects and it is interesting to see how the display size impacts onh the identified aspects of spatial memory and how results of this study match results of previous research (Tan, Gergle et al. 2006). Therefore, I decided to work with display size and finalized to investigate the influence of display size on the concluded aspects of spatial memory. Finally, I have come up with following mentioned research question and I am addressing this overall research question by investigating the following research objectives.

Research Question

Does display size influence spatial memory?

Research Objectives

Does display size influence the *Visual Recall* aspect of spatial memory?

Does display size influence the *Rotation* aspect of spatial memory?

Does display size influence the *Navigation & Recall* aspect of spatial memory?

Does display size influence the *Assembling/Disassembling* aspect of spatial memory?

I aim to address the four research objectives separately, as each research objective reflects one of the aspects. The combination of the results for the separate research objective will then provide the basis to answer the overall research question.

I have planned an experiment with four tasks based on these four research objectives. In the next chapter, I have summarized the conduction of the pilot study with the planned experiment with a few participants to test whether the lab set-up, the display sizes, the task and the analysis worked as planned.

4 Pilot Study

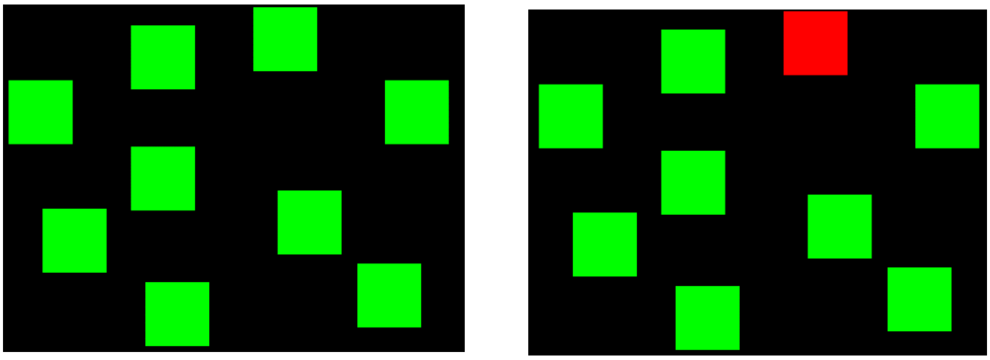
Before going ahead with the experiment, I conducted a pilot study (Jain 2017) with three participants in one of the labs of the HCI group at the University of Konstanz. In this pilot study, I performed the whole experiment in order to test whether the lab set-up, the selected display sizes, the implemented tasks, the procedure of the experiment and the analysis of the collected data worked as planned. The pilot study allowed me to test technical issues related to the selected input modalities and the display size. In addition, I could investigate how participants' experience the experiment, and whether they understand the tasks. Finally, I could experience the whole session myself, and test in a preliminary analysis whether the collected data could be processed and analyzed as planned. Based on my experiences from the pilot study, I could identify parts of the experiment that should be modified for the final experiment. The following chapters provide a brief overview over the pilot study that includes the information about tasks and technical setup used in the experiment, the results of the preliminary analysis, demographic information about participants and the procedure followed in the whole session of pilot study. A detail description can be found in (Jain 2017).

4.1 Tasks

I have planned an experiment with four tasks that are reflecting the identified four aspects of spatial memory. In this section, I explain the tasks used in the pilot study separately, for further information regarding the implementation of these tasks, please refer to (Jain 2017).

4.1.1 Spatial Span task

The *Spatial Span task* was included in the pilot study to reflect the *Recall* aspect of spatial memory. A visual arrangement of green 9 non-highlighted blocks was shown on the screen. Then, a few of them flashed (in red) in a pre-defined order. The participant's task was to remember that sequence of flashing blocks and to click on these in the same order (using the mouse). The non-highlighted 9 blocks and one highlighted red block among non-highlighted block is shown in figure 2 (a) and figure 2 (b) respectively.



(a) (b)

Figure 2: (a) nine blocks (in green colour) on the screen in non-highlighted state. (b) One flashing red block among non-highlighted blocks.

The task had eight different difficulty levels, ranging from level B2 to level B9. In each level, the number of flashing blocks gradually increased. For example, at level B2 two blocks flashed, at level B3 three blocks flashed, increasing up to level B9 where all the nine blocks flashed in a pre-defined sequence. In each level, there were two attempts that participants were asked to respond to. If participants at least managed to correctly perform one of these two attempts, he/she proceeded to the next level. In case, participants were not able to correctly respond to at least one of the two attempts, (s)he did not proceed to the next level and the task was terminated.

The task was programmed in Inquist (Inquist 2015) using JavaScript and html. The implementation process is explained in detail in my project report (Jain 2017).

4.1.2 Histogram Rotation task

The *Histogram Rotation task* was implemented in the experiment to reflect the *Rotation* aspect of spatial memory. In the histogram rotation task, participants were asked to decide between congruency and in-congruency after watching the 'target histogram' (in vertical upright position) and the 'comparison histogram' (rotated by either 90-degree, 180-degree or by 270-degree) – one after each other. At first, a target histogram appeared on the display for three seconds, then a blank screen was shown for one second before another comparison histogram was shown (rotated) on the display (see Figure 3). Participants were then asked to decide whether the two shown histograms were congruent or incongruent by pressing two specified keys on the keyboard.

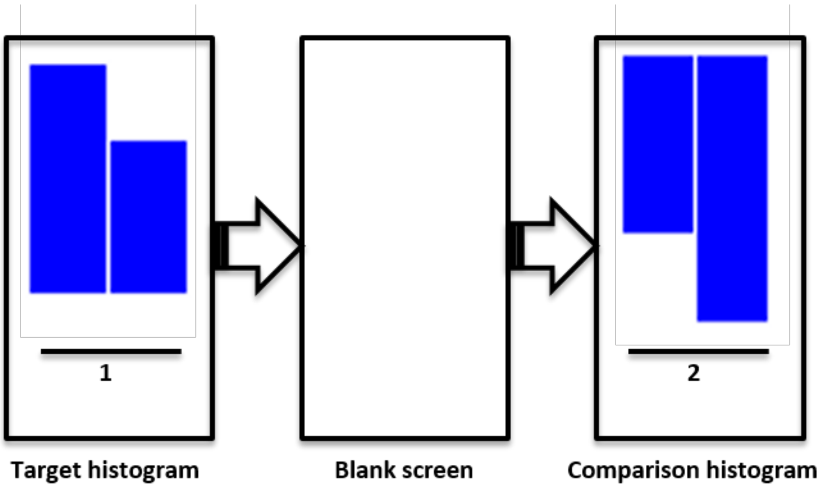


Figure 3: The sequence of screens in the Histogram rotation task.

The histograms were shown in blue colored bars on white background. There were three types of histograms: 2-bar, 4-bar and 6-bar histograms. The task was planned for 6 minutes i.e. 2 minutes for each type of bars. The task started with showing 2-bar histograms, followed by 4-bar histograms and 6-bar histograms. The rotation of the 'comparison histogram' was counterbalanced (it may be rotated by either 90-degree, 180-degree or by 270-degree).

Figure 4, figure 5 and figure 6 show how the 2-bar, 4-bar and 6-bar histograms will look like in the three types of rotation respectively where (a) is showing the vertical upright position, (b) is showing the rotation by 90-degree, (c) is showing the rotation by 180-degree and (d) is showing the rotation by 270-degree in all the three figures.

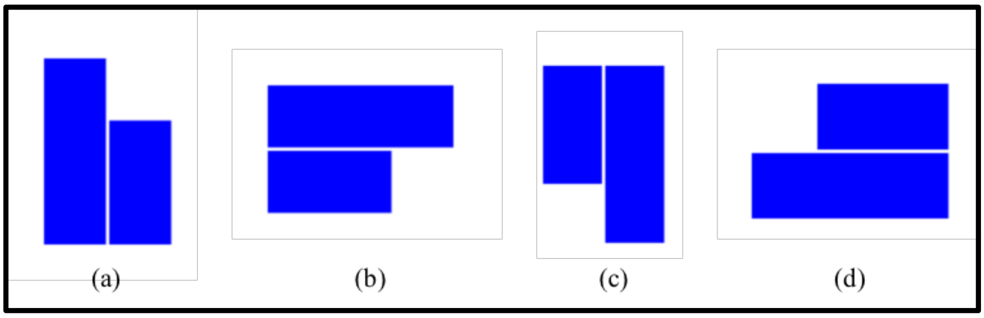


Figure 4: 2-bar histograms and its rotation in three direction.

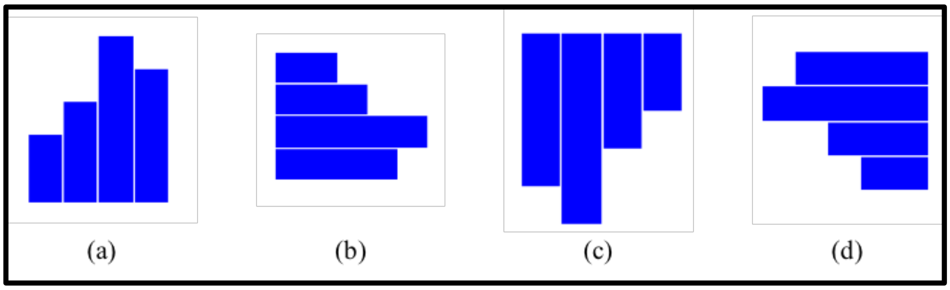


Figure 5: 4-bar histogram and its rotation in three direction.

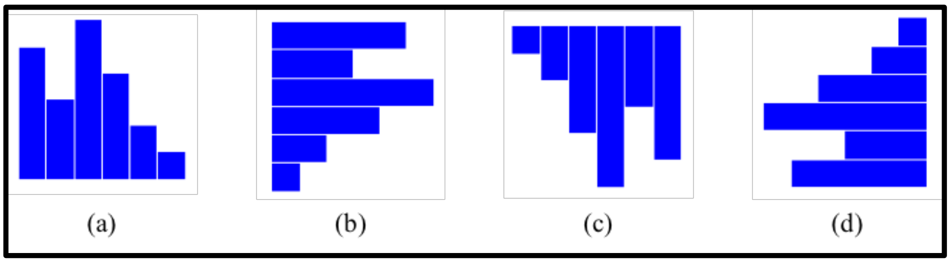


Figure 6: 6-bar histograms and its rotation in three direction

The task is programmed in Inquist (Inquist 2015) using JavaScript and html. The implementation process is explained in detail in my project report (Jain 2017).

4.1.3 Navigation task

The *Navigation task* was included in the experiment to reflect the *Navigation & Recall* aspect of spatial memory. I have chosen a task as a navigation task for my experiment that is well-established and used in other studies (Jetter, Leifert et al. 2012). In this task, 18 different items were spatially distributed over a canvas whose centre was empty at the beginning and was serving as home position (see Figure 7). When the task was started, an item appeared at the centre of the screen and participants were asked to navigate to this item on the canvas by panning the landscape with a mouse and overlap the item to be searched with the centre of the screen with a tolerance of 100 pixels.

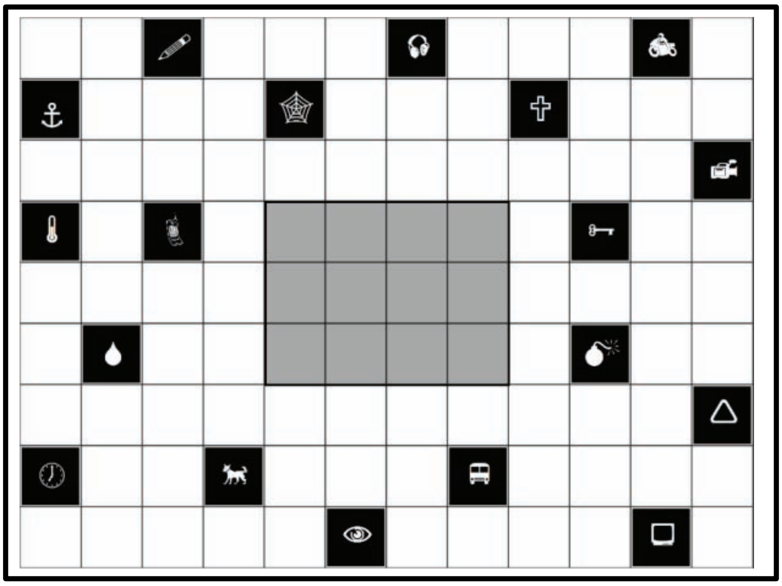
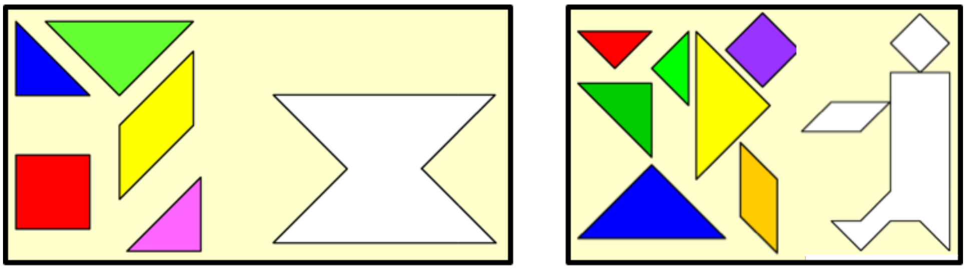


Figure 7:The Navigation task: A canvas with 18 items scattered over it and with an empty center position

4.1.4 Tangram task

The *Tangram task* was included in the experiment in order to reflect the *Assembling/Disassembling* aspect of spatial memory. The Tangram task[3] required participants to assemble few multicolour geometrical shapes into a target image. All geometrical shapes were placed at the left side of the screen and the target image was placed at the right side of screen. The task consisted of four trails. In the first two trails, participants were instructed to assemble 5 shapes into target image whereas in the second trail, they were instructed to assemble 7 shapes (see Figure 8 (a) and 8 (b)). These two different versions represent two different difficulty levels of the task.

[3] http://www.tangramgames.co.uk/



(a) (b)

Figure 8: Tangram task where (a) is showing the difficulty level one with 5 shapes to assemble and (b) is showing the difficulty level two with 7 shapes to assemble in target image.

For the user interaction, the mouse was chosen to drag and move the shapes onto the target image and two rotation symbols at the bottom left of the screen to rotate the shapes by 90-degree, 180-degree and 270-degree. A complete picture of the screen with the rotation symbols is presented in figure 9.

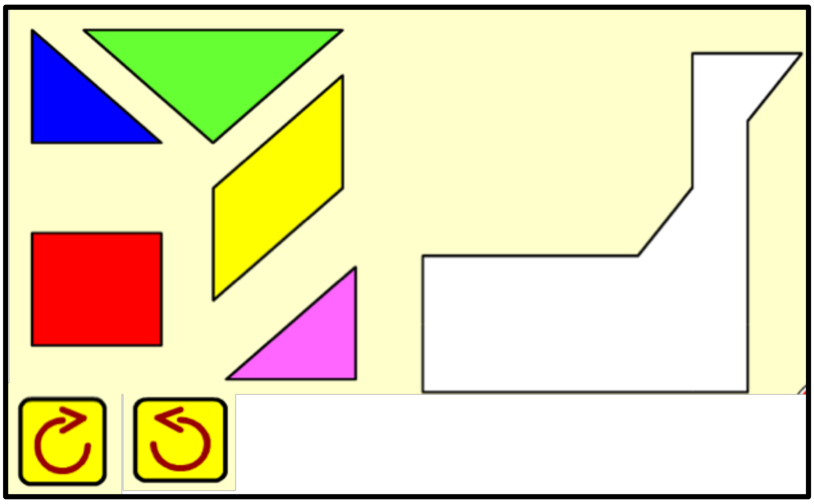


Figure 9: Five geometrical shapes and two symbols for rotation at the left side of the screen and one empty target shape at the right side of the screen.

4.2 Technical setup

The selection of display size (the Perspective pixel 55” display as LARGE and the Microsoft surface 2 pro 10.6” display as SMALL) and the lab setup did not change after the pilot study, and therefore are identical in the final experiment (see section 6.1). Therefore, I only provide a sketch of the setup here in this section (in figure 10). The rationale behind the choice of display size and the experimental setup is described in detail in section 6.1.

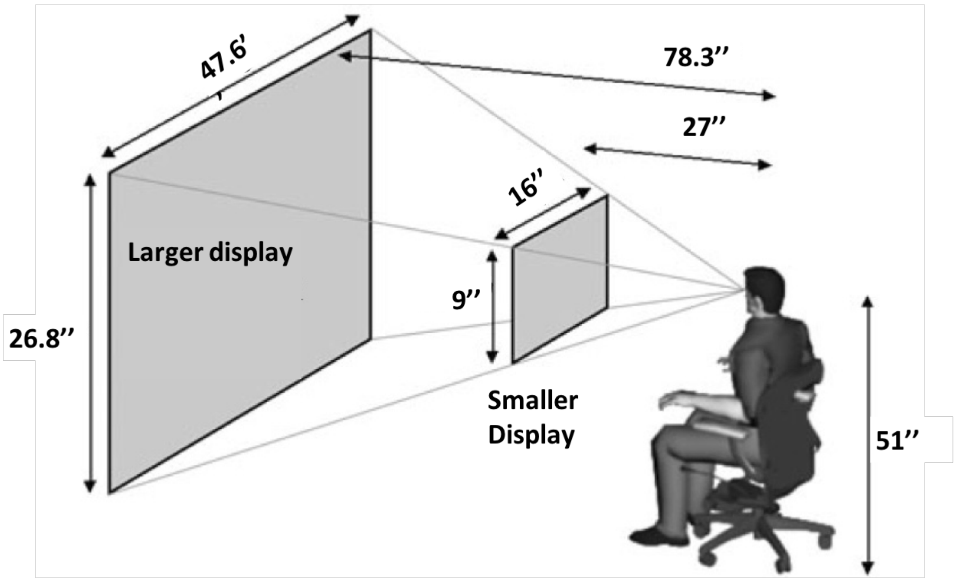


Figure 10: A sketch to represent the total settings for maintaining the visual angles between two display sizes.

4.3 Participants

Pilot-study was conducted with 3 participants (2 females and 1 male). The cultural background of participants was diverse. One participant belonged to Azerbaijan, one was German and one participant was Indian. The mean age of all the participants was 24.3 years. All the three participants belonged to computer science department and were Master's students. They have mentioned their expertise in the use of mouse and in keyboard (in *Demographic questionnaire*). No one of them has any vision problem.

4.4 Procedure

The whole session of pilot-study was conducted in one of the interaction lab of HCI group in the University of Konstanz. All the three participants were invited at the same day at different times. The whole session was planned for the 65 minutes. There was no money as reward for their participation in the pilot-study.

The whole procedure of the pilot study is shown in figure 11. Each session was started with a welcome session and followed by filling in the *Demographic-questionnaire* and the *Letter of*

Consent. Firstly, one type of task was performed in front of one type of display size and later the same task was performed in front of other type of display size. In the whole session of pilot-study, I have switched the both display size total four times. The selection of the tasks and the selection of the display size was counterbalanced (Jain 2017). The sequence of display size and tasks were repeated a total four times in a counterbalanced manner.

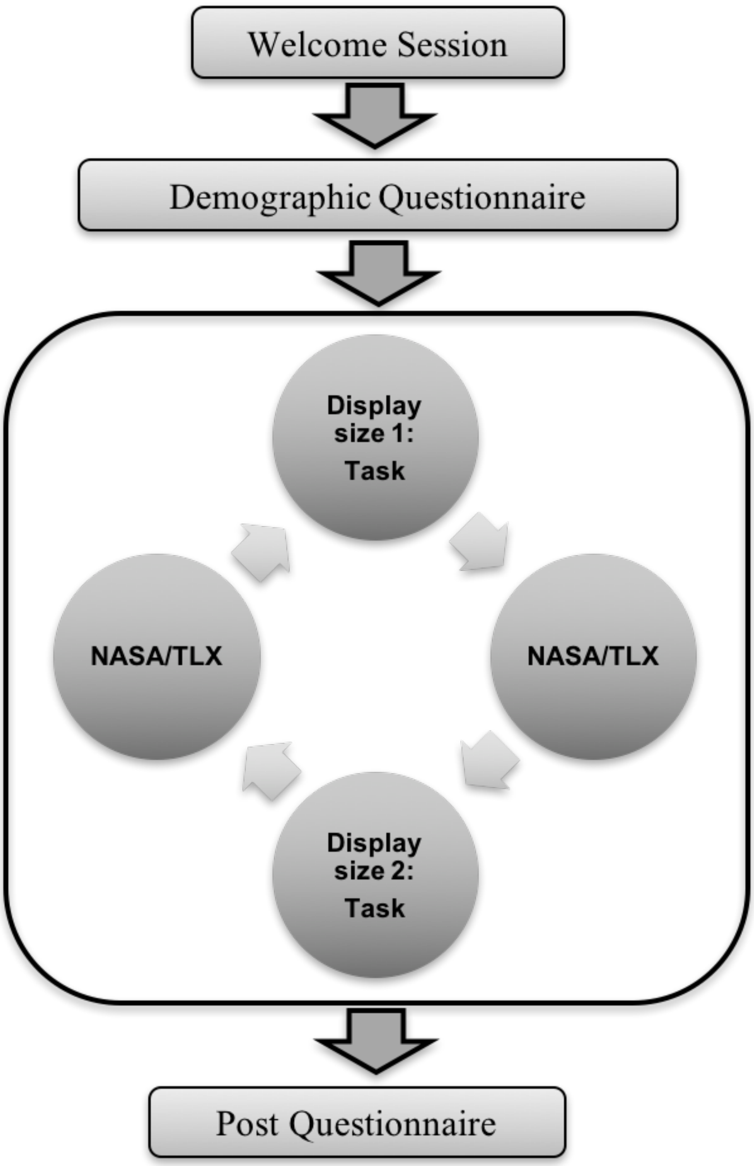


Figure 11: Complete picture of pilot study that is showing all the steps followed during the whole session.

The technical set-up was same for the study as mentioned above (in section 4.2). For each of the tasks, I prepared a demonstration to show to the participants. Before the beginning of each task, I gave this short demo of that particular task so that participant was able to become familiar with the task easily and could perform better. I was also instructing them about the input modalities before each task, like which input device they were asked to use for performing a particular task.

After the demo presentation, participants were asked to work on the task in front of one type of display and the data collection was started. At the end of one task, participant was asked to fill out *NASA/TLX* (Hart 2006). In the meanwhile, I was arranging the second type of display size. Now the participant was ready to perform again the same task in front of second type of display. Again, after the task completion, participant was asked to fill out the *NASA/TLX*.

After the completion of all the four tasks in front of both the display sizes, the participant was asked to fill in the *Post-questionnaire* and it was the end of the experiment.

4.5 Preliminary analysis

The *Spatial Span task* has some challenges as participants quit the task at different stages which posed a problem when analysing the data ('Problem of abandon'). In addition, the choice of colour (red and green on black background) was discovered to be problematic for colour blind participants. To overcome these challenges, some changes have been implemented for the final version of this task.

In the *Histogram Rotation task*, mainly three challenges were faced: the duration of the task was too long and tiring for participants, the colour of histograms might have influenced participants' performance (blue colour bars), and the type of the histograms (2-bar, 4-bar, and 6-bar) was not counterbalanced leading to a learning effect. These challenges were addressed in the final task.

Concerning the *Navigation task*, it was identified that the two item sets were not counterbalanced across the two display sizes. This was addressed in the final version of the task.

In the *Tangram task*, mainly challenges faced were: influence of colours (multicolour shapes might have biased results), long duration of the task and the counterbalancing between the two target images.

All four tasks with their challenges and the required changes are summarized in Table 1. All the changes made regarding the tasks as well as the set-up are described in detail in chapter 5.

SN	Tasks	Challenges	Changes in task	Changes in set-up
1	Spatial span	• Problem of abandon • Color blindness	• Participant can perform till end of last block • Monochromatic theme	
2	Histogram rotation	• 6 minutes' time is longer • Color blindness • Performance evaluation	• 3.3 minutes total time • Monochromatic theme	Counterbalancing with types of bars
3	Navigation task	-	-	Counterbalancing with two item sets
4	Tangram	• Concluded strategies • Color blindness • Time	• Only 7 shapes to assemble • Monochromatic theme	Counterbalancing

Table 1: Collection of all the challenges faced in all the four tasks with their respective changes regarding the task and the set-up.

Chapter 4 was all about the conduction of pilot-study. In this pilot study, I performed the whole experiment in order to test whether the lab set-up, the selected display sizes, the implemented tasks, the procedure of the experiment and the analysis of the collected data worked as planned. Based on my experiences from the pilot study, I could identify parts of the experiment that should be modified for the final experiment, that is mentioned in the next chapter 5.

5 Task Implementation

In chapter 4, I have discussed the challenges that I have faced during the planning, conduction and analysis of the tasks in the pilot study (Jain 2017) with three participants. To address these challenges, I have provided suggestions to overcome them and changes for a re-implementation of the tasks, so that I could conduct the final experiment with a higher number of participants in my thesis.

In this chapter, I describe in detail how I re-implemented the four tasks for my experiment. I divide this chapter into four sections corresponding to the four tasks. Each section is again divided into two sub-sections: at the beginning of each section, I provide a brief introduction into the old version of the task (as used in the pilot-study) and the challenges faced. In the second part, I describe the new version of the task after addressing the challenges.

5.1 Spatial Span task

The task is reflecting the *Visual Recall* aspect of spatial memory. Participants are shown visual patterns on the screen (different blocks flashing one after another). Participants are expected to learn the pattern first and then recall it by clicking on the blocks in the order that they previously flashed. In the following, I provide a brief introduction about the old task, followed by the challenges faced and solutions applied. Lastly, I describe the final version of the task.

5.1.1 Old version of Spatial Span task

Nine green colored blocks were appearing on the black background over the screen and some of them flashed in red in a pre-defined sequence (see figure 2 in section 4.1.1). Participants were asked to remember the sequence of the flashes and click on the blocks in the same order.

There are eight different difficulty levels within the task, ranging from B2 up to B9. At the first level (B2), only two blocks flash and at the last level (B9), all the nine blocks flash. Each level includes two attempts. In level 1, two blocks flash in the first attempt and in the second attempt, again two different blocks flash.

In the original version, moving to next level of the task depended on the performance in the previous level. If participants completed both the attempts within a level successfully, then they proceeded to the next difficulty level. If participants respond wrongly in both attempts of a level, then they were not able to move to the next level and the task was terminated. There was no time limit to respond in the task.

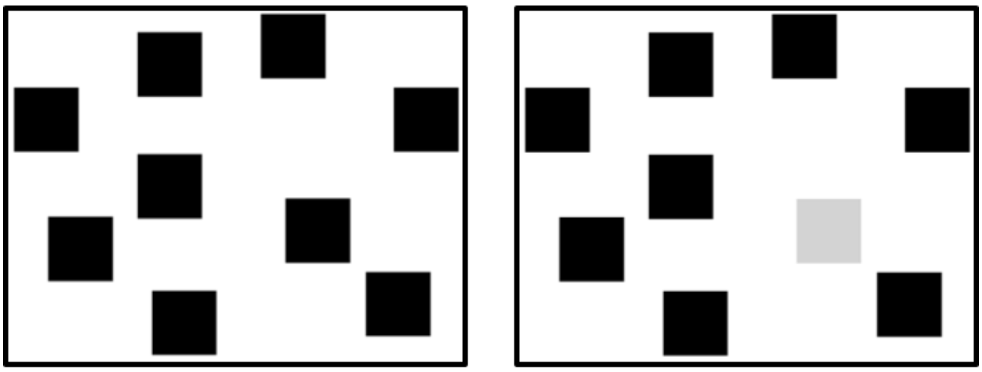
5.1.2 Challenges faced

The problem faced during the analysis was the **'**problem of abandonment'. The reason for this problem was that participants dropped out of the task at different stages. Some participants performed well and therefore faced more levels and dropped out later in the task. Others, that did not perform well dropped out at the beginning of the task. The problem that participants dropped out at different stages of the task led to problems in the analysis of the time that participants needed to perform the task (due to missing values for higher levels). One more challenge that I faced was concerning the choice of colour for this task, as the choice of green and red was very sub-optimal for colour-blind participants.

5.1.3 Final Spatial Span task

I aimed to remove the 'problem of abandonment' and to consider the factor of colour blindness while preparing the final *Spatial Span task* for the experiment. In the newer version of the task, the concept of the automatic drop-out after two wrong attempts in a level was discarded. Participants are confronted with all possible attempts until B9 irrespectively of the number of wrong responses. Hence, there are no earlier dropouts at different levels in the task and all participants have to perform all difficulty levels (B2-B9). Subsequently, counting the number of right and wrong responses throughout the task was more meaningful that resulting the analysis of 'time' in a straightforward and correct way.

In order to overcome the problem of colour blindness, the design of the task was based on a monochromatic theme, resulting in a black-white version of the task. In the new version of the task, all the nine blocks appear in black colour over white colour background and flash in grey colour (as shown in figure 12 (a) & (b)).



(a) (b)

Figure 12: The new version of the spatial span task with monochromatic theme where in (a) all nine blocks are in non-highlighted state and in (b) one block is flashing in grey colour among all nine blocks.

5.2 Histogram Rotation task

The *Histogram rotation task* reflects the *Rotation* aspect of spatial memory. In this task, participants are instructed to decide whether two histograms – shown separately and rotated – are congruent or incongruent. The following sections provide a brief description of the old version of this task used in the pilot-study, the challenges faced in that and the new variant of the task.

5.2.1 Old version of the Histogram Rotation task

The *Histogram Rotation task* (Jain 2017) is the task where participants are asked to decide whether two histograms are congruent or incongruent. At the beginning of the task, a first histogram (target histogram) is shown on the display in a vertical up-right position for a few seconds, then another histogram (comparison histogram) is shown in an rotated way (see figure 3 in section 4.1.2). After the appearance of the second histogram, participants have to decide whether these two histograms are congruent or incongruent by pressing two specified keys on the keyboard.

There are three types of histograms in the task; histograms consisting of two bars (2-bar), histograms consisting of four bars (4-bar) and histograms consisting of six bars (6-bar). In addition, the comparison histograms can be rotated in three ways: rotation by 90-degree, 180-degree and 270-degree. Originally, the whole task was planned to last a total of 6 minutes. In the first two minutes, 2-bar histograms were shown on the screen, for the next two minutes, 4-bar histograms were shown followed by another two minutes in which 6-bar histograms were shown [Jain 2017]. The order of the orientation was counterbalanced.

5.2.2 Challenges faced

After analyzing the collected data of the pilot study, I faced the following challenges in the task as well as in its set-up:

Challenges faced in task

Working on the rotation task for a total of 6 minutes was sometimes perceived to be frustrating by participants and some of them lost motivation to complete the task until the end. Thus, I decided to reduce the duration of the task. In addition, I also experienced that the colorful design of the histograms might lead to biased results due to color blindness or preference.

Challenges faced in set-up

The main problem faced in the experimental set-up was the fact that the different types of histograms (2-bar, 4-bar and 6-bar) were not counterbalanced leading to a learning effect which resulted in a very low performance in the 2-bar condition. However, as this condition was always the first condition that people worked with during the task, this low performance was not due to the fact that histograms consisted of 2 bars, but due to the fact that these histograms were always the first histograms that participants worked with. In order to overcome this problem, the sequence of the three types of histograms is counterbalanced in the final experimental design.

5.2.3 Final Histogram Rotation task

In the final version of the task, a monochromatic theme is applied. Now the target histogram and comparison histogram are shown in black color on a white background (see Figure 13). The number of the trials is reduced in order to fix the final duration of the task to 3.15 minutes. Now in the newer version of the task, each type of histogram will appear total 8 times on the screen, 1.05 minutes for each type of histogram with counter-balanced order of rotation. The order of the types of histograms is counterbalanced (starting either with a 2-bar, 4-bar or 6-bar histogram followed by the remaining two types of histograms) as well as the order of rotation.

The sequences of histograms are mentioned in the figure 13. Target histogram is appearing for 3 seconds on the screen that is followed by the appearance of blank screen for 1 second and at last comparison histogram is appearing on the screen (till participants' response).

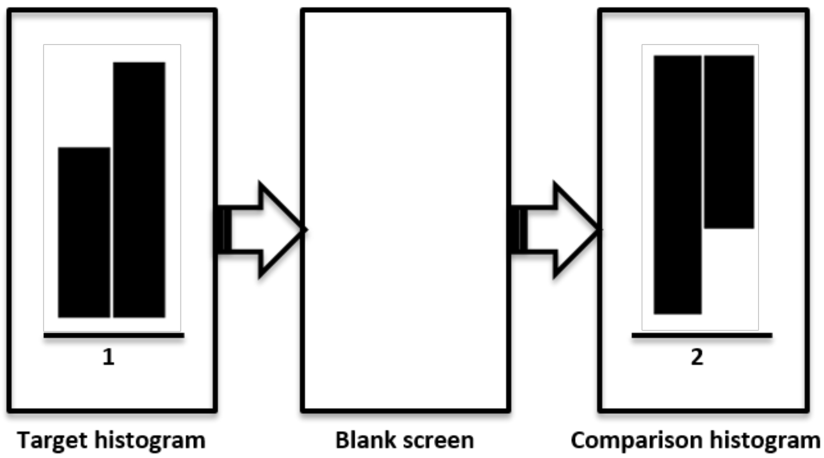


Figure 13: Newer version of histogram rotation task with monochromatic theme.

5.3 Navigation task

The *Navigation task* addresses the *Navigation & Recall* aspect of spatial memory. To complete the task, participants are asked to navigate to six items on a canvas in a specific sequence six times. There were no major challenges discovered regarding this task in the pilot study. Therefore, I briefly describe the old version of the task (including the minor changes performed) as well as the new version in the following sections.

5.3.1 Older version of Navigation task

Participants were instructed to find six items for a repetition of six times (36 navigational trails) on a 12 by 9 grid canvas on which 18 different items (same size and colour) were spatially distributed. The centre of the canvas served as a home position and was empty. Once the task started, a destination item was displayed in the centre of the canvas and participants were instructed to navigate to this item on the canvas. Participants then had to navigate to this item on the canvas by panning the screen using the mouse. Once they found the item, they had to overlap the item to be searched with the centre of the screen with a tolerance of 100 pixels around the centre. When participants navigated to the item successfully, a beep sound indicated that the search was successful and the next item is going to appear at the home position of the canvas.

5.3.2 Challenges faced

In the pilot study, I didn't face any challenges regarding the collection and analysis of the data during the *Navigation & Recall* task. However, in the pilot study, I always used one item set on one display size and another item set on another display size. This could potentially have biased the results as one item set might have been easier/more difficult to perform than the other. Thus, for the final experimental design, I aimed to counterbalance (see Appendix A6) the item sets with regard on which display size they were performed with.

5.3.3 Final Navigation task

In the new version of the task, I decided to counterbalance the two-different item sets across the display size. I counter-balanced the item sets so that participants always started the task with different item sets on different display size (see Appendix A6). Remaining aspects like the distribution of the items on the canvas and the number of item to be search (6) as well as the number of repetitions (6) were kept the same as in the original task. Figure 14 represents the distribution of the items on the canvas and the empty home position on the canvas.

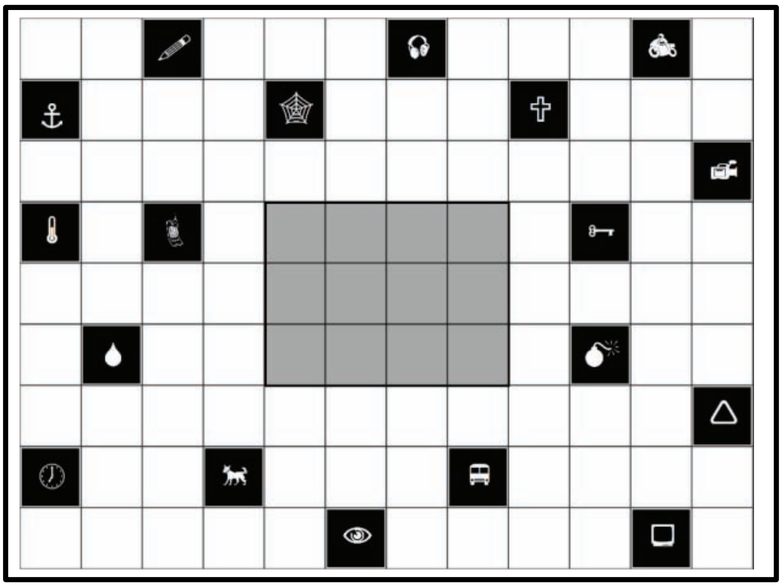


Figure 14: A canvas with 18 items scattered over it and with an empty centre position

5.4 Tangram task

The main focus of the *Tangram task* is to assemble a collection of separate pieces into a target image. Therefore, this task addresses the *Assembling/Disassembling* aspect of spatial memory. In the following sections, an overview over the old version of this task, challenges faced in the pilot study as well as the resulting changes in the new version of the task are discussed.

5.4.1 Old version of the Tangram task

The tangram task[4] involved the assembling of 5 or 7 multicolour geometrical shapes into a target image. All geometrical shapes were originally placed at the left side of the screen and the target image was shown the right side of screen (see Figure 8) The goal of the task was to assemble the geometrical shapes into the target image by moving and rotating them. For the interaction, a mouse is used to drag and move the shapes on the target image. There were also two symbols at the bottom left of the screen which were used to rotate the given shapes either in clockwise or in anti-clockwise direction with the orientation of 90-degree, 180-degree and 270-degree. For doing so, participants were asked to click on those symbols. All shapes could be rotated, moved or dragged an unlimited number of times.

The task consisted of four trails. In the first two trails, participants were instructed to assemble 5 shapes into a target image whereas in the third and fourth trail, they were instructed to assemble 7 shapes into a target image each. To assemble two different number of shapes, represent two different difficulty levels in the task. The appearance of both difficulty levels was counterbalanced so that sometimes participants had to start with 5 shapes, and sometimes participants had to start with a 7 shapes image. There was no given time limit to complete the task.

5.4.2 Challenges faced

I faced challenges regarding the task procedure as well as regarding the set-up of the tangram task after analysing the collected data in the pilot-study. In the following, I focus on these two aspects separately.

Challenges in task procedure

I consider the influence of colours on user's performance as a challenge of this task. Multicolours might have influenced participant's behaviour and biased results.

Difficulty levels complicated analysis and lengthened the duration of the task. With respect to time, the whole task (2 trials with 2 different difficult levels each) was consuming a lot of time in which case I aimed to shorten the duration of the task in the final experimental design.

Challenges in set-up

In the final version of Tangram task, I have two different item sets. Between these item sets, one may be easier or tougher than the other item sets. So, to avoid any biased participant's performance, I aimed to have counterbalancing between two item sets across two display size (see Appendix A6).

[4] http://www.tangramgames.co.uk/

5.4.3 New version of the Tangram task

In the newer version of the task[5], participants were required to assemble a collection of uni-coloured shapes into a target shape. In detail, they were instructed to assemble seven brown colour geometrical shapes into the target image and this was repeating three times with three different target image. The number of shapes was kept constant at 7 for all three trials. For the user interaction, mouse and keyboard were used simultaneously. Two specified keys on the keyboard were used to rotate the geometrical shapes whereas the mouse was used to drag the shapes onto the target image. The complete picture of the new version of the tangram task is shown in Figure 15 where all seven shapes are shown in a single brown colour and target image is shown in grey colour over the light colour background.

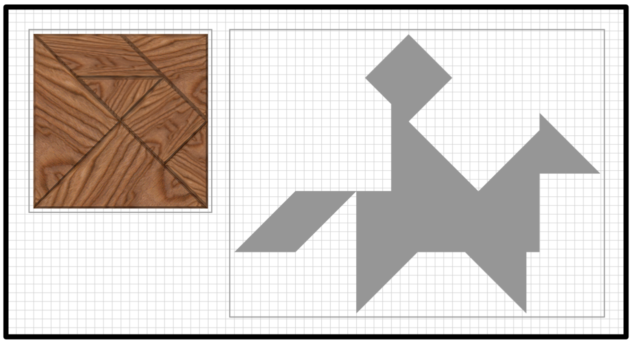


Figure 15: An image of newer version of the tangram task.

In the new version of the task, participants are asked to assemble the target shape a total of three times which resulted in a shorter task duration compared to the old task. Sticking to a single colour for the geometrical shapes reduce the influence of colours on users' task performance.

There are two different item sets in the task and in each item set there are three different target images to assemble. Only the two item sets were counterbalanced by following Latin Square counterbalancing design, across the two-display size (see Appendix A6). No counterbalancing was applied on the three target images in the two item sets.

Briefly description about older version of all the four tasks, challenges faced with those tasks during the pilot study and the final tasks (newer version) for the experiment after overcoming the proposed challenges are mentioned in this chapter. Next chapter is focused on the followed experimental design in the experiment.

[5] http://download.cnet.com/Tangram-7/3000-2111_4-10973833.html

6 Experimental Design

I explained the implementation of the tasks in the previous chapter, I want to focus here on the design of the experiment. The experiment is designed as a within-subject study design. Each task is performed two times by all participants, once in front of the large display and once in front of the small display. An overview of the tasks, apparatus used, recruitment of the participants and the procedure followed to conduct the study is mentioned in the following chapter.

6.1 Apparatus

In the choice of the displays, I aimed to reflect display sizes that resemble real life devices: devices which we use in our daily life like tablets or monitors (comfortable viewing size for routine things) and devices that we see in public (To show the images bigger and from long distance). I have decided to work with two types of display sizes and named them LARGE display size and SMALL display size.

I have selected the Perspective pixel 55" display as LARGE and the Microsoft surface 2 pro 10.6" display as SMALL. The height and width of the both display sizes with their resolutions are mentioned in table 2.

Apparatus	**Diagonal**	**Height**	**Width**	**Resolution**	**Aspect Ratio**
Microsoft surface 2 pro (Small display)	10.6´´	9´´	16´´	1920 × 1080	16´´/9´´ = 1.77´´
Perspective pixel (Large display)	55´´	26.8´´	47.6´´	1920 × 1080	47.6´´/26.8´´= 1.77´´

Table 2: Description of the selected display sizes in terms of diagonal, height, width, resolution and aspect ratio.

The images of both devices are shown in figure 16. Both displays run at a resolution of 1920 × 1080 and were calibrated to be of roughly equivalent colour, brightness and contrast.

Figure 16: Images of larger screen (Perspective pixel, 55'') and smaller screen (Microsoft surface 2 pro, 10.6'') for the experiment.

To keep the visual angle constant and to keep a constant viewing distance, I have set the distance from the user to the small display to 27''. This was done to allow for a comfortable viewing distance that is comparable to the distance of the user to the display in a common office setting. In order to keep the visual angle constant, this resulted in a distance between the large display and the user of 78''. Therefore, the distance between the two displays is 51''.

In order to keep the visual angle constant, the eye-height and the centre of both displays from the ground is set for 51'' since the environmental context around each display could potentially affect participants.

A sketch showing the total setting is shown in figure 17 where the distance between the small display and the participant and the distance between the large display and the participant is shown. Figure 17 also shows the position of the eye-height from the ground.

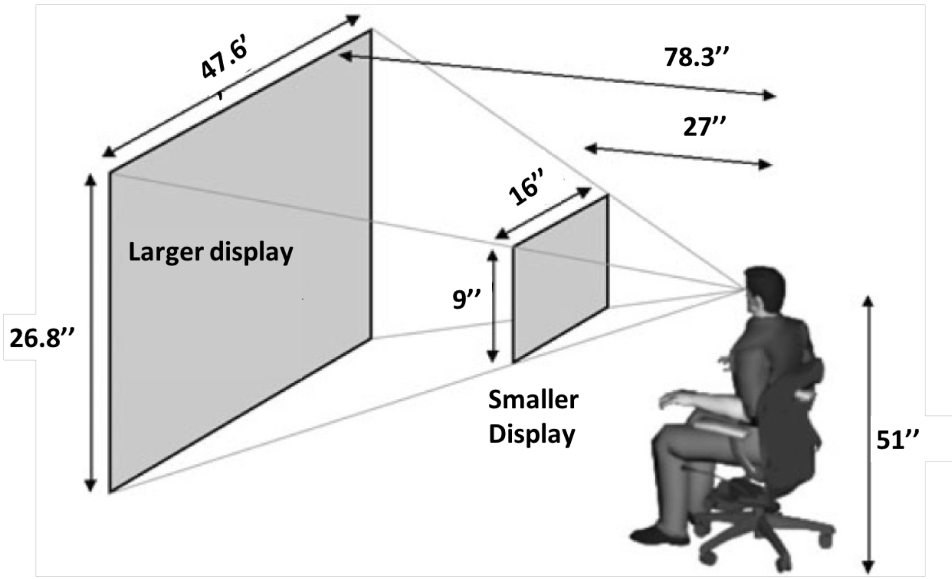


Figure 17: A sketch to represent the total settings for maintaining the visual angles between two display sizes.

The complete picture of the lab set-up with the large and the small display and with a revolving chair in front of them (where participant will sit during the experiment) is shown in the Figure 18.

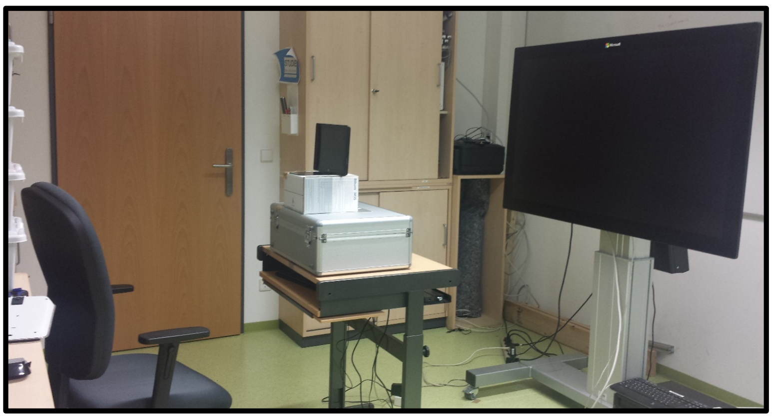


Figure 18: The picture of lab with the two displays and the position of the participant to keep the visual angle constant.

6.2 Tasks

To accomplish my research goal, I have planned an experiment with four different tasks where each of the tasks reflects one of the aspects of spatial memory (and therefore one of the research objectives). The detailed description of the tasks and their implementation has been mentioned in chapter 5. Table 3 summarizes the four tasks, outlining the aspect they address and providing a brief figure for all of them.

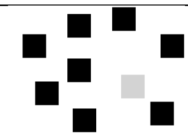
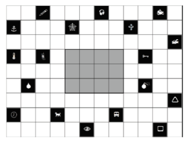
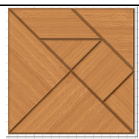
No.	Task	Name of task	Aspect of spatial memory	Figure
1	Task 1	Spatial Span task	Visual recall	
2	Task 2	Histogram Rotation task	Rotation	
3	Task 3	Navigation task	Navigation & Recall	
4	Task 4	Tangram task	Assembling/Disassembling	

Table 3: A summary of the tasks and related aspects of spatial memory with respective figures.

6.3 Procedure

The final experiment conduction is started with 35 participants in the interaction lab of the Human Computer Interaction group[6] at the University of Konstanz. The total duration of the experiment session was 70-75 minutes. The whole procedure of the experiment session is mentioned in Figure 19.

[6] http://hci.uni-konstanz.de/

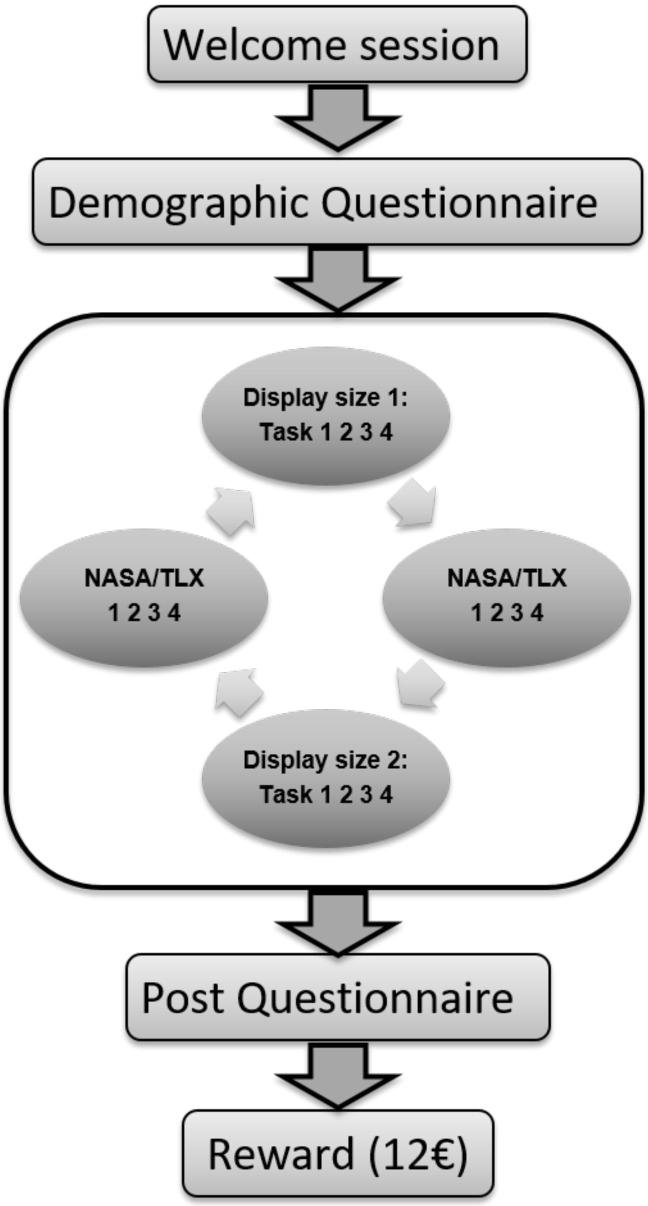


Figure 19: Complete picture of all the steps followed in the final experiment.

Participants were asked to make themselves comfortable in the lab and after a brief introduction, I started the session. At the beginning of the session, participants were instructed to sit comfortably on a chair and asked neither to change his/her body position and head position nor to drag the chair. Later, I adjusted the height of the chair according to participants' height to full-fill the above-mentioned lab settings (see section 6.1).

Then, participants received a *Welcome Letter* (see Appendix A1) containing a brief overview of the experiment and some information regarding the whole session (duration of session, reward money, role and rights of participants). Subsequently, they were asked to sign *Letter of Consent* (see Appendix A2), a declaration that participants are allowing me to record, process and publish the collected and anonymised data in future publications to contribute to the HCI research community. As a last step of the introductory phase of the session, participants were handed out a *Demographic questionnaire* (see Appendix A3) where questions related to their personal information (age, gender, nationality, education and many more) and related to computer usage were asked.

After completing these forms, the experiment was started. At the beginning of each task, I showed a demonstration of the task so that they could become familiar with the task easily and could ask questions (if occurred) before his/her performance. I was also instructing them about the input devices before the tasks, for example when to use the keyboard or the mouse. The selection of display size and the selection of first task & the sequence of remaining tasks is counterbalanced (see Appendix A6). For each task, I showed the demo of the task on that display size, that the participant was going to perform on first and the similar sequence (first demo then performance) is following for the remaining tasks. Once the participant was done with all the four tasks in front of one type of display then all tasks were performed again in front of another display. Before the participant's performance, I have asked them whether they like to receive another demo on the second display.

At the end of each task in front of one type of display, participants were instructed to fill in the *NASA/TLX*[7] questionnaire (see Appendix A4). In the *NASA/TLX*, there are six subscales to measure the workload in form of Mental demand, Physical demand, Temporal demand, Performance, Effort and Frustration (Hart, 2006). Participants filled total 8 *NASA/TLX* during the whole experiment that is 2 times for each task because of two displays.

Once the participants had performed all 4 tasks in front of both displays, they were handed out a *Post questionnaire* (see Appendix A5) related to their overall feedback of experiment in form of three questions.

At the end of the session, participants received reward money of 12 Euro and signed a receipt of confirmation of getting money. Participant's farewell is the last phase of the session.

[7] NASA/TLX: National Aeronautics and Space Administration Task Load Index

6.4 Participant's recruitment

Participants were recruited by creating a Doodle poll[8] over Facebook in the group of students of University of Konstanz (Uni Konstanz - International - Winter 2016/2017) where a brief introduction of the experiment, the money of reward, and the total duration of the experiment were explained. No special requirement or restriction was asked for from the participants. All participants belonged to the University of Konstanz and are currently involved in different departments, either as students or employees. As the experiment consists of 4 different tasks with different difficulty levels in front of two displays so in that order after following 'Latin Square' counterbalancing design, I counterbalanced the four tasks across two display size with total 40 participants and performed the experiments with 35 random versions of these 40 defined one. I managed to recruit 40 participants and finally conducted the experiment with 35 participants (I scheduled the experiment between the 1st week to mid of the 4th week in December but due to vacations 3 participants cancelled the appointment and 2 of them didn't show up on the scheduled date and time).

6.5 Participants

35 participants (20 females and 15 males) took part in the experiment. The cultural background of the participants was very diverse. 8 participants belong to Romania, 4 were Germans, 4 were Ukrainians, 2 were Italians, 2 were Russians, 2 were Egyptians, 2 were Azerbaijanis' and from remaining participants, one participant belonged to Pakistan, Bangladesh, Lithuania, Czech, China, Australia, Serbia, Spain, Franc, Poland, and Slovenia each. The mean age of participants was 23.74 years (SD=2.7). 20 out of 35 were graduate students (bachelor's) Whereas the remaining 15 belonged to post graduation (master's). Most of the participants were students (30 out of 35) and remaining 5 were employees (research associates). Most of the participants were not suffering from vision problem but 5 participants were suffering from far-sightedness and 9 were suffering from short-sightedness. However, they all used glasses or contact lenses in order to have corrected-to-normal vision. Participants have used a computer for an average of 13.34 years and use their computers daily more than three hours. They have rated themselves on 5-point Likert scale as an expert in the use of the keyboard (Mean = 4.22) and the mouse (Mean = 4.34).

In this chapter, I explain the whole experimental design based on apparatus used, performed tasks, followed procedure during the experiment and some demographic information about the participants. Next chapter is all about the analysis of the collected data during the experiment.

[8] http://doodle.com/poll/i7dfs76wwqkp54z6

7 Analysis

In order to obtain the results from the collected data and to see the outcome of the research objective, I went to analyze the collected data. In this chapter, I discuss the analysis for each of the four tasks describing their respective measures for the task performance. Before going to the analysis process in detail, I mention here a brief overview on outset of my study planning about statistical tests which I follow during the analysis of collected data.

I have collected the data in the form of Score and the collected data is measured on either ratio or interval scale. A brief overview of different variables and their respective scales on which they are measured in the experiment is mentioned in table 4.

Task name	Variables	Scale
Spatial Span task	Time	Ratio
	Error	Ratio
Histogram Rotation task	Reaction time	Ratio
	Accuracy	Ratio
Navigation task	Path length	Interval
	Time	Ratio
Tangram task	Time	Ratio

Table 4: Kind of scale used to measure the dependent variables with their respective tasks.

There is only one independent variable (display size) of fixed category is present in the experiment. Similarly, there are two dependent variables in each of the task so in short, total 8 dependent variables are measured throughout the experiment (7 out of them are summarized in the above-mentioned table). The goal of the analysis is to find the difference of performance on dependent variables so I follow the *Experimental method* deign in my experiment. All participants are taking part in all of the conditions in the experiment so I follow *Repeated measures* with two groups. In the same order, I focus to select between parametric and non-parametric test; as a parametric test, I select dependent t-test after checking following mentioned conditions:

- I check for dependent variables; they are measured on continuous scale. In my case, all the dependent variables are measured either on ratio or on interval scale.
- Independent variable has two matched pairs. In my case, there are two display size.

- Later I check for outliers and find that there are no significant outliers in the difference between the two matched pairs (large and small display size).
- I check for the normality; the distribution of the differences in the dependent variable between the two matched pairs is normally distributed.

As a non-parametric test, I choose Wilcoxon signed-rank test after checking that following mentioned conditions are fulfilling or not:

- Dependent variables are measured at the continuous level (ration and interval scale in the experiment).
- Independent variable has two related groups (large and small display size in experiment).
- The distribution of the differences between the two related groups are symmetrical in shape.

After a brief introduction about the planning and selection of statistical tests now I mention the structure of this chapter in form of sections and sub-sections. This chapter is divided into 6 sections and their respective sub-sections. Each sub-section is again divided depending on the measures used in particular task and used documents. Afterwards, I report how the data obtained from the NASA/TLX and the post-questionnaire was analyzed.

7.1 Spatial Span task

Time and the error are the two dependent variables for this task. A detailed description of the analysis procedure for the *Spatial span task* is mentioned below.

7.1.1 Time

The time is measured by calculating the time taken by all the participants from the flash of the last block until the first click on a block. Then the next time is taken between this click and next one and so on until the participant clicked on the required number of blocks. This results in a separate measure for each block click that can be summed to one measure from the flash of the last block until the participant clicked on the required number of blocks. For example, in level B3 (there are 8 blocks from B2-B9 in the task) three blocks flash in a specific order and the participant is asked to click on these blocks following the same order. So, for the first block, the time is measured from the last flash until the participant clicks on the first block. After finishing the first click on first block, the time is reset to zero. Again, once he clicks on the second block, the time is recorded again and same with the third block. As the trials consist of different numbers of blocks (B2-B9), the sum of all clicks would result in larger numbers for the trials consisting of more blocks. Therefore, I do not measure the sum of the time taken for all clicks but the average time for each block (e.g. the sum divided by the number of blocks). This results in an average number of click time (in millisecond) for each block and for both of the investigated display sizes. I calculated the average time for each block in front of both the display sizes. For example, I compare the average time taken in block B2 in front of the large display versus the average time taken in block B2 in front of the small display. The same is done for blocks B3, B4, B5, B6, B7, B8 and B9.

Results of this analysis are plotted in a bar-graph (see Figure 21). In order to test for significant differences, I performed statistical tests on the collected data using IBM SPSS statistics[9]. In order to decide which tests to used, I checked the collected data with respect to normal distribution using the Shapiro-Wilks test ($p > .05$). If the data is normally distributed, I perform a dependent t-test to analyse for statistical difference between the two display sizes. In cases in which the data is not normally distributed, I perform the Wilcoxon signed-rank test to analyse for statistical differences. Following mention table 5 is giving a short overview which test is performed on which block.

Blocks	*p*-value	Name of Test
B2	$p = .000$	Wilcoxon signed-rank test
B3	$p = .000$	Wilcoxon signed-rank test
B4	$p = .000$	Wilcoxon signed-rank test
B5	$p = .000$	Wilcoxon signed-rank test
B6	$p = .806$	dependent t-test
B7	$p = .002$	Wilcoxon signed-rank test
B8	$p = .681$	dependent t-test
B9	$p = .002$	Wilcoxon signed-rank test

Table 5: An overview on the selected statistical test with p-value in each of the different blocks (B2-B9).

7.1.2 Error

In the *Spatial Span task*, error is measured by counting the right and wrong responses in both the attempts of all the levels, from B2 to B9. Thus, error is measured in terms of a success rate and this success rate is measured by counting the number of right responses (in percent) for each of the blocks in front of both display size. There are two attempts to response in each of the block so the maximum number of responses in the whole task is sixteen.

At the beginning of the analysis phase of the error, I perform a comparison between average percent of success in front of larger and smaller display size irrespectively of the number of blocks. I checked for significant difference using IBM SPSS statistics. The collected data was normally distributed, as assessed by Shapiro-Wilks test ($p > .05$) and I choose to perform the dependent t-test for the calculation of significant difference.

[9] IBM SPSS Statistics: www-01.ibm.com/software/de/analytics/spss/

In order to check for significant differences in error, firstly I have checked for the normality. The data was not normally distributed so I have decided to analyze the collected data by using Wilcoxon signed-rank test. Following mentioned table 6 is giving a short overview on the test performed with *p*-values.

Blocks	*p*-value	Name of Test
B2	$p = .000$	Wilcoxon signed-rank test
B3	$p = .000$	Wilcoxon signed-rank test
B4	$p = .000$	Wilcoxon signed-rank test
B5	$p = .001$	Wilcoxon signed-rank test
B6	$p = .000$	Wilcoxon signed-rank test
B7	$p = .000$	Wilcoxon signed-rank test
B8	$p = .000$	Wilcoxon signed-rank test
B9	$p = .000$	Wilcoxon signed-rank test

Table 6: An overview on the selected statistical test with p-value in each of the different blocks (B2-B9).

7.2 Histogram Rotation task

In the Histogram Rotation task, the two dependent variables are reaction time and the accuracy. In the following section, I will discuss how these two measurements were analysed.

7.2.1 Reaction time

Reaction time is measured by calculating the time taken to respond for congruency and in-congruency in the *Histogram Rotation task*. The calculation of this reaction time starts from showing the second histogram until the response is received from the participant.

I start the analysis of collected data by comparing the average time taken (in milliseconds) with all the three types of histograms in front of large and the small display. To represent the result of this general comparison, I select the bar charts (see Figure 24). In order to test for significant differences between the reaction time with the small and the large display, I performed some statistical test. Again, I checked for the normality first and realize that data sampled is normally distributed (using the Shapiro-Wilks test ($p < .05$)). Therefore, I conducted a dependent t-test to check for significant difference between the two display sizes.

To further analyze the reaction time for the histogram-task, I separated the analysis into two parts: (a) depending on the type of histogram (2-bar, 4-bar and 6-bar) and (b) depending on the type of rotation (by 90-degree, 180-degree and by 270-degree).

I start with the analysis of reaction time based on the type of histogram where I compare the reaction time individually for each type of histogram in front of both the display sizes (reaction time taken with 2-bar histogram in front of larger and smaller display and same for 4-bar and 6-bar). At the starting of this analysis, I perform basic comparison between both the display size based on the type of histogram and later I perform the test for the significant differences by using IBM SPSS statistics. I start this test from checking for the normality of the collected data and based on the result of normality I select the further test which are mentioned below in the table 7.

Types of histogram	***p*-value**	**Name of test**
2-bar	$p = .020$	Wilcoxon signed-rank test
4-bar	$p = .846$	dependent t-test
6-bar	$p = .196$	dependent t-test

Table 7: An overview on selected statistical tests with their p-values in 2-bar, 4-bar and 6-bar histograms.

The next analysis was based on the type of rotation where I perform basic comparison in front of both the display size depending on order of rotation (comparison between rotation by 90-degree in front of larger and smaller display and same type of comparison for 180-degree and 270-degree). As this basic comparison seemed interesting, I decided to perform some further test to check for significant differences. Firstly, I checked for normality using the Shapiro-Wilk test. Based on the result of normality, I perform either Wilcoxon signed-rank test or a dependent t-test on the identified data. Table 8 provides a short overview of selected tests.

Types of orientation	***p*-value**	**Name of test**
90-degree	$p = .000$	Wilcoxon signed-rank test
180-degree	$p = .000$	Wilcoxon signed-rank test
270-degree	$p = .385$	dependent t-test

Table 8: An overview on selected statistical tests with their p-values in 90-degree, 180-degree and 270-degree orientation in histograms.

7.2.2 Accuracy

Accuracy can be measured by counting the right and wrong attempts during the decision for congruency and in-congruency between the histograms. The participant's response was recording in the binary order; 1 is for right response and 0 for the wrong response. By calculating the total number of 1's and 0's in all the three types of histograms, I compare the accuracy (success rate) in front of both the display sizes. It was calculated as percentage of right response among all responses (total 8 responses are possible with each type of histogram). After the basic comparison, I perform the Wilcoxon signed-rank test (data was not normally distributed – include results from Shapiro Wilk test here) to test for significant difference regarding the success rate in front of the large and the small display.

As with the analysis of the reaction time, accuracy is also analysed in two ways: based on the type of histograms and type of rotation. Firstly, I perform the basic comparison based on the type of histogram in front of both the display size. In other words, I compared here the success rate in front of both the display size with 2-bar, 4-bar and 6-bar histograms. I also checked whether there is significant difference in accuracy in front of both display size selecting either the dependant t-test or the Wilcoxon signed-rank test depending on whether the data was normally distributed or not. A detailed overview over the selected tests is provided in Table 9.

Types of histogram	***p*-value**	**Name of test**
2-bar	$p = .000$	Wilcoxon signed-rank test
4-bar	$p = .069$	dependent t-test
6-bar	$p = .007$	Wilcoxon signed-rank test

Table 9: An overview on selected statistical tests with their p-values in 2-bar, 4-bar and 6-bar histograms.

In the same order, I analyzed the accuracy again depending on the types of rotation this time. Firstly, I perform some basic comparisons depending on type of rotation (90-degree rotation in front of one display size versus 90-degree rotation in front of the other display size and same type of comparison with other two orientation types). I obtained some interesting results from this comparison (see Figure 29) and I decide to test for significant differences. The detailed description of the selected tests is shown in Table 10.

Types of orientation	***p*-value**	**Name of test**
90-degree	$p = .020$	Wilcoxon signed-rank test
180-degree	$p = .205$	dependent t-test
270-degree	$p = .278$	dependent t-test

Table 10: An overview on selected statistical tests with their p-values in 90-degree, 180-degree and 270-degree rotation in histograms.

One problem that was faced while analyzing this task was the ‘problem of timeout’. The total time to complete the task was fixed to 3.15 minutes that also means 1.05 minutes for each type of histogram to response to total 8 attempts. If participant is taking much time at the beginning to response, then sometimes they were not able to response for the last attempts because of timeout or could get chance to face only less than eight attempts. This type of problem is analyzed in front of both the display but it was observed more in front of large display.

This drawback of the task is kept in mind carefully while calculating the accuracy. The accuracy is measured by calculating the percentage of right responses among all responses. There are total 24 responses are possible in the task, that is 8 responses in each type of histogram. If during the analysis, it was observed that participant has faced the ‘problem of timeout’ then the all responses (that was 24) were considered according to his performance. For example, if s(he) has faced total 23 attempts instead of 24 throughout the task then his/her performance is calculated based on 23 attempts.

7.3 Navigation task

Concerning the *Navigation task*, the path length as well as the time taken to navigate the path was measured. The analysis of the collected data is presented below in this section.

7.3.1 Path length

The path length is the total path covered by one participant to navigate from the start position (centre of the canvas) to the respective item. Participants have to navigate the 6 items with 6 repetitions. It means total 36 navigational trials for each of the participant.

At the end of the task, I had huge amount of collected data to analyse and the goal of the analysis was to extract six different path lengths (one for each repetition) and their respective time to cover those path lengths. This huge amount of data was sorted easily with the help of Macros in Microsoft excel. The data was recorded in the .csv file, then later for the analysis, it was converted to. xxl and where the Macros were applied. For each search, the actual path length is divided through the optimal path length to account for the fact that the shortest path to navigate to an item was different for each item. This results in a ratio of covered path/optimal path that results in 1 if the participant navigated the shortest possible path.

At the beginning of the analysis, I have compared the total path length in front of both the display size that is the average covered distance, calculated in all 30 navigational trials by each participant (average of 30 * 35 trails). For example, D2 is representing the average covered distance in the second navigation trial of all the items and by all the participants. Same type of calculation is applicable for D3, D4, D5 and D6. The first navigation distance (D1) is not considered because at the very first time that a participant navigated to an item. (s)he searches the canvas for the item, thus the result of this search is not based on spatial memory. This is the reason why I have considered only 30 navigational trial here instead of 36. To check for the significant differences between the large and the small display, I have performed the dependent t-test because the collected data was normally distributed that is assessed by Shapiro-Wilks test (p> .05).

In addition, I was interested to analyse whether there is any difference between the repetitions (e.g. between D2, D3, D4, D5 and D6) for each of the two displays. For doing so, I have compared each navigational distance with respect to each other in front of the large and the small display where first navigation distance is considered as D2 and the last navigation distance is D6. D1 is not considered in the analysis because of the above-mentioned reason. To test for significant differences in each navigational distance (D2-D6), some statistical tests have performed whose results are mentioned in the table 11. The selection of the tests on different navigational distances are based on the normality. If the identified data is normally distributed, then dependent t-test is applied otherwise Wilcoxon signed-rank test is applied on the data sets.

Different distances	***p*-value**	**Name of test**
D2	$p = .771$	dependent t-test
D3	$p = .819$	dependent t-test
D4	$p = .215$	dependent t-test
D5	$p = .040$	Wilcoxon signed-rank test
D6	$p = .000$	Wilcoxon signed-rank test

Table 11: A description about chosen statistical tests with p-value in different navigational trials (D2-D6).

7.3.2 Time

In addition to the path length, the time was taken to investigate how long participants take to navigate to the items on the canvas. As mentioned above, participants were asked to navigate to 6 items in 6 repetitions. For each of the repetitions, the time taken was recorded separately. In other words, for six navigation distances (D1-D6), there are six respective times (T1-T6). To calculate the time separately for each of the trail, a Macro has been used in the analysis that results in six different times for each of the item i.e. T1, T2, T3, T4, T5 and T6. In the analysis, T2 is representing the average time taken in the second navigation trial of all the items and by all the participants. The same is applicable for T3, T4, T5 and T6. The time T1 is not considered in the analysis because of the above-mentioned reason.

The analysis starts with a basic comparison of the times taken in front of both the display size in all navigational trials. To check for the significant differences, I perform the Wilcoxon signed-rank test on not normally distributed data (Shapiro-Wilks test (p<.05)).

Also, I have performed some comparison to see whether there is any time difference in the separate trails (T2-T6) in front of both the display sizes. I perform some statistical tests on the collected data to check for the significant differences. Again, the statistical test started with the check for normality using the Shapiro Wilks test. I have chosen the dependent t-test and the Wilcoxon signed-rank test for normally distributed and not normally distributed data respectively. A complete overview on test selection on different time trails is shown in Table 12.

Different time	*p*-value	Name of test
T2	p = .785	dependent t-test
T3	p = .007	Wilcoxon signed-rank test
T4	p = .005	Wilcoxon signed-rank test
T5	p = .001	Wilcoxon signed-rank test
T6	p = .000	Wilcoxon signed-rank test

Table 12: An overview of selected statistical tests with p-values on different time trials (T2-T6).

7.4 Tangram task

In the *Tangram task*, the time taken and the strategies followed to complete the task are considered as a dependent variable. In the following section, I explain the analysis process of both measures. I am going to start with the first measure 'time' that is followed by the second measure 'strategies'.

7.4.1 Time

Time measures the total time taken to assemble the shapes into the target image. A comparison is performed to see the performance differences concerning the time taken in front of the large and the small display size. For doing so, the average of all the times taken (seconds) in the three trails for each display size are considered. Obtained result showed tendencies that motivated me to check for significant differences between the two display sizes. In the same order, firstly I check for the normality and after getting the result form Shapiro-Wilks test (p<.05), it is concluded that data sets are not normally distributed that results I choose the Wilcoxon signed-rank test to test for statistically significant differences.

For this task, only data from 32 participants have been considered due to some technical problems as the data of three of the participants could not be collected in front of smaller display. Four participants could not complete to assemble one of the target image among three in front of large display and similarly one participant could not assemble two out of three target images in front of small display size. They tried to assemble the shapes into the target image but after trying for few minutes they quitted the task. Because of un-completed performance of those participants (quit the task), the time is considered for them until they have tried to assemble the target image; for example, one of the participant tried to assemble the target image for 3.15 minutes and after this time immediately quit the task so the recorded time for that participant is 3.15 minute.

7.4.2 Strategies

To see whether there is any difference in strategies selection while performing the task in front of both the display size. For doing so, I have recorded the session by using a screen recorder[10] so that later I can analyze their chosen strategies to complete the task. I aim to see the performance difference in front of large and small display in two conditions; strategies followed to complete the task and strategies followed to overcome the challenges faced during the task completion.

I was focused to see the strategies followed to complete the task. To accomplish the goal are the participants are making any pre- strategies (pre-planning) before starting to attempt with the given shapes? In other words, are they thinking first (how to start or from where to start) after the appearance of task (target image and seven shapes) on the screen).

I was also interested to see, how they are dealing with the challenges (if they stuck in between the task)? For doing so, are they starting the assembly of shapes from beginning (putting all the assembled shapes again back to its source place from the target image) Or are they re-shuffling/adjusting the shapes over the target image? To complete the task, are they changing the strategies followed and adopting any new technique or are quitting the task?

7.5 NASA/TLX

To calculate the task load of each task, the NASA/TLX was filled out by all participants at the end of each task in front of both displays. Two parameters have taken to represent the results of NASA/TLX of each task; display size and the scale. Display size is representing the data in front of larger and smaller display. Scale is dividing the total workload in the six 20-stage sub-scales those are *Mental Demand, Physical Demand, Temporal Demand, Performance, Effort* and *Frustration*. To evaluate the different questionnaire, the 20-stage scales were translated to number between 0 and 100 where 0 and 100 are representing the low and high demand respectively. Firstly, I analysed the data for some basic comparison and later for statistically significant differences for each of the tasks between the small and the large display. Table 13 shows the tests used for this analysis, depending on the distribution of the data.

[10] ActivePresenter: https://atomisystems.com/download/

Name of the task	*p*-value	Name of test
Spatial span task	$p = .031$	Wilcoxon signed-rank test
Histogram rotation task	$p = .447$	dependent t-test
Navigation task	$p = .778$	dependent t-test
Tangram task	$p = .226$	dependent t-test

Table 13: A description of selected statistical tests with p-values and their respective tasks in NASA/TLX.

7.6 Post-questionnaire

At the end of the experiment, participants were asked to fill a post-questionnaire to give the overall feedback on their experience with the tasks. Based on the experiment, there are total three questions to answer.

In first question, participants were asked to mention about the display size which they found best to perform the experiment? I read the answers given by participants and analysed their choice manually, just by counting how many of them preferred to perform in front of Large display and how many of them preferred to perform in front of smaller display. I have mentioned the reason of their preference of display size of few participants in the section 8.6.

In the second question, participants handed out the ranking for the tasks depending on the easiness to toughness which they felt during performance. They are asked to mention 1 for the easiest task and 4 for the toughest task. To analyse the collected data, I have taken the average rank of each task which is mentioned by all participants and presented the result over bar charts.

In the third and last question, participants were asked to mention their preference of display size to perform the particular task. Participants were mentioning either larger or smaller display size (their preference) in front of the image of tasks in the post questionnaire. I have counted the total number of responses in favour of larger and smaller display size for each of the task and presented the result over bar chart.

In the chapter 7, I describe the analysis procedure followed in all the four tasks. I discuss the analysis for each of the four tasks describing their respective measures for the task performance. Results obtained from the analysis is mentioned in the next chapter 8.

8 Results

This section provides an overview of the experiment's results. Various measurements were taken in the course of the study. Multiple types of information were gathered throughout the experiment with regard to participants' performance, their work load and their preferences of display size to perform the tasks. In the following, the results of these analyses are summarized for each task separately.

8.1 Spatial Span task

As an independent variable, display size is considered. As a dependent variable, I measure time and the error. In the following, I discuss them separately.

8.1.1 Measure – Time

The time is measured by calculating the time taken by all the participants from the flash of the last block until the first click on a block. Then the next time is taken between this click and next one and so on until the participant clicked on the required number of blocks. This results in a separate measure for each block click that can be summed to one measure from the flash of the last block until the participant clicked on the required number of blocks. There are total eight difficulty levels in the task (B2-B9). With the increasing difficulty levels, the number of clicks are also increasing. Thus, I do not measure the sum of the time taken for all clicks but the average time for each block (e.g. the sum divided by the number of blocks). In order to compare the two conditions (small display vs. large display) I calculated the time taken (in milliseconds) to complete the task. Figure 20 shows the results of this comparison.

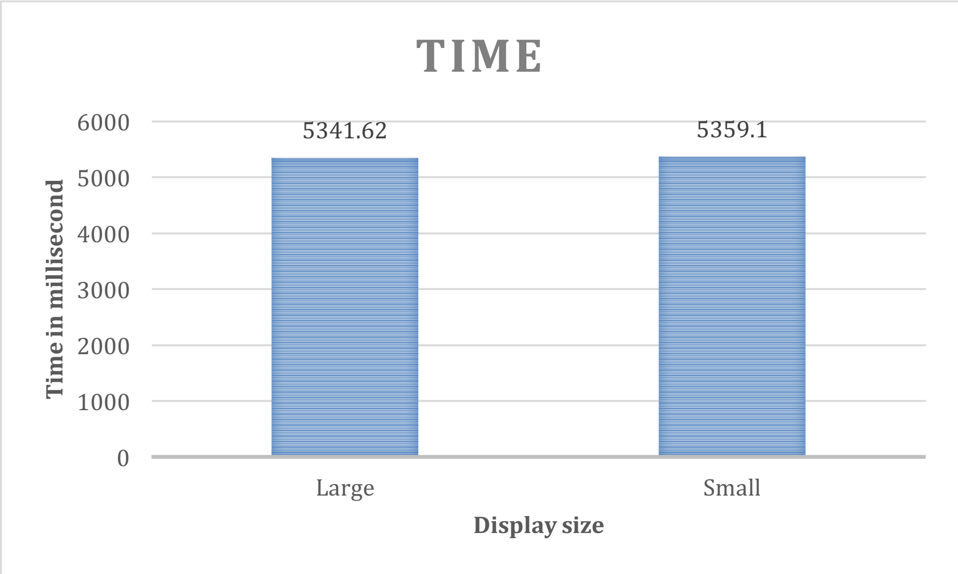


Figure 20: Representation of total time taken by the all participants in front of two displays.

The figure 20 is representing the total average time taken by all the participants to perform the task in front of both the display size where display size is represented on the x-axis and time (in milliseconds) is represented on the y-axis. A dependent t-test was performed to determine if there is any significant difference in performance of time in front of both the display size. Data scores for both the display size is normally distributed, as assessed by Shapiro-Wilks test ($p > .05$). The test results show that there is not significant difference in the time in front of both the display size with t (34) = -.093, $p = .927$.

Figure 21 shows the time for each of the eight blocks separately. Different blocks (B2-B9) are shown on the x-axis and time in milliseconds is shown on the y-axis. Two colors of bars on the x-axis are representing the two display sizes where blue bars represent the data in front of the large display and green bars represent the data in front of the small display.

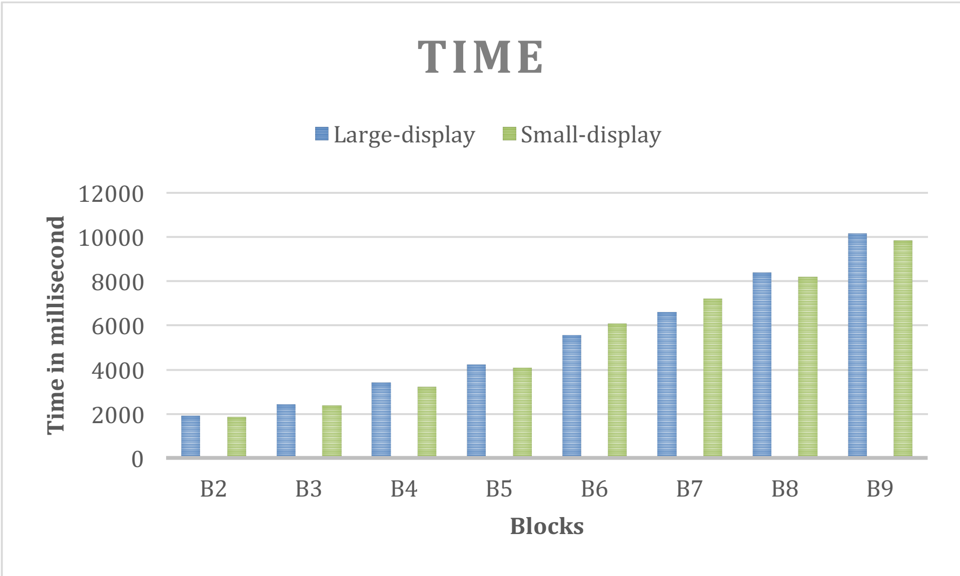


Figure 21: Representation of time taken to response in each block by all the participants in front of large and small display.

Looking at Figure 21, we can see that the time taken to response in each block (B2-B9) is gradually increasing with the difficulty level in front of both the display size. The amount of time taken to response was lowest at the level B2 and it was highest at the level B9. This is due to the increasing number of blocks participants have to click on.

I have performed some statistical tests for each block separately to test whether there is any statistically significant difference concerning the time in front of both displays. The results of these statistical tests are summarized in Table 14.

Blocks	Normality	Name of Test	Exact *p*-Value	Distribution Value
B2	Not normal (p = .000)	Wilcoxon signed-rank test	p = .432	Z = -.786
B3	Not normal (p = .000)	Wilcoxon signed-rank test	p = .889	Z = -.139
B4	Not Normal (p = .000)	Wilcoxon signed-rank test	p = .756	Z = -.311
B5	Not normal (p = .000)	Wilcoxon signed-rank test	p = .743	Z = -.328
B6	Normal (p = .806)	dependent T- test	p = .087	t = -1.762
B7	Not normal (p = .002)	Wilcoxon signed-rank test	p = .351	Z = -.934
B8	Normal (p = .681)	dependent t-test	p = .698	t = .391
B9	Not normal (p = .002)	Wilcoxon signed-rank test	p = .935	Z = -.082

Table 14: Result of the statistical test to measure the time in each block separately, showing the data related to the normality, performed significant test and their respective exact p-value and distribution value.

There are 5 columns in Table 14 where the first column shows the name of the blocks, and the second column identifies whether the data of the block is normally distributed or not as assessed by the Shapiro-Wilks test ($p > .05$). Next columns show the tests that were applied to test for significant differences, and the resulting exact p-values and distribution values for each block. It can be seen in Table 14 that for each of the blocks the value of p is always greater than 0.05 and it can be said that there is no significant difference in the time in any of the blocks in front of both the display size.

8.1.2 Measure- Error

Errors are measured by counting the right and wrong attempts and by considering which one was wrong. To see the performance difference (if any) in front of the large and the small display, I calculated the success rate (% of right attempts) across all blocks. Figure 22 shows the success rate in front of both displays where the display size is shown on the x-axis and the success rate is represented on the y-axis.

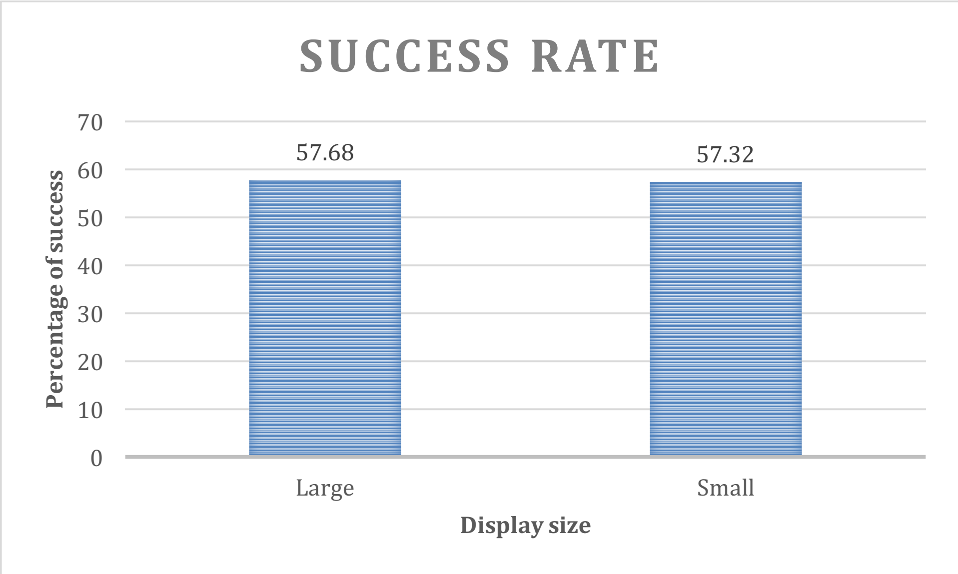


Figure 22: Representation of error (success rate) in percentage by the all participants in front of two displays.

A dependent t-test was performed to determine if there is any significant difference in the performance of success rate in front of both displays. The data scores for both display sizes are normally distributed as assessed by Shapiro-Wilks test ($p > .05$). The test results show that there is not significant difference in the success rate in front of both display with $t(34) = .314$, $p = .756$.

Figure 23 shows the total rate of success divided into the separate blocks. Figure 23 shows the success rate in front of both displays where blue color bars represent the success rate in front of the large display and red color bars represent the success rate in front of the small display. All the eight blocks (B2-B9) are shown on the x-axis and the success rate (in % of right attempts) are shown on the y-axis.

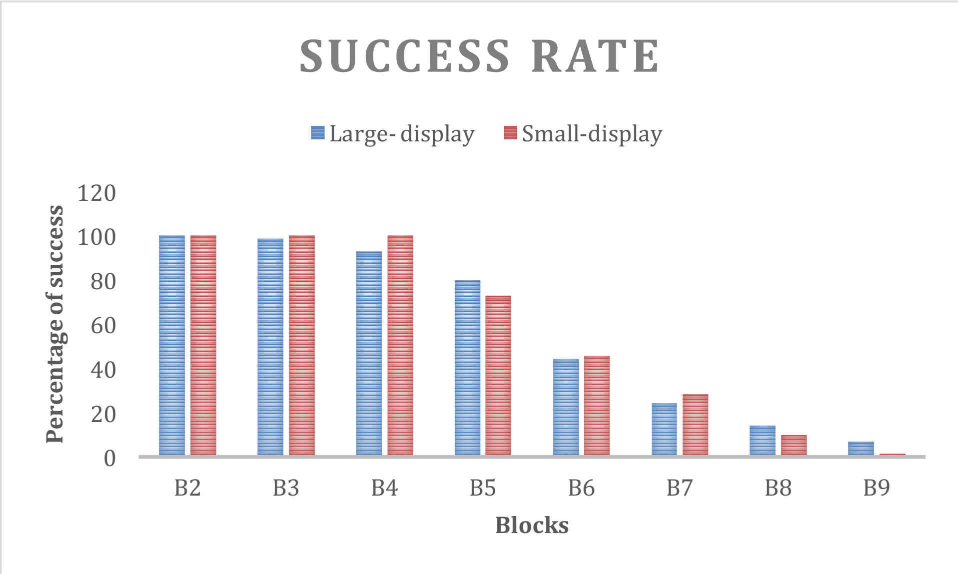


Figure 23: Representation of rate of error (success rate) in each block separately in front of two display size.

As Figure 23 shows, none of the participant made a mistake in block B2 because the success rate is 100% for both displays. Similarly, in blocks B3 and B4 the success rate is 100% for the small display but not for the large display. From Figure 23, it seems that the success rate is gradually decreasing as the difficulty level (number of blocks) is increasing.

I have performed some statistical tests to check the significant difference in the rate of error (success rate) in each of the block separately. The detailed description of the applied tests and related outcomes are mentioned in the table 15.

Blocks	Normality	Name of Test	Exact *p* - value	Distribution value
B2	Not normal (*p* = .000)	Wilcoxon signed-rank test	*p* = 1.000	Z = .000
B3	Not normal (*p* = .000)	Wilcoxon signed-rank test	*p* = .317	Z = -1.000
B4	Not Normal (*p* = .000)	Wilcoxon signed-rank test	*p* = .059	Z = -.1.890
B5	Not normal (*p* = .001)	Wilcoxon signed-rank test	*p* = .302	Z = -.1.031
B6	Normal (*p* = .000)	Wilcoxon signed-rank test	*p* = .538	Z = -.615
B7	Not normal (*p* = .000)	Wilcoxon signed-rank test	*p* = .448	Z = -.758
B8	Normal (*p* = .000)	Wilcoxon signed-rank test	*p* = .593	Z = -.535
B9	Not normal (p = .000)	Wilcoxon signed-rank test	P = .102	Z = -1.633

Table 15: Result of the statistical test to measure the rate of error in each block separately, showing the data related to the normality, performed significant test and their respective exact p-value and distribution value.

It can be seen in Table 15 that for each of the blocks the value of *p* is always greater than 0.05 and it can be said that there is no significant difference in the rate of error in any of the blocks in front of both the display size.

According to participants as they have mentioned at the end of task that after the B5 block it was being tougher to remember the flashing sequences of blocks and in blocks B8 and B9 it was almost impossible. During the analysis, it is found that few of the participants are not clicking on the instructed number of blocks before clicking on the 'Done' button. For example, suppose participant is in B6 that means he has to click on 6 blocks before clicking on the 'Done' button but he is either clicking on less than 6 blocks or on more than 6 blocks. So this type of participant's performance is affecting the results of the task. It can also be seen from the results that as number of blocks are increasing, participants have to remember the more and more difficult it gets that results the more mistakes in their performance.

8.2 Histogram Rotation task

In this task, the independent variable is the display size. As dependent variables, I measure reaction time and accuracy. In the following, I will discuss them separately.

8.2.1 Measure- Reaction time

Reaction time is measured by calculating the time taken to respond for congruency and in-congruency (the time from showing the second histogram until the response from the participant). Figure 24 shows the average time taken for each response (in millisecond) separately for the large and the small display. On the x-axis, the two types of display sizes are shown whereas the y-axis shows average time for response in milliseconds.

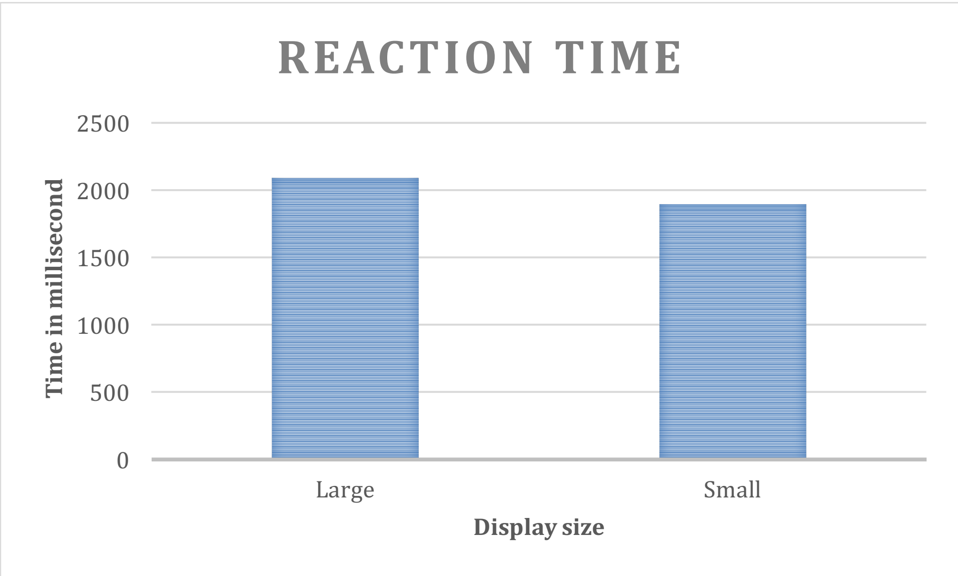


Figure 24: Representation of average reaction time (in milliseconds) by all the participants in front of two display size.

A dependent t-test was performed to determine whether there are any statistically significant differences regarding the time taken to respond between the large and the small display. The data scores for both conditions are distributed normally, as assessed by Shapiro-Wilks test ($p>$ .05). The participants took longer time in front of Large display (M= 2086.44, SD= 689.76) as compared to smaller display size (M= 1896.54, SD= 502.80). The test shows that this difference was statistically significant with t (34) = 2.238, p = .032.

As Figure 24 shows, participants take longer when working on the large display compared to the small one. The task has three difficulty levels based on the types of histogram (2-bar, 4-bar and 6-bar). Figure 25 shows the time taken to respond for each condition separately.

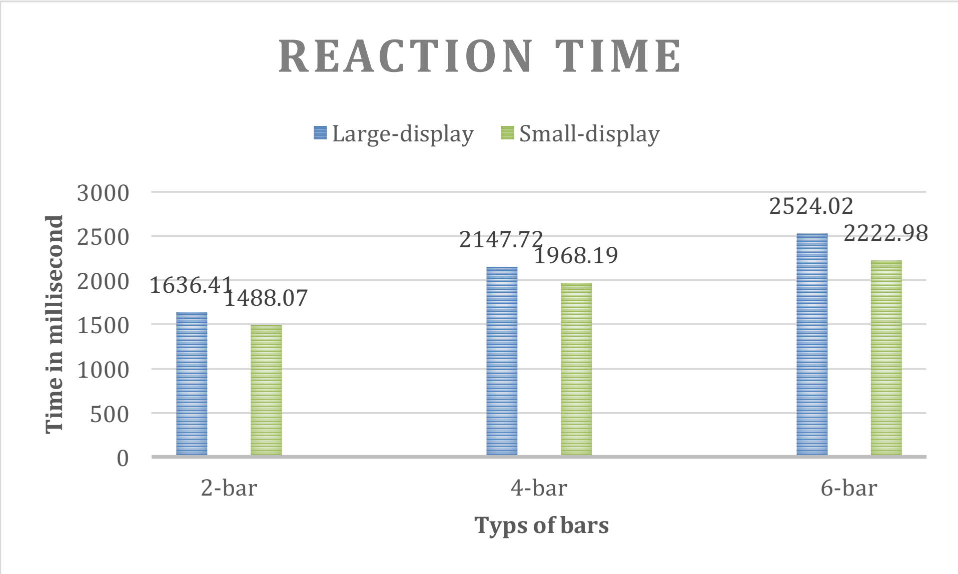


Figure 25 : Representation of reaction time to response for each difficulty level (based on types of histogram).

Figure 25 shows the total average time for the three types of histograms in front of both displays. 2-bar, 4-bar and 6-bar histograms are shown on the x-axis where blue coloured bars represent the data for the large display and green coloured bars represent the data for the small display. The time is shown on the y-axis as the average time taken for each type of bars.

In order to test for statistically significant differences, I performed the following tests summarized in Table 16.

Types of histogram	Normality	Name of test	Exact *p* -value	Distribution value
2-bar	Not normal (p = .020)	Wilcoxon signed-rank test	p = .302	Z = -1.032
4-bar	Normal (p = .846)	dependent t-test	p = .156	t = 1.452
6-bar	Normal (p = .196)	dependent t-test	p = .032	t = 2.240

Table 16: Result of the statistical test to measure the reaction time in 2-bar, 4-bar and in 6-bar histograms separately, Table is showing the data related to the normality, performed significant test and their respective exact p-value and distribution value.

The data shown in Table 16 indicates that there are significant differences in the time that participants took to respond to the task in front of the two display with 6-bar histograms only because its *p*-value is showing the result as $p = .032$ which is less than .05.

The histogram task also includes three types of rotation; rotation by 90-degree, 180-degree and by 270-degree. Figure 26 shows the time participants took for each type of rotation separately.

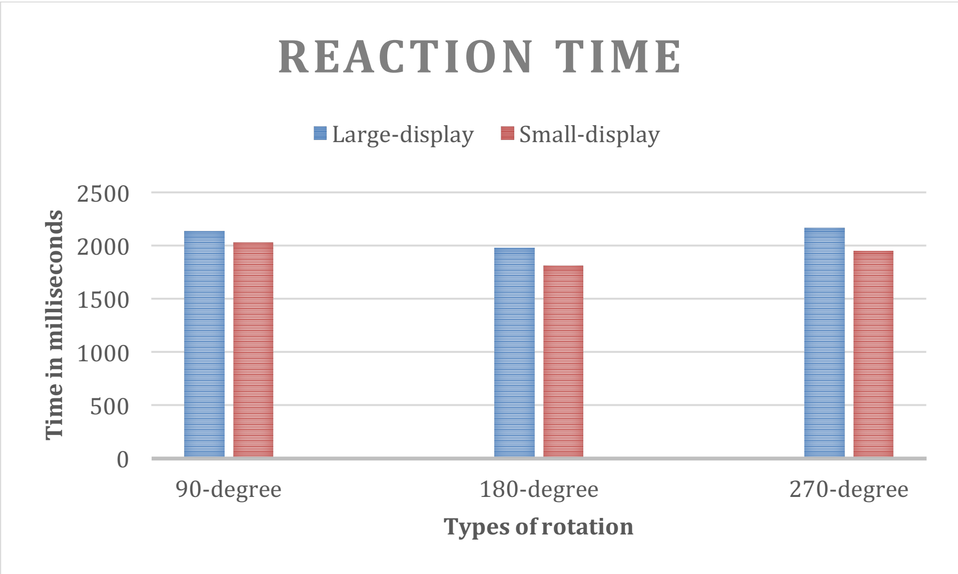


Figure 26: Representation of reaction time to response in orientation by 90-degree, 180-degree and by 270-degree.

Figure 26 shows the average time taken for one response for the three types of rotation in front of the large and the small display separately. The types of rotation are shown on the x-axis where the blue and red bars represent the data for the large and small display respectively. The time in millisecond is shown on the y-axis.

Investigating Figure 26, it seems that participants take longer responding to 270-degree rotation tasks in front of the large display. In contrast to that, it seems that they might take slightly longer for the 90-degree rotation tasks in front of the small display. To check for statistically significant differences concerning the response time based on the type of rotation the following tests shown in Table 17 have been performed.

Types of orientation	Normality	Name of test	Exact *p* - value	Distribution Value
90-degree	Not normal (p = .000)	Wilcoxon signed-rank test	p = .085	Z = -1.720
180-degree	Not normal (p = .000)	Wilcoxon signed-rank test	p = .201	Z = -1.27
270-degree	Normal (p = .385)	dependent t-test	p = .027	t = 2.304

Table 17: Result of the statistical test to measure the reaction time in three types of orientation separately, Table is showing the data related to the normality, performed significant test and their respective exact p-value and distribution value.

The results in Table 17 show that there is a significant difference in time between the large and the small table for the 270-degree rotation task.

8.2.2 Measure- Accuracy

Accuracy is measured by counting the number of right and wrong attempts when deciding if two histograms are congruent or in-congruent. Figure 27 shows the percentage of right trials over all histograms for the large and the small display.

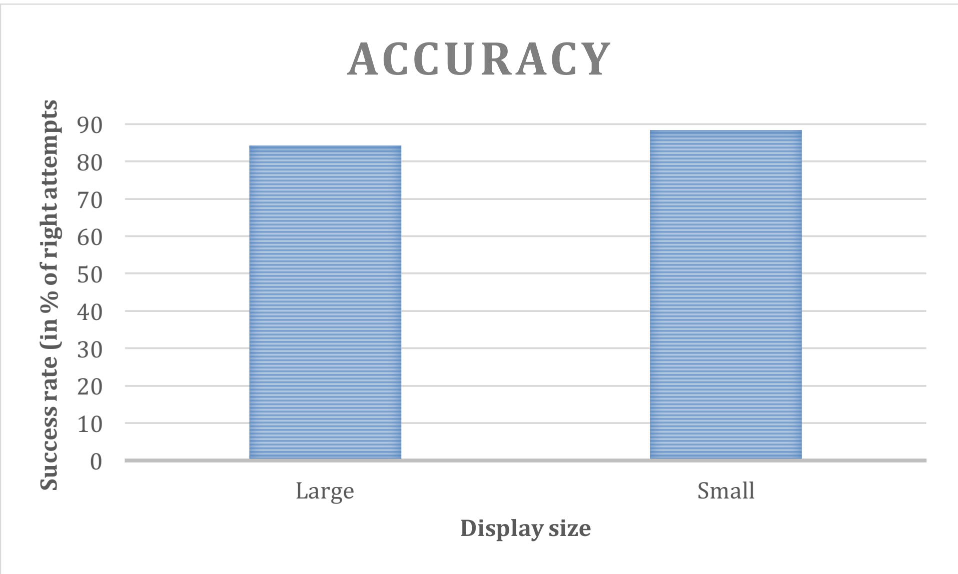


Figure 27: Representation of accuracy (success rate measured in percentage of right attempts) in front of large and small display.

A Wilcoxon signed-rank test was run to test for statistically significant differences between the accuracy for the large and the small display. The test showed that there is not significant difference in accuracy in front of both the display size with Z = -1.431 and p = .152.

As previously discussed, the *Histogram Rotation task* consists of three types of histograms; 2-bar histograms, 4-bar histograms and 6-bar histograms. Figure 28 shows accuracy for each of the three types of histograms separately.

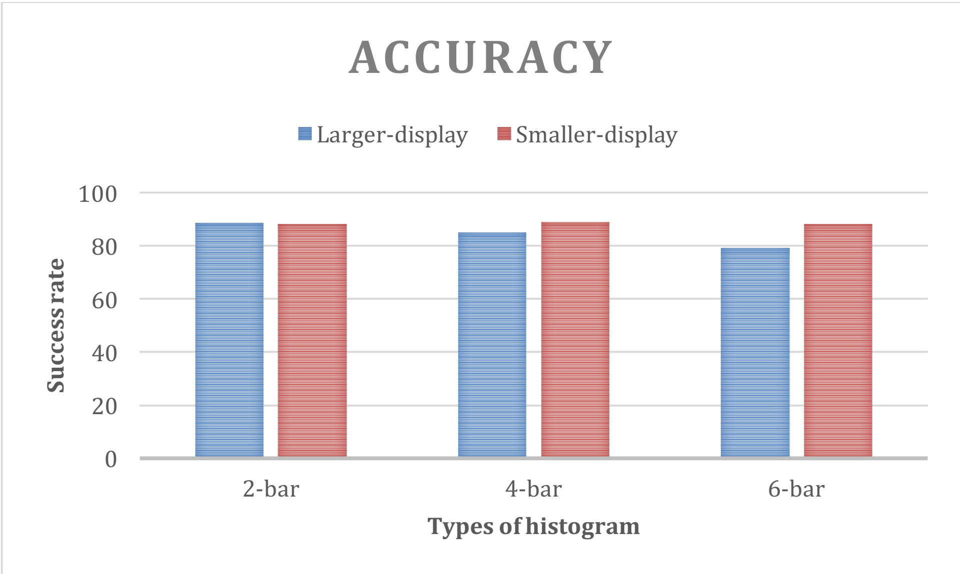


Figure 28: Representation of accuracy with three types of histograms in front of large and small display.

Types of histograms (2-bar, 4-bar and 6-bar) and success rate are mentioned on x-axis and on y-axis respectively in Figure 28. On the x-axis, the blue colored bars represent the data collected when working on the large display whereas the red colored bars represent the data collected when working with the small display. As Figure 28 shows, the success rate (accuracy) is very similar between the two displays concerning the 2-bar histograms which is not the case with 4-bar and 6-bar histograms.

To see whether there is any statistically significant difference in accuracy based on the type of histogram, Table 18 summarizes the statistical tests that have been performed on the data.

Types of histogram	Normality	Name of test	Exact *p* - value	Distribution Value
2-bar	Not normal (p = .000)	Wilcoxon signed-rank test	p = .489	Z = -.692
4-bar	Normal (p = .069)	dependent t-test	p = .263	t = -1.138
6-bar	Not normal (p = .007)	Wilcoxon signed-rank test	p = .034	Z = -2.114

Table 18: Result of the statistical test to measure the accuracy in three types of histograms separately, Table is showing the data related to the normality, performed significant test and their respective exact p-value and distribution value.

As the results reported in Table 18 show, there is a statistically significant difference between the two display sizes in accuracy when participants worked with 6-bar histograms. No significant difference in accuracy is found with 2-bar and 4-bar histograms.

The histograms also support three types of rotation. To see the difference in accuracy based on the type of rotation, I have plotted the data in a graph (see Figure 29).

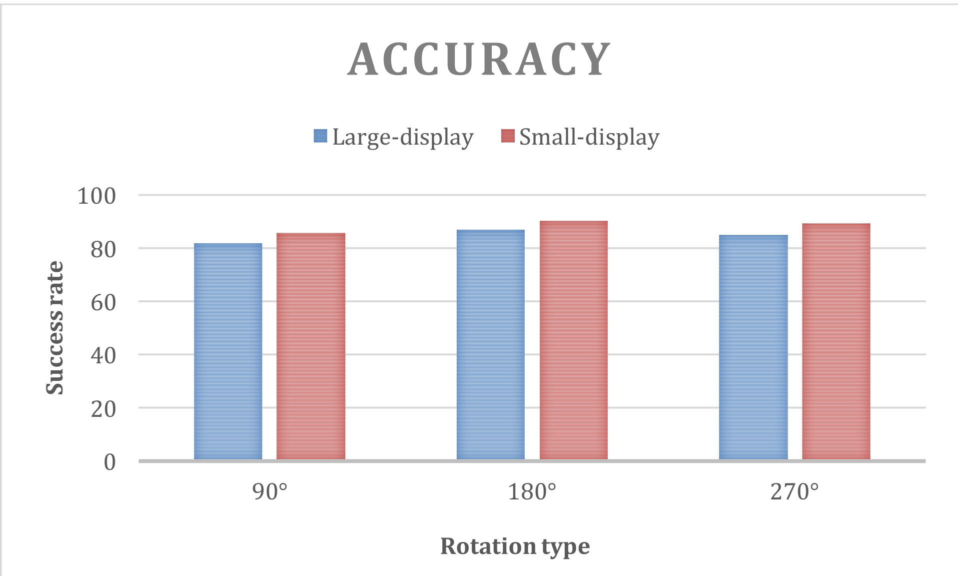


Figure 29: Representation of accuracy with three types of orientation in histograms in front of both the display size.

Figure 29 represents the success rate with respect to the type of rotation in front of both displays. The x- axis lists the three types of rotation, i.e. rotation by 90-degree, 180-degree and by 270-degree, whereas the y-axis plots the percentage of right responses while deciding between congruency and in/congruency. The blue bars and red bars represent the data for the large and small display respectively.

In order to test whether there are any statistically significant differences between the large and the small display concerning the different types of orientation of the histograms, I have performed some statistical test reported in Table 19.

Types of orientation	**Normality**	**Name of test**	**Exact *p* - value**	**Distribution Value**
90-degree	Not normal (p = .020)	Wilcoxon signed-ran test	p = .657	Z = -.444
180-degree	Normal (p = .205)	dependent t-test	p = .343	t = -.962
270-degree	Normal (p = .278)	dependent t-test	p = .150	t = -1.474

Table 19: Result of the statistical test to measure the accuracy in three types of orientation in histograms separately, Table is showing the data related to the normality, performed significant test and their respective exact p-value and distribution value.

Results of the statistical analysis shows that there are no statistically significant differences between the large and the small display when comparing specific types of rotation of histograms.

8.3 Navigation task

Again, the independent variable for this task is the display size. As dependent variables, I measured the path length and the time taken to navigate to the items. In the following, I will discuss them separately.

8.3.1 Measure – Path length

Path length is the path that participants have taken to navigate to the items to be searched on the canvas divided by the shortest possible distance between the start position (center of the canvas) and the respective item. Thus, the optimal path results in a score of 1. The average scores across all navigation trials for the small and the large display are shown in Figure 30. For each search, the actual path length is divided through the optimal path length, and the measurement "covered distance" shows the result of this calculation

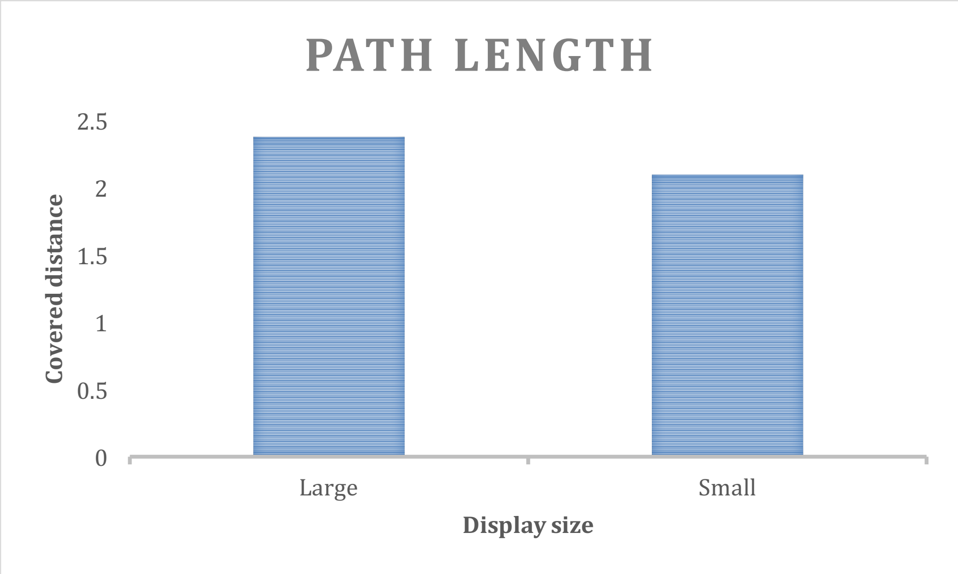


Figure 30: Representation of path length in front of two display size.

Figure 30 shows the scores of actual paths length in relation to the shortest possible path length for the small and the large screen separately. The display size is represented on the x-axis and navigation score is represented on the y-axis.

A dependent t-test was performed to test for statistically significant differences in path length between the large and the small display. The data for both displays is distributed normally, as assessed by the Shapiro-Wilks test ($p > .05$). The analyzed result (path length) in front of large display is (M = 2.376, SD = .927) and in front of small display is (M = 2.098, SD = .762). The test shows no significant differences in path length between the large and the small display.

In the task, participants have to follow 36 navigation trials (6 items and 6 repetitions). This means, that each item is navigated to a total of six times and there are six path lengths for each of the items during the whole task. Figure 31 shows the path length for each of the runs/repetitions/ separately. For each search, the actual path length is divided through the optimal path length, and the measurement “covered distance” shows the result of this calculation.

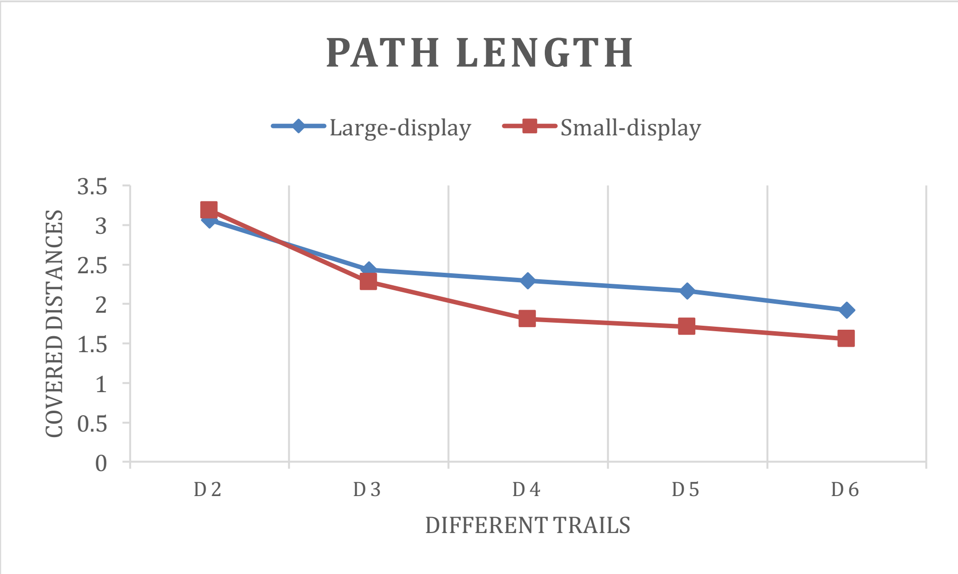


Figure 31: Representation of path length (D2-D6) in front of large and small display.

Figure 31 shows the covered distance in front of both the displays where the blue and red lines represent the covered distance for the large and small display respectively. The x- axis divides the data based on the number of times that the item had to be searched (D2-D6) and thus represent the different repetitive navigation trials. The first navigation trail (D1) is not considered because at the beginning, the participant is unfamiliar about the position of the item on the canvas and the result represent the search accuracy and not the navigation performance based on spatial memory.

As expected, Figure 31 shows that the covered distances gradually decrease with the increasing number of navigation trails. For both displays, the covered distance is highest in the first trail (D2) and lowest in the last trail (D6). As can be seen in Figure 31, the learning curve is steeper when working with the small display, especially in the first few repetitions.

To test for significant differences in each navigation trail (D2-D6), some statistical tests have performed whose results are summarized in Table 20.

Different distances	**Normality**	**Name of test**	**Exact *p* -value**	**Distribution Value**
D2	Normal (p = .771)	dependent t-test	p = .766	t = -.300
D3	Normal (p = .819)	dependent t-test	p = .591	t = .543
D4	Normal (p = .215)	dependent t-test	p = .070	t = 1.872
D5	Not normal (p = .040)	Wilcoxon signed-rank test	p = .112	Z = -1.589
D6	Not normal (p = .000)	Wilcoxon signed-rank test	p = .351	Z = -.934

Table 20: Result of the statistical test to measure the path length separately (D2-D6), Table is showing the data related to the normality, performed significant test and their respective exact p-value and distribution value.

Table 20 shows, there are no statistically significant differences between the path lengths for the large and the small screen in the separate navigation trails (D2-D6).

8.3.2 Measure – Time

This section analyzes the time that participants took to navigate to the items. Figure 32 shows the average time that participants took to navigate to an item when working with the large and the small display.

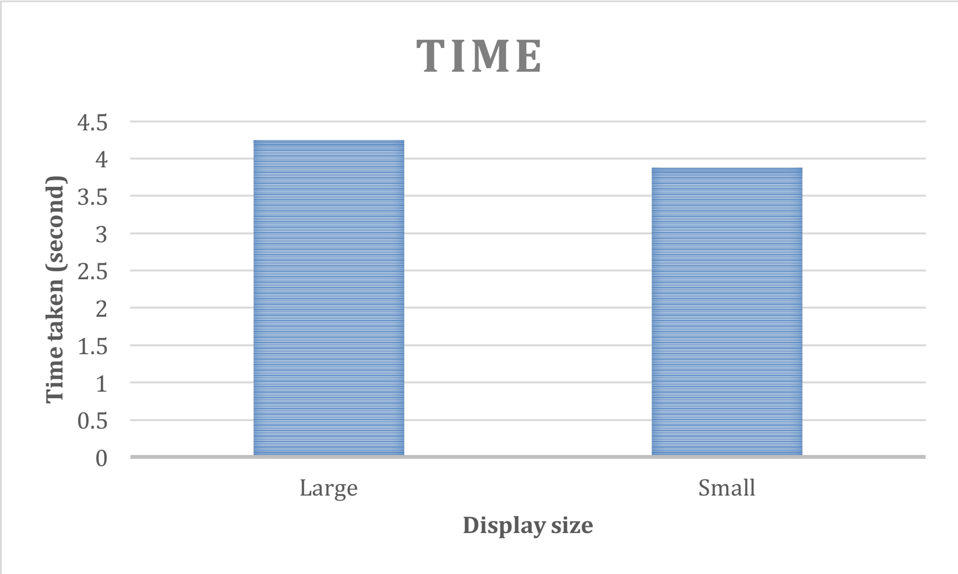


Figure 32: Representation of time (second) in front of two displays.

In Figure 32, the display size is displayed on the x-axis and the time taken is plotted on the y-axis.

A Wilcoxon signed-rank test was run to test for statistically significant differences in the time between the two displays. The data for both displays is not normally distributed, as assessed by Shapiro-Wilks test ($P<.05$). Obtained values are indicating that participants have taken time in front of large display is ($M = 4.246$, $SD = .2.304$) and time taken in front of small screen is ($M = 3.871$, $SD = 1.999$). The test showed that there is no significant difference in time taken in front of both the display size with $Z = -.932$ and $p = .351$.

As mentioned above that there are total 36 trials to navigate the 6 items over the canvas. Therefore, I analyzed the separate times i.e. T1, T2, T3, T4, T5 and T6 for each of the repetitions. Figure 33 shows the differences in time taken to perform the task in front of both displays. The times for each repetition (e.g. T2, T3, T4, T5 and T6) are presented on the x-axis and the total time taken to navigate to the items is plotted on the y-axis. T2 is representing the average time taken in the second navigation trial of all the items and by all the participants. Same is applicable for T3, T4, T5 and T6. Time T1 is not considered because of above mentioned reason (same as D1). The blue and green lines represent the time taken in front of the small and the large display size respectively.

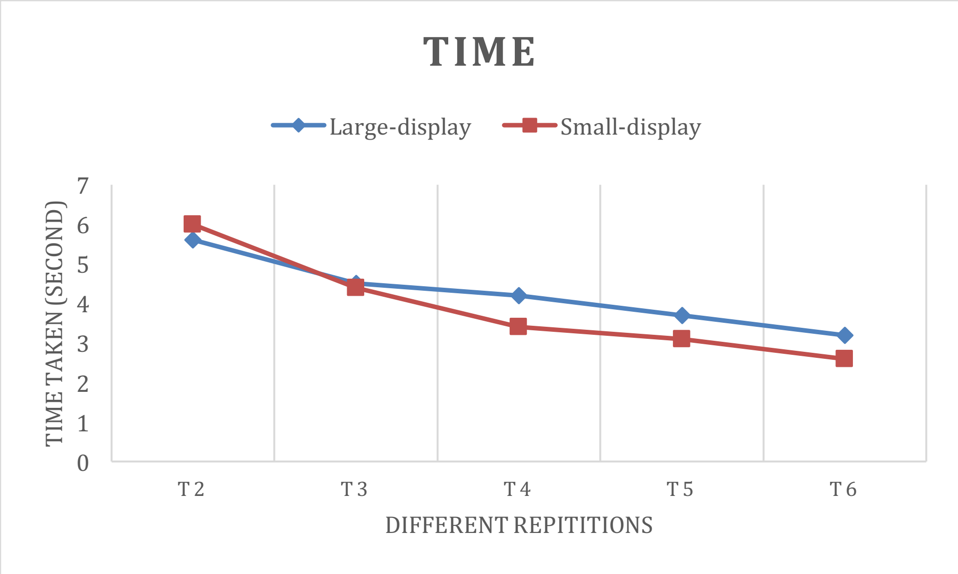


Figure 33: Representation of time in different repetitions (T2-T6) on small and large display.

As with the path length, Figure 33 shows that the time taken to navigate to an item is decreasing with the increasing number of repetitions. In other words, T2 has higher value than T3, T4, T5 and T6. Figure 33 also shows that the learning curves steeper when working with the small display which indicates that learning is better with time in front of smaller display size.

To test for significant differences in different times (T2-T6) in each navigation trail, some statistical tests have performed and reported in Table 21.

Different time	Normality	Name of test	Exact *p* - value	Distribution Value
T2	Normal (p = .785)	dependent t-test	p = .510	t = -.665
T3	Not normal (p = .007)	Wilcoxon signed-rank test	p = .578	Z = .557
T4	Not normal (p = .005)	Wilcoxon signed-rank test	p = .092	Z = -1.685
T5	Not normal (p = .001)	Wilcoxon signed-rank test	p = .070	Z = -1.815
T6	Not normal (p = .000)	Wilcoxon signed-rank test	p = .280	Z = -1.081

Table 21: Result of the statistical test to measure the time separately from T2-T6, Table is showing the data related to the normality, performed significant test and their respective exact p-value and distribution value.

The result from Table 21 indicate that there is no statistically significant difference between the large and the small display concerning T2, T3, T4, T5 and T6.

8.4 Tangram task

For the Tangram task, time and the strategies followed to solve the task are the dependent variables. The analysis of both the dependent variables is discussed separately below.

8.4.1 Measure – Time

For the Tangram Task, the total time taken to complete the task is measured. Figure 34 shows the average time participants needed to finish the Tangram Task in front of the small and the large display.

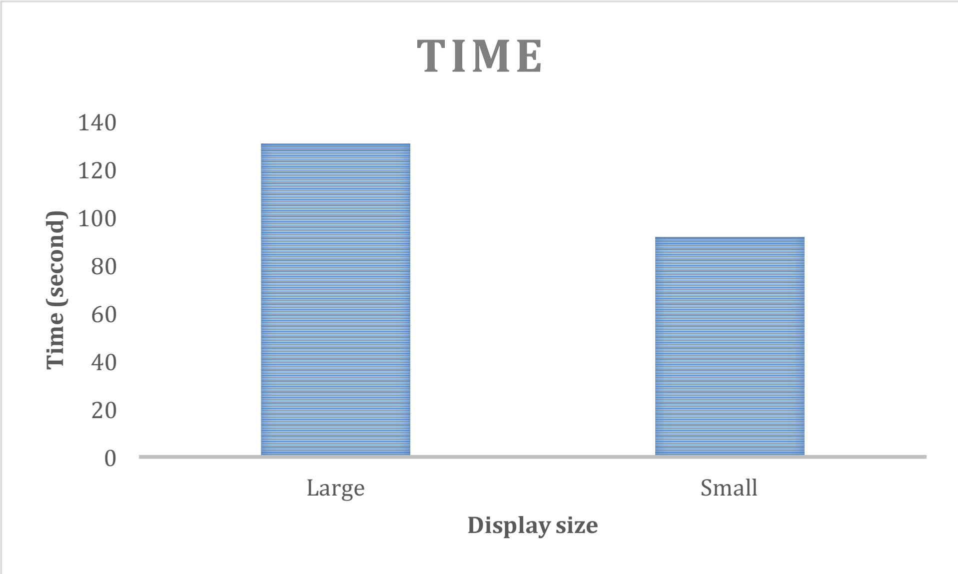


Figure 34: Representation of average time taken to finish the task on large and small display.

Figure 34 shows the display size on the x-axis and the time in second is mentioned on the y-axis. The time is shown as the average time taken to assemble all the three shapes in front of one display type. To analyze the results in detailed, a Wilcoxon signed-rank test was run to determine whether there is a statistically significant difference in the time that participants need to assemble the shapes between the small and the large display. The data is not normally distributed, as assessed by Shapiro-Wilks test ($p<.05$). The participants have taken around 130 seconds (SD= 99.135) to assemble the shape in front large display. In front of small display, it took them 91.7 (SD= 67.24) seconds to finish the assembly. The test showed that time taken was higher in front of large display size with $Z = -2.225$ and $p = .026$.

8.4.2 Strategies

The main aim of the analysis was to see the performance difference in front of large and small display while choosing the strategies to complete the task and strategies followed to deal the challenges faced during the task completion.

There was not any influence of display size was observed in the strategies selection to complete the task and to overcome the challenges in task completion. The most commonly used strategies to complete the task that is followed by most of the participants in front of both display size is mentioned below.

- Participants picked one shape from its source position, tried to put on different places on target image (by sliding or by rotating the shapes over target) but at last kept back the shape into its source. By this strategy, participant was taking the overview of the target image.

- Participants had chosen one position on the target image, picking the shapes randomly from the source position and trying to fit into target image on the selected position.
- Firstly, putting all the shapes randomly anywhere on the target image and later adjusting the shapes on proper position.
- To find the exact location of the shape on the target image, firstly chosen one shape and tried to make this shape overlap with each corner and edges of target image.
- Firstly, covering the extreme points of the target image (top most, bottom most, left and right most), later focusing on to fill middle shapes.
- Tried to assemble the shapes in a straight line (from left most to right most).

In the same manner, following mentioned strategies are used by most of the participants when after some point, they found the task difficult to complete:

- Tried to complete the half-assembled target image by moving here and there assembled shapes over the target. At this moment, participant is not selecting the new shape from source only re-adjusting the existing shapes on the target image.
- When he/she was stuck at a point then he has un-assembled all assembled shapes of the target image and then starting again to re-assemble the same shape.
- Throwing out assembled bigger shapes to the source position and trying to assemble the smaller shapes first and at last assembling the bigger shapes.
- After struggling few minutes with few of the shapes, started again from the beginning (empty target image and all the seven shapes on source).
- Quit the task.

8.5 NASA/TLX

In order to analyse whether participants rated the task load for the separate tasks differently when working on the small or the large screen, I plotted for each task the average NASA/TLX scores for each of the two display conditions. Then, statistical tests were performed in order to check for significant differences. The following sections summarize the analysed results for each task separately.

8.5.1 Spatial Span task

Figure 35 shows the task load for the spatial span task in front of the small and the large display.

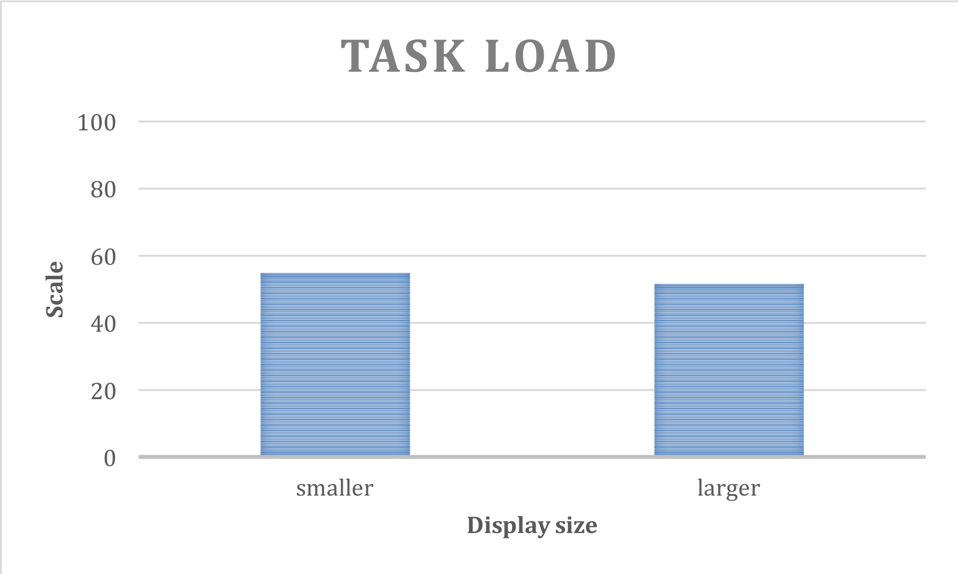


Figure 35: Representation of task load, obtained from the analysis of NASA/TLX for Spatial Span task on large and small display..

In order to test whether there are any statistically significant differences in the NASA/TLX score for the spatial span task between the two display conditions, a Wilcoxon signed-rank test was run. Data scores for larger and smaller display size are not normally distributed, as assessed by Shapiro-Wilks test (p<.05). Data represents that participant's performance with scores (M = 51.57, SD = 16.74) in front of large display and with scores (M = 54.80, SD = 16.37) in front of small display. The test showed that there is no significant difference in task load between the two display sizes with Z = -1.599 and p = .110.

8.5.2 Histogram Rotation task

Figure 36 shows the NASA/TLX scores for the Histogram task working with the small and the large display.

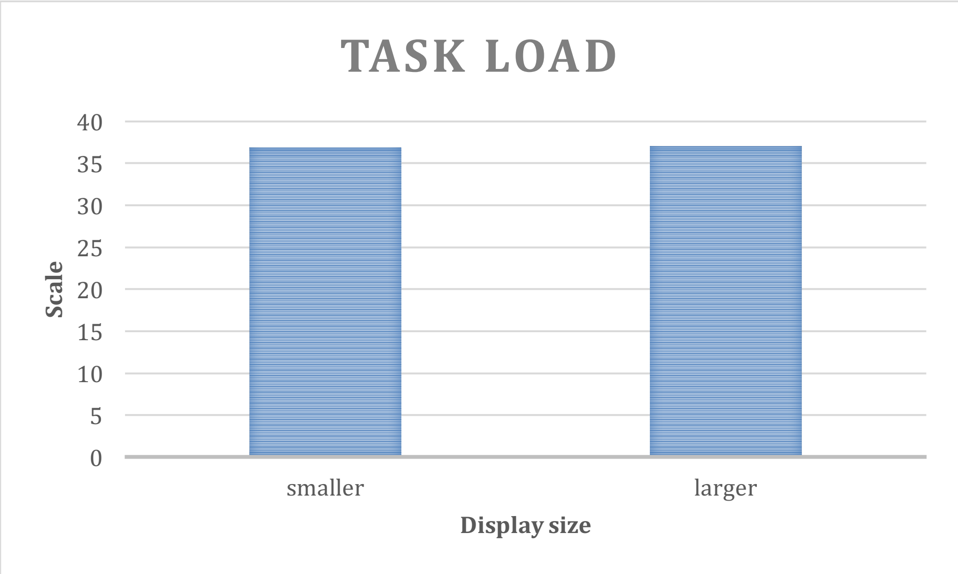


Figure 36: Representation of task load in front of large and small display for Histogram Rotation task.

In order to test whether there are any statistically significant differences between the two display conditions, a dependent t-test was performed. The data for both display sizes is distributed normally, as assessed by Shapiro-Wilks test ($p > .05$). Score of task load was recorded (M = 37.04, SD = 16.98) on large display whereas it was observed (M = 36.90, SD = 16.14) on small display size. The test shows no significant differences in task load performing the Histogram Rotation task between the two display sizes with t (34) = .050, p = .960.

8.5.3 Navigation task

Figure 37 shows the task load for the Navigation task when working on the small and the large display.

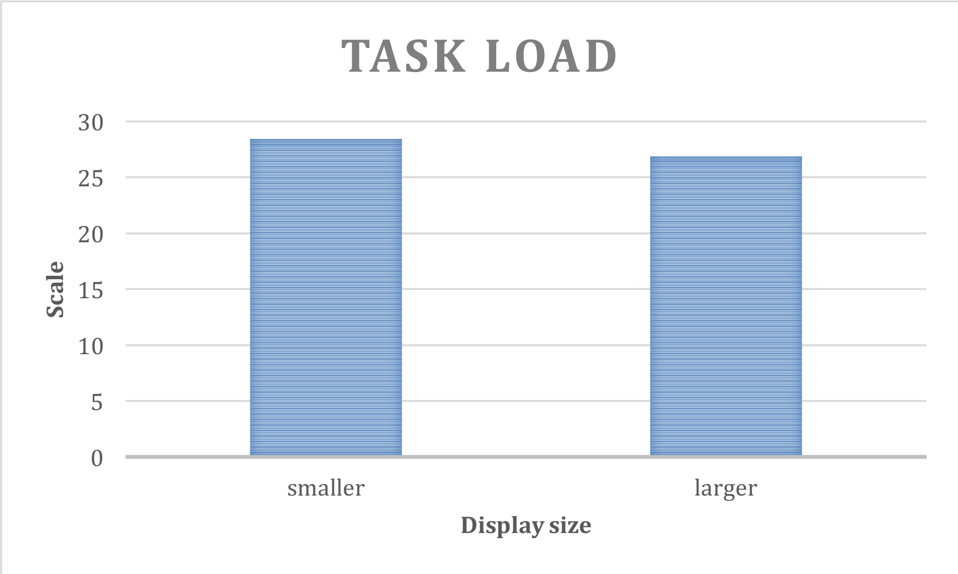


Figure 37: Representation of task load for Navigation task in front of two display size.

A dependent t-test was performed to determine whether there is a significant difference in the task load in front of the large and the small display. The data for both display sizes is distributed normally, as assessed by Shapiro-Wilks test ($p > .05$). The participants have (M = 28.41, SD = 11.10) task load scores in front of small display and (M = 26.87, SD = 12.40) task load scores in front of large display. The test shows no significant differences in task load in front of both the display size with t (34) = -.767 and $p = .449$.

8.5.4 Tangram task

Figure 38 shows the average NASA/TLX score for the Tangram task in front of the small and the large display.

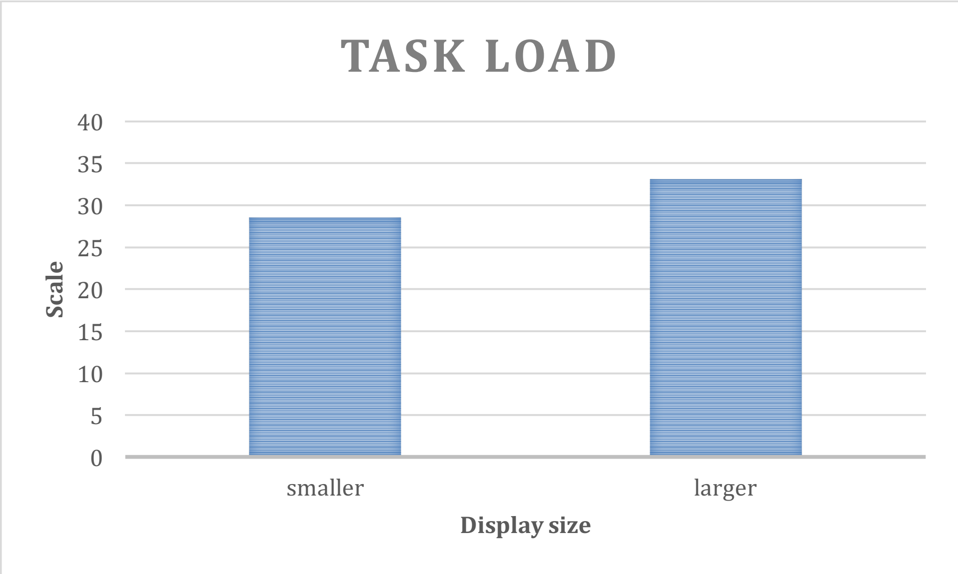


Figure 38:Representation of average task load scores for Tangram task in front of two displays.

To check whether there is significant difference in the NASA/TLX score between the two display sizes, a dependent t-test was performed to determine the significant difference in task load in front of both the display size. The data scores for both the display size are distributed normally, as assessed by Shapiro-Wilks test ($p > .05$). The average task load on large display is (M = 33.04, SD = 16.86) and on small display is (M = 28.47, SD = 16.94). The test shows no significant differences in task load in front of both displays with t (34) = 1.423 and $p = .164$.

8.6 Post-questionnaire

In the post-questionnaire, participants were asked to answer three questions with regard to their personal preference. When asked which display size they preferred, 20 participants stated that they preferred the large display, 10 participants preferred the small display and 5 participants did not have a preference.

Table 22 lists the few mentioned reasons that participants mentioned to explain their preference of the large display. Identity of participants are presented as [P 3] (third participant).

SN.	Reason
1	More field of vision. [P4, P22]
2	Large display keeps fully engaged. [P25]
3	Tasks are easy to follow in front of Large display size. [P16, P7]
4	Feeling of able to see more and clear. [P9]
5	Easy to see and follow the tasks. [P18]
6	Easy to follow and recognize. [P23]
7	Easy to complete the task. [P33]
8	Complete view and better perception of space. [P1]
9	Comfortable watching in front of Large display. [P27]
10	More clear & bigger picture so easy to concentrate. [P20]
11	Easy to estimate the size and shape. [P10]
12	Better overview and gives better understanding to perform. [P25]
13	No need to focus more. [P19]
14	Easy to remember. [P31, P12]

Table 22: List of the reasons for the preference for large display by some participants.

As an explanation for the preference of the small display, participants mentioned the statements summarized in Table 23.

SN.	Reasons
1	More natural. [P21, P34]
2	No need to move eye. [P32]
3	Feeling of yourself while in front of Large display you feel like inside the screen. [P8]
4	Less attention requires, can see everything in a glance. [P11]
5	Faster and better, no need to move head or eyes. [P35]

Table 23: List of the mentioned reasons by the few participants about their preference for small display.

The second question of the post-questionnaire was concerned with the difficulty level of the tasks. Participants were asked to sort the task in the order of their difficulty (1 = easiest and 4 = toughest). The analysed result from the collected data is represented in Figure 39.

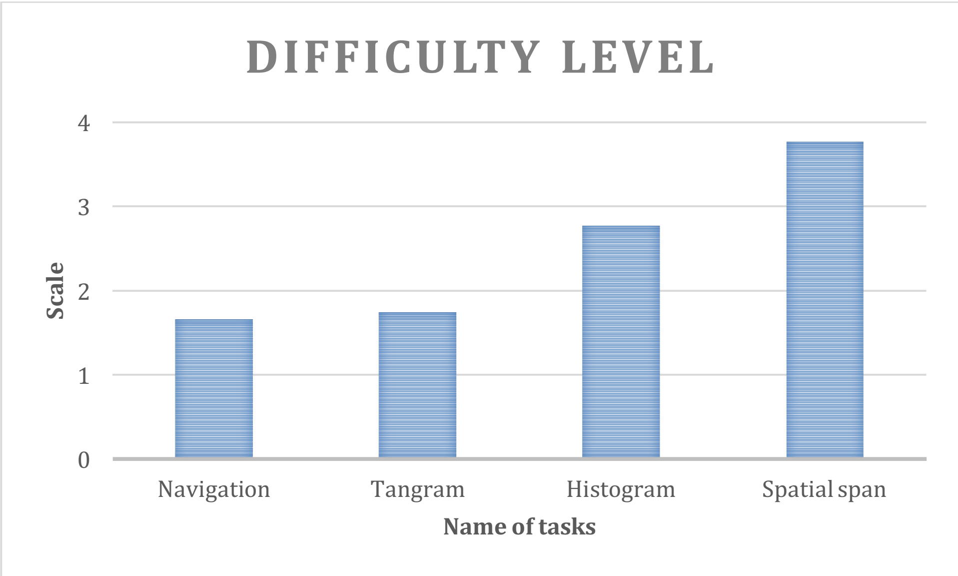


Figure 39: Representation of different difficulty levels for each of the task

In Figure 39, the name of the task is mentioned on the x-axis whereas the average of the scale (ranges between 1-4) is represented on the y-axis. Figure 39 shows that participants rated the *Navigation task* and the *Tangram task* easiest task and the *Spatial Span task* as the toughest task.

The third question in the post-questionnaire was concerned with the display preference in relation to the particular tasks

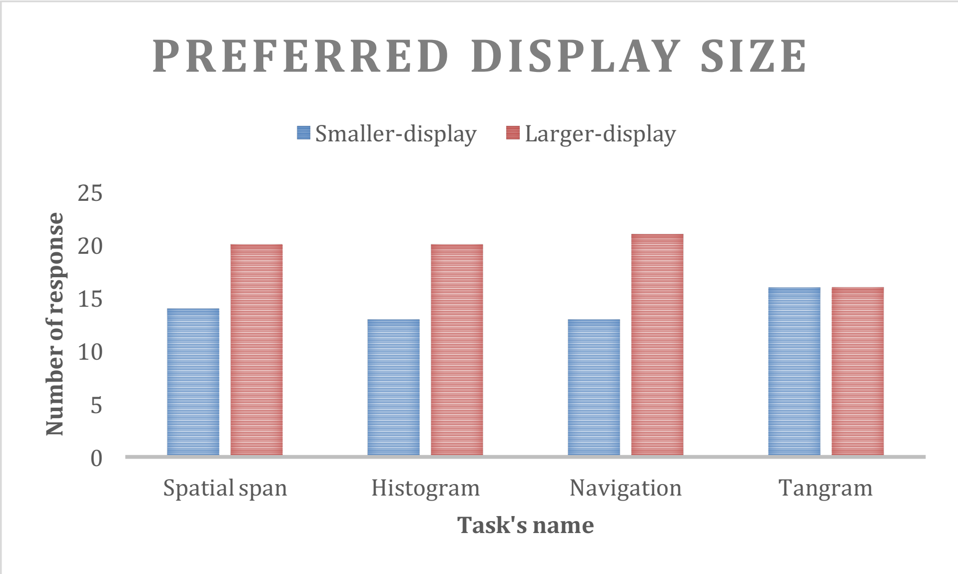


Figure 40: Representation of preference for display size in relation to particular task.

Figure 40 shows participants' preferences for the display size in relation to the particular tasks. The names of the all four tasks are listed on the x-axis and the number of responses (how many participants prefer to perform it in front of the large and the small display) are shown on the y-axis. On the x-axis, blue and red coloured bars represent the data in front of the small and the large display respectively. From the collected data, it seems that participants prefer to perform the *Spatial Span task*, the *Histogram rotation task* and the *Navigation task* in front of Large display size where as it seems there is equal preference of display size for the *Tangram task*. One participant for the *Spatial Span task*, two participants for the *Histogram Rotation task* and *Navigation task* and three participants for the *Tangram task* mentioned that they didn't have any preference of the display size to perform the tasks.

Results of each task is summarized in this chapter. All the information related to participants' performance, their work load and their preferences of display size to perform the tasks are mentioned here. In the next chapter, I discuss these results to determine the answers for specified research question.

9 Discussion

In the previous chapter, the experiment's results were elaborated on. This chapter discusses the presented data concerning participant's performance in front of the large and the small display with respect to each task, in order to determine answers for the specified research question and research objectives. A brief summary of the participant's performance is shown in table 24 listing the statistical findings of the different measures with their respective task and aspect of spatial memory.

SN.	**Aspects of spatial memory**	**Tasks name**	**Measures**	**Statistical results**
1	*Visual Recall*	Spatial Span task	Time	No significant difference
			Error	No significant difference
2	*Rotation*	Histogram rotation task	Reaction time	Participants were significantly faster on small display with difficult histogram (6-bar and 270-degree rotation.
			Accuracy	Participants were significantly more accurate for difficult tasks (6-bar) on large display.
3	*Navigation & Recall*	Navigation task	Path length	No significant difference
			Time taken	No significant difference
4	*Assembling/ Disassembling*	Tangram task	Time	Participants were significantly faster on small display.
			Strategies	No difference.

Table 24: An overview of the findings of the experiment showing the statistical result of each task with their respective measures and aspects.

The goal of this thesis is to investigate the influence of display size on aspects of spatial memory. I formulated four research objectives. In the following sections, I will discuss each research objective separately, and discuss how the results address the overall research question.

9.1 Research objective 1

Does display size influence the *Visual Recall* (aspect of) spatial memory?

The first research objective is addressed by the analysis and results of the *Spatial Span task* (see section 8.1). Time and error are the two dependent variables that were measured in the task. The time was measured by calculating the time taken by all the participants from the flash of the last block until the first click on a block. Then the next time is taken between this click and the next one etc. until the participant clicked on the required number of blocks. Error is measured in terms of a success rate and this success rate is measured by counting the number of right responses (in percent) for each of the blocks (B2-B9)) in front of both displays.

As a result, it is observed that the average response time increases with the increasing difficulty level in front of both the display size where the increment in difficulty levels is represented by an increasing number of blocks that participants had to click on. Therefore, this increase in time was at least partly due to the higher number of clicks that participants had to perform. When comparing the overall response time for the small and the large display, the results show no significant difference between the two display sizes. Although, there was not any significant difference in overall response time on small and large display between the different number of blocks that participant had to respond to, figure 21 shows that the response time slightly increased with the increasing number of blocks.

In addition, when comparing the success rate between the large and the small display size, no significant difference was found between the two displays. Although, there was no statistical differences in error between the different number of blocks that participants had to respond to, Figure 23 indicates that participants make more errors with increasing difficulty level in the task in front of both the display. I guess the reason for this type of participant's performance is that the more block they have to remember, the more difficult it gets and therefore the more mistakes they make.

In terms of task load (from NASA/TLX), no any significant difference is measured to perform the task in front of large and small display size.In short, there is not any significant difference is identified in response time and in error in front of both the display size.

A few studies have been done with the traditional version (*Corsi blocl-tapping task*) of this task (Berch, Krikorian et al. 1998) (Teixeira, Zachi et al. 2011) with the aim to investigate the affect of age, gender and educational level on the performance of *Corsi blocl-tapping task* (traditional version of *Spatial Span task*)where Berch et al. (Berch, Krikorian et al. 1998) has measured the attention and working memory by counting the recall of sequences. Results show no gender difference in performance and they also show that children who are more than or equal to six-year-old and go to school are better in recalling the sequences than who are less than six years and don't go to school. In addition to this task, Riccardo et al. (Brunetti, Del Gatto et al. 2014) have performed the experiment on the digitized version (eCorsi) of the task. In their experiment, they have compared types of span (forward-span, backward-span and supra-span) and measured the span length, correctly recalled sequences and faster first tapping latency. As a result, they show that the forward span condition has an advantage over the other two spans in terms of higher span, more number of correctly recalled sequences and faster First Tapping Latency in correct sequences.

After going through the previous studies, I elicited that researchers have performed many studies based on traditional and digitized version of *Spatial Span tasks* where they are focused to measure correctly recall sequences of taps, span length, time and gender-age difference in performance and many more but I could not find any study where they have performed this type of task (who reflects *Visual Recall* aspect of spatial memory*)* to see the influence of different display size.

Miller (Miller 1956) has investigated if there is a span of 7 digits or seven sequences then neighborhood of span 7 (plus or minus 2 in 7 that is either 5 or 9) puts some limits on our capacity to process the information. In my *Spatial Span task*, there is a span of 9 blocks (B2-B9) and from concluded result of this study, error (success rate) was observed higher after the level B6 to level B9 in the task where participants were instructed to click on the seven, eight and nine blocks respectively. The reason for this type of performance is following the logic given by Miller (Miller 1956) that brain limits the capacity to process the information in the neighborhood of span of 7. In my task, the span length of the task was 9 so the error (success rate) was observed more between level B6 to B9 compared to previous levels (B2, B3, B4, B5).

One problem that was observed during the analysis, with the increasing difficulty level sometimes participants do not click on the required number of blocks (if they are expected to click on the 6 blocks in B6, they are clicking on either less than or more than six blocks). The reason for this may be that as the task is being tougher and being hard to recall the block sequences, participants are losing their grip and interest in performing the task accurately. These types of attempts are counted as inaccurate which resulted in a higher number of errors. This may be the one of the reason which might influence the measured error (success rate).

It can be concluded form the analysed results that there is no influence of display size on the *Visual Recall* (aspects of) spatial memory.

9.2 Research objective 2

Does display size influence the *Rotation* (aspect of) spatial memory?

The *Rotation* aspect of spatial memory was reflected by the Histogram rotation task. Thus, the second research objective can be addressed by discussing the results of this task. Reaction time and accuracy is measured as a dependent variable in the Histogram Rotation task. Reaction time is measured by calculating the time taken to respond for congruency and in-congruency between the histograms. The calculation of this reaction time starts from showing the second histogram until the response is received from the participant. Accuracy is measured by counting the right and wrong attempts during the decision for congruency and in-congruency between the histograms.

The result of the *Histogram Rotation task* indicates that participants perform significantly better in front of the small display. They have responded faster when working with the smalldisplay compared to the large display. In addition, participants responded significantly faster in front of the small display when working with 6-bar histograms specifically. Also, they were significant faster on the small display compared to the large one when working with histograms that were rotated by 270-degree. This shows that the advantage of the small display is particularly important when working on difficult histograms (6-bars, 270-degree orientation).

Concerning the accuracy of responses, I observed significant difference between the large and small display only when participants worked with 6-bar histograms. In these cases, participants made significantly more errors while performing on the small display. Even though they were significant faster with the small display, they made significantly more errors when the task was difficult. Thus, it cannot be concluded overall that smaller was better but participants were faster on small display and in terms of accuracy large display size was better for difficult tasks.

Findings of this task are partly in line with previous research. For example, Tan et al. (Tan, Gergle et al. 2006) have performed the experiments on *Rotation* aspects of spatial memory and have considered the display size as an independent variable in their study. They have shown that participants perform 26% better in mental rotation tasks in front of large displays. In my case, participants were better with the large display, but only for difficult histograms. At the same time, they were quicker with the small display. So partly my findings are similar than Tan.

Tan and his co-workers (Tan, Gergle et al. 2006) has chosen one projector (76" wide and 57" tall) as a large display and one desktop (18") as a small display for their experiment and shown the combination of three tasks (card test, cube test and Shepard Metzler test) based on mental rotation. In my experiment, I have performed a *Histogram Rotation task* on Microsoft perspective pixel (55") as a large display and on Microsoft surface 2 pro (10.6") as small display. Also, distances between user and both the displays & the eye-height form the ground was different in my and Tan's experimental set-up (because of different display sizes used and to maintain the visual angle constant). As I have told previously, that my lab set-up work was inspired by the Tan's work but it was bit different in terms of dimensions of display size and the distances between the two displays and the distances between two displays and user. Slight difference in this lab set-up may be the reason of this difference in findings that leads to result that partly my findings are similar than Tan. Participants were faster but less accurate with the small display, the reason may be that they found it easy to response faster on small display because it allows us to see everything in a glance and less attention is required to solve the task and leads the comparison faster between two histograms but large display gives more feeling of immersion, results better accuracy in decision making.

In terms of task load (from NASA/TLX), no any significant difference is measured to perform the *Histogram Rotation task* in front of large and small display size

At last, I can summarize that display size influence the *Rotation* aspect of spatial memory. Participants perform faster (time) on small display and in terms of accuracy, they performed better on large display for difficult tasks.

9.3 Research objective 3

Does display size influence the *Navigation & Recall* (aspect of) spatial memory?

Concerning the *Navigation & Recall* aspect of spatial memory, the path length as well as the time taken to navigate to items was measured in the *Navigation task*. The path length is the total path covered by one participant to navigate from the start position (centre of the canvas) to the respective item. In addition to the path length, the time was taken to investigate how long participants take to navigate to the items on the canvas.

In the *Navigation task,* participants have to navigate 6 items with 6 repetitions. Thus, there are 6 different path lengths (D1-D6) and 6 different times (T1-T6). The first navigation distance (D1) and its respective time (T1) is not considered because at the very first time that a participant navigated to an item, (s)he searches the canvas for the item, thus the result of this search is not based on spatial memory. There are no statistically significant differences between the path lengths for the large and the small screen in the separate navigation trails (D2-D6). Although, the line graph shows that the covered distances gradually decrease with the increasing number of navigation trails. For both displays, the covered distance is highest in the first trail (D2) and lowest in the last trail (D6) and the learning curve is steeper when working with the small display, especially in the first few repetitions.

In addition to the time, there is no statistically significant difference between the large and the small display concerning T2, T3, T4, T5 and T6. Although, the time taken to navigate to an item is decreasing with the increasing number of repetitions. In other words, T2 has higher value than T3, T4, T5 and T6. The learning curves are steeper when working with the small display which indicates that learning is better with time in front of smaller display size.

Tan et al. (Tan, Gergle et al. 2004) have investigated the performance difference and showed that participants perform 17 % better in 3D navigation tasks involving path integration in front of large display. As their experimental set-up, Tan and his co-workers have chosen one projector (76" wide and 57" tall) as a large display and one desktop (18") as a small display and performed a triangle completion task on both the display size. As a dependent variable, they have measured distance to origin error, angle-turned error and distance-moved error. In addition to this research, Tan and his co-workers (Tan, Gergle et al. 2006) have performed a mental map formation and memory task in the same experimental set-up. Users have to explore a virtual world in order to form a cognitive map of the environment. The task had two phases: in the learning phase, participants were instructed to explore and learn the environment and in recall phase, they were instructed to find to specified targets as sooner as possible. In short, after exploring the environment, participants have to build a mental map to find the specified items. As dependent variables, distance covered and time taken to find the target were considered. Result of the experiment show that users performed better on the large display in map formation and in memory task.

According to my analysed results, I could not investigate any significant difference in covered path length and time taken by participants to navigate the items. My task is reflecting to the previous studies (that I discussed above). Although, my task is 2D navigation task which is bit different from the task performed by Tan (Tan, Gergle et al. 2004). In some order, his mental map formation and memory task is reflecting *the Navigation & Recall* aspect of spatial memory and calculated measures are also almost same in my *Navigation task* and in his mental map formation and memory task. The reason for the contradictory results may be different dimension of display size and followed lab-settings or may be the difference of environment, I conducted the task in real world whereas Tan's task was performed in virtual world. In terms of task load (from NASA/TLX), no any significant difference is measured to perform the task in front of large and small display size.

It can be concluded form the analysed results that there is no influence of display size on the *Navigation & Recall* (aspects of) spatial memory.

9.4 Research objective 4

Does display size influence the *Assembling/Disassembling* (aspect of) spatial memory?

The Assembling/Disassembling aspect of spatial memory was reflected by the *Tangram task*. Thus, the last research objective can be addressed by discussing the results of this task. In the *Tangram task*, the time taken and the strategies followed to complete the task are considered as a dependent variable. Time is measured as the total time taken to assemble the multiple shapes into the target image. This is done by calculating per participant the average time (in second) in the three trails in front of each display size.

The result of the Tangram task showed that participants performed significantly faster when working on the small display. Thus, I conclude that the display size influence the *Assembling/Disassembling* aspect of spatial memory, as a small display increases the performance (speed) in which participants finished the task. Overall, participants found it easier to assemble (move, drag, rotate) the shapes in front of small display resulting in a faster completion of the task on the small display. This finding contradicts existing studies investigating the influence of display size on spatial memory. For example, Tan et al. (Tan, Gergle et al. 2006) have conducted experiments based on spatial tasks where they have chosen display size as an independent variable. Their results show that participants performed better in front of large display size. Similarly, Simmons et al. (Simmons 2001) have performed a series of productive tasks (to find out miss-spell words among words or to browse some links to answer specified questions and many more) in front of four display size; 21", 15", 17", and 19" display sizes. Their findings also claimed that users perform better in front of large display during productivity tasks. However, neither of these studies included a task that addressed the assembling or dissembling aspect of spatial memory. After going through previous research, I investigated that nobody has investigated the influence of display size on the performance of such a type of task (addressing the *Assembling/Disassembling* aspect). The contradicting findings might therefore be due to the specific characteristics of the assembling/disassembling tasks that was used in my study.

Regarding the strategies followed, there was not any significant difference is found in front of large and small display. I could not recognise any research based on the measurement of strategies followed in the field of display size and aspects of spatial memory. In terms of task load (from NASA/TLX), no any significant difference is measured in *Tangram task to* perform the task in front of large and small display size.

As a summary, I conclude that display size influence the *Assembling/Disassembling* aspect of spatial memory and participants perform faster (time) in front of small display while attempting assembling or disassembling of shapes, types of tasks.

As an answer to my research question, it is not easy to say whether display size influences spatial memory in general, because it depends on what aspect of spatial memory we talk about. If we talk about *Rotation* and *Assembling/Disassembling* aspect, then indeed display size influences spatial memory. If we talk about *Navigation & Recall* and *Visual Recall*, then no, display size does not influence spatial memory. For the future, I therefore recommend to investigate spatial memory in a more nuanced way as HCI techniques have different kinds of influences on spatial memory, depending on what aspect of spatial memory we focus on.

10 Conclusions and Outlook

Within this thesis, I summarized the theoretical background and the related work in the field of spatial memory and display size in HCI. I discussed the experimental set-up of the planned experiment (implemented tasks and followed apparatus), and explained in detail the newer/modified version of the tasks and the complete procedure of the experiment. Based on the results, I can conclude that the display size had different influences on the four identified aspects of spatial memory, and that some of the results contradict findings from previous studies. The *Rotation* and the *Assembling/Dissembling* (aspects of) spatial memory showed significant differences in the performance depending on the display size. For these two aspects, participants performed significantly faster (time) when working on small display. Participants performed significantly more accurate on large display in the task of *Rotation* aspect whereas for the *Visual Recall* and *Navigation & Recall* aspects, no significant difference in performance between the two display sizes were identified.

In my thesis work, I investigate the influence of display on aspects of spatial memory separately. As a future work, I suggest to investigate the co-relation among these aspects, to what extent these different aspects provide similar results. It can be investigated whether participants' performance varies based on different aspects and different display sizes? It would be interesting to observe whether participants performing equally bad/well in the different task resembling different aspects and whether there is a correlation in task performance across the different aspects. The same analysis can be done with display size, for example by investigating if one participant is doing well in front of the large display while solving mental rotation task whether this results in the same type of (good) performance for all the aspects on the large display? This type of analysis can be done by analysing the collected data participant-wise with respect to all the aspects in front of both the display size. The outcomes of the analysis can contribute to research in the domain of spatial memory.

In comparison to previous studies, I have found partly contradictory results. Tan et al. (Tan, Gergle et al. 2006) have shown that participants perform better in front of large display when they perform the mental rotation tasks whereas my findings indicate that participants have performed faster in front of small display and better in front of large display for difficult histograms while performing the rotation task (related to *Rotation* aspects of spatial memory). I would suggest as future work, it would be interesting to investigate the reason of this performance difference. The reason for this discrepancy may be the different lab-settings followed by Tan (large display: LCD projector and small display: 18´´ desktop) and myself (large display: 55´´ and small display: 10.6´´) or may be due to the type of tasks (Tan has used combination of 2D and 3D tasks whereas my task is in 2D only) performed.

In my thesis, I aimed to investigate the influence of output modalities (display size) on aspects of spatial memory. As a future work, I recommend to investigate the influence of input modalities (mouse, touch, gesture) on identified aspects of spatial memory. Although, this type of work has been tackled by Zagermann et al. (Zagermann, Pfeil et al. 2017) where they have investigated the influence of display size on Navigation and object location task (reflects

Navigation & Recall aspect of spatial memory) only. It would be interesting to investigate the combined influence of input and output modalities on other identified aspects (*Assembling/Disassembling, Rotation and Visual Recall*) of spatial memory. The findings of these types of studies can contribute to investigate the effects of different levels of embodiment on participant's spatial memory and the influence on different identified aspects of spatial memory.

To work further on this field, I recommend to plan similar type of experiment in virtual environment by using any VR devices. It would direct the research to conclude the influence of display size on different aspects of spatial memory in virtual world and to compare the performance between real and virtual world.

I have elicited the previous work based on spatial memory and on few aspects of this memory. I have focused on both the direction of the HCI (presentation and interaction) where information flows between user and the system and at last chosen the presentation facet of the HCI in the form of two display size (large and small) for my research work. After going through extensive literature review, I have identified four aspects of spatial memory. Based on those four aspects, have planned an experiment with four tasks (reflects identified aspects) on large and small display size. I have analysed different measures respected to different tasks/aspects. As a result, I have summarized that Assembling/Disassembling and Rotation aspects significantly influencing the user's performance on small display.

A fundamental understanding of the influence of the physical size on different aspects of spatial memory (that relates different visual and spatial tasks) and the experimental settings can be used for performing further research based on physical display size. A set of design principles and planning for conducting/implementing the experiment can be followed by other researchers analysing the influence of large displays. I have proposed some new criteria and metrics (various aspects based on spatial and visual tasks) that evaluate the advantage of large display size, I expect that these new paradigms (influence of *Rotation* and *Assembling/Disassembling* aspects of spatial memory) will contribute to research community and system designers who are aimed the evaluation (based on spatial memory performance) on larger scale.

Bibliography

Allen, G. L. (2004). Human spatial memory: Remembering where, Psychology Press.

Baddeley, A. D., et al. (1974). "The psychology of learning and motivation." The psychology of learning and motivation **8**: 47-90.

Baddeley, A. D., et al. (2003). The handbook of memory disorders, John Wiley & Sons.

Berch, D. B., et al. (1998). "The Corsi block-tapping task: Methodological and theoretical considerations." Brain and cognition **38**(3): 317-338.

Brunetti, R., et al. (2014). "eCorsi: implementation and testing of the Corsi block-tapping task for digital tablets." Frontiers in Psychology **5**(939).

Carpenter, M. and D. R. Proffitt (2001). "Comparing viewer and array mental rotations in different planes." Memory & Cognition **29**(3): 441-448.

Darken, R. and B. Peterson (2001). Spatial Orientation, Wayfinding and Representation. Handbook of Virtual Environment Technology. K. Stanney. Mahway, NJ, Lawrence Erlbaum Associates.

Dudfield, H. J., et al. (2001). Big is better? Human factors issues of large screen displays with military command teams. 2001 People in Control. The Second International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres.

Foreman, N. and R. Gillett (1997). A handbook of spatial research paradigms and methodologies, Psychology Press.

Grabe, M. E., et al. (1999). "The role of screen size in viewer experiences of media content." Visual Communication Quarterly **6**(2): 4-9.

Guilford, J. and W. S. Zimmerman (1948). "The Guilford-Zimmerman Aptitude Survey." Journal of applied Psychology **32**(1): 24.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. Proceedings of the human factors and ergonomics society annual meeting, Sage Publications.

Hegarty, M., et al. (2006). "Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning." Intelligence **34**(2): 151-176.

Holtzblatt, K. and H. Beyer (1993). "Making customer-centered design work for teams." Commun. ACM **36**(10): 92-103.

Inquisit 4 (2015). [Computer software]. Retrieved from http://www.millisecond.com.

Jain, V. (2016). Master Seminar: Creating a task taxonomy of spatial memory in HCI.

Jain, V. (2017). Master Project: Investigating the Influence of Display Size on Aspects of Spatial Memory.

Jetter, H.-C., et al. (2012). Does (multi-)touch aid users' spatial memory and navigation in 'panning' and in 'zooming & panning' UIs? Proceedings of the International Working Conference on Advanced Visual Interfaces. Capri Island, Italy, ACM**:** 83-90.

Klinkhammer, D., et al. (2013). Body Panning: a movement-based navigation technique for large interactive surfaces. Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces. St. Andrews, Scotland, United Kingdom, ACM**:** 37-40.

Leifert, S. (2011). The influence of grids on spatial and content memory. CHI '11 Extended Abstracts on Human Factors in Computing Systems. Vancouver, BC, Canada, ACM**:** 941-946.

Miller, G. A. (1956). "The magical number seven, plus or minus two: some limits on our capacity for processing information." Psychological Review **63**(2): 81-97.

Mueller, J., et al. (2015). Virtual Objects as Spatial Cues in Collaborative Mixed Reality Environments: How They Shape Communication Behavior and User Task Load. To appear in Proceedings of the 34th annual ACM conference on Human factors in computing systems (CHI'16)}, ACM}.

Muller, J., et al. (2015). An Experimental Comparison of Vertical and Horizontal Dynamic Peephole Navigation. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Seoul, Republic of Korea, ACM**:** 1523-1526.

Patrick, E., et al. (2000). Using a large projection screen as an alternative to head-mounted displays for virtual environments. Proceedings of the SIGCHI conference on Human Factors in Computing Systems. The Hague, The Netherlands, ACM**:** 478-485.

Rädle, R., et al. (2014). Bigger is not always better: display size, performance, and task load during peephole map navigation. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.

Scarr, J., et al. (2013). Supporting and Exploiting Spatial Memory in User Interfaces, Now Publishers Inc.

Shepard, S. and D. Metzler (1988). "Mental rotation: effects of dimensionality of objects and type of task." Journal of Experimental Psychology: Human Perception and Performance **14**(1): 3.

Simmons, T. (2001). "What's the Optimum Computer Display Size?" Ergonomics in Design **9**(4): 19-25.

Tan, D. S. (2004). Exploiting the cognitive and social benefits of physically large displays, DTIC Document.

Tan, D. S., et al. (2003). With similar visual angles, larger displays improve spatial performance. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Ft. Lauderdale, Florida, USA, ACM**:** 217-224.

Tan, D. S., et al. (2006). "Physically large displays improve performance on spatial tasks." ACM Trans. Comput.-Hum. Interact. **13**(1): 71-99.

Tan, D. S., et al. (2004). Physically large displays improve path integration in 3D virtual navigation tasks. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Vienna, Austria, ACM**:** 439-446.

Tan, D. S., et al. (2002). Kinesthetic cues aid spatial memory. CHI '02 Extended Abstracts on Human Factors in Computing Systems. Minneapolis, Minnesota, USA, ACM**:** 806-807.

Tan, D. S., et al. (2001). The Infocockpit: Providing location and place to aid human memory. Proceedings of the 2001 workshop on Perceptive user interfaces, ACM.

Teixeira, R. A. A., et al. (2011). "Memory span measured by the spatial span tests of the Cambridge Neuropsychological Test Automated Battery in a group of Brazilian children and adolescents." Dementia & Neuropsychologia **5**: 129-134.

Tripathi, K. (2011). "A study of interactivity in human computer interaction." International Journal of Computer Applications **16**(6): 1-3.

Tyndiuk, F., et al. (2005). Cognitive comparison of 3D interaction in front of large vs. small displays. Proceedings of the ACM symposium on Virtual reality software and technology. Monterey, CA, USA, ACM**:** 117-123.

Winner, E., et al. (2000). "Dyslexia and visual-spatial talents: No clear link." International Dyslexia Association Newsletter: 27-30.

Zagermann, Johannes; Pfeil, Ulrike; Fink, Daniel; Bauer, Philipp; Reiterer, Harald (accepted) Memory in Motion: The Influence of Gesture- and Touch-Based Input Modalities on Spatial Memory. To appear in: Proceedings of the 35th annual ACM conference on Human factors in computing systems (CHI ’17), ACM Press, May 2017.

A1 Welcome letter

Welcome

First, I would like to thank you for agreeing to participate in my study. Before we start, I want to provide you with a brief introduction, which includes my study goals and the role you will play.

I want to evaluate the user's performance in front of the larger display and the smaller display, to this end, the participants (you) will perform the same task two times in front of two different size of screen. In addition, I also investigate the usability of the size of screen in order to improve the performance. It is important to remark it is the influence of display size on spatial memory what is being tested and not the participant, so there is no right or wrong answer.

You will have to solve four tasks during the experiment. All the interactions with the system will be automatically recorded. You will have to fill out a short questionnaire before the experiment and a short interview afterwards if necessary. The session will last approximately 65-70 minutes.

During your participation in my study, you are allowed to withdraw at any time, so feel free to communicate it to me with any other question or concern that you may have.

Finally, I wish you lots of fun and I would like to thank you again for your participation.

A2 Letter of Consent

Dear participant,

I would like to thank you for participating in the study. As usual with any study, I will analyse the generated data and publish them anonymously in future publications to contribute to HCI research community. I guarantee absolute discretion, so no information can be traced back to you.

The generated data will proceed from the recorded interaction between the user and system, and also from the filled-out questionnaires /interviews.

During the test you are allowed to withdraw at any time and still you are allowed to receive your compensation. You can cancel the trial at any time! If you need a break, do not be shy, let me know. If you have questions about the general course or the system, you can ask them at any time. However, please understand that I can respond to task-specific questions until after the study in order to prevent a distortion of the data.

Please confirm with your name and signature that you agree with the recording and processing of confidential data mentioned above.

Place/Date: Konstanz,

Participant Name: ____________________

Signature: ____________________

Place/Date: Konstanz,

Evaluators Name: ____________________

Signature: ____________________

A3 Demographic questionnaire

Participant Id...............

Thank you for agreeing to participate in this investigation. Before we begin, I still need some information about you and your experience with computers. I would like to inform you that all data will be treated confidentially.

Personal Information

Age: ______________ Nationality: ______________

Sex: ◯ male ◯ female

Education: ◯ High school ◯ Graduation ◯ Post graduation ◯ Others:

Current job / profession: ______________________

Do you have any kind of visual weakness?

◯ Myopia - how strong? _______ ◯ Long-sightedness - how strong? ______ ◯ Other:

Computer usage

For how many years have you been using a computer?

How many hours do you spend on your computer every day?

◯ Up to 1 hour

◯ More than 1 hour, up to 2 hours

◯ More than 2 hours, up to 3 hours

◯ More than 3 hours

Rate your expertise with the following input devices

	Beginner				Expert
◯ Keyboard	◯	◯	◯	◯	◯
◯ Mouse	◯	◯	◯	◯	◯

A4 NASA/TLX

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

A5 Post questionnaire

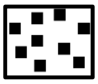
University of Konstanz

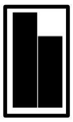
Participant Id ……….

1. Please mention which screen size did best for the tasks? Explain why?

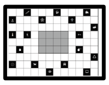
   ………………………………………………………………………………………

2. Can you rank the tasks (easiest =1 to toughest=4)? Explain why?

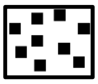
 ………………………………………………………………………

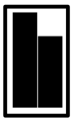
 ………………………………………………………………………

 ………………………………………………………………………

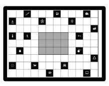
 ………………………………………………………………………

3. Please mention your preferred screen size to perform the task?

 ………………………………………………………………………

 ………………………………………………………………………

 ………………………………………………………………………

 ………………………………………………………………………

A6 Counterbalancing in four tasks across two displays

Participants	**Counter balanced 4 tasks and 2 displays**			
P1	Ad BD	Bd CD item b	Dd (TI 1) AD	Cd item-a DD (TI 2)
P2	Bd AD	Cd item-b BD	Ad DD (TI 1)	Dd (TI 2) CD item a
P3	Cd item-a DD (TI 2)	Dd (TI 1) AD	Bd CD item b	Ad BD
P4	Dd (TI 2) CD item a	Ad DD (TI 1)	Cd item-b BD	Bd AD
P5	AD Bd	BD Cd item a	DD (TI 2) Ad	CD item-b Dd (TI 1)
P6	BD Ad	CD item-b Bd	AD Dd (TI 1)	DD (TI 2) Cd item a
P7	CD item-b Dd (TI 1)	DD (TI 2) Ad	BD Cd item a	AD Bd
P8	DD (TI 2) Cd item a	AD Dd (TI 1)	CD item-b Bd	BD Ad
P9	Ad BD	Bd CD item a	Dd (TI 1) AD	Cd item-a DD (TI 2)
P10	Bd AD	Cd item-a BD	Ad DD (TI 2)	Dd (TI 1) CD item b
P11	Cd item-a DD (TI 2)	Dd (TI 1) AD	Bd CD item b	Ad BD
P12	Dd (TI 1) CD item b	Ad DD (TI 2)	Cd item-a BD	Bd AD

P13	AD Bd	BD Cd item a	DD (TI 2) Ad	CD item-b Dd (TI 1)
P14	BD Ad	CD item-b Bd	AD Dd (TI 1)	DD (TI 2) Cd item a
P15	CD item-b Dd (TI 1)	DD (TI 2) Ad	BD Cd item a	AD Bd
P16	DD (TI 2) Cd item a	AD Dd (TI 1)	CD item-b Bd	BD Ad
P17	Ad BD	Bd CD item b	Dd (TI 1) AD	Cd item-a DD (TI 2)
P18	Bd AD	Cd item-a BD	Ad DD (TI 2)	Dd (TI 1) CD item b
P19	Cd item-a DD (TI 2)	Dd (TI 1) AD	Bd CD item b	Ad BD
P20	Dd (TI 2) CD item a	Ad DD (TI 1)	Cd item-b BD	Bd AD
P21	AD Bd	BD Cd item a	DD (TI 2) Ad	CD item-b Dd (TI 1)
P22	BD Ad	CD item-b Bd	AD Dd (TI 1)	DD (TI 2) Cd item a
P23	CD item-b Dd (TI 1)	DD (TI 2) Ad	BD Cd item a	AD Bd
P24	DD (TI 2) Cd item a	AD Dd (TI 1)	CD item-b Bd	BD Ad
P25	Ad BD	Bd CD item a	Dd (TI 1) AD	Cd item-a DD (TI 2)
P26	Bd AD	Cd item-a BD	Ad DD (TI 2)	Dd (TI 1) CD item b

P27	Cd item-a DD (TI 2)	Dd (TI 1) AD	Bd CD item b	Ad BD
P28	Dd (TI 2) CD item a	Ad DD (TI 1)	Cd item-b BD	Bd AD
P29	AD Bd	BD Cd item a	DD (TI 2) Ad	CD item-b Dd (TI 1)
P30	BD Ad	CD item-b Bd	AD Dd (TI 1)	DD (TI 2) Cd item a
P31	CD item-b Dd (TI 1)	DD (TI 2) Ad	BD Cd item a	AD Bd
P32	DD (TI 2) Cd item a	AD Dd (TI 1)	CD item-b Bd	BD Ad
P33	Ad BD	Bd CD item b	Dd (TI 1) AD	Cd item-a DD (TI 2)
P34	Bd AD	Cd item-b BD	Ad DD (TI 1)	Dd (TI 2) CD item a
P35	Cd item-a DD (TI 2)	Dd (TI 1) AD	Bd CD item b	Ad BD
P36	Dd (TI 2) CD item a	Ad DD (TI 1)	Cd item-b BD	Bd AD
P37	AD Bd	BD Cd item a	DD (TI 2) Ad	CD item-b Dd (TI 1)
P38	BD Ad	CD item-b Bd	AD Dd (TI 1)	DD (TI 2) Cd item a
P39	CD item-b Dd (TI 1)	DD (TI 2) Ad	BD Cd item a	AD Bd
P40	DD (TI 2) Cd item a	AD Dd (TI 1)	CD item-b Bd	BD Ad

1. d- Small screen
2. D- Large screen
3. A,B,C,D- four tasks where:
    - A- Spatial Span task
    - B- Histogram Rotation task
    - C- Navigation task
    - D- Tangram task
4. Item a and item b are item sets for Navigational task.
5. TI 1 and TI 2 (Target Image 1 and 2 or item set 1 and 2) are two sets in Tangram task.

A7 Digital media

- Digital version of thesis
- Tasks
- Documents used in experiment
  - Welcome letter
  - Letter of Consent
  - Demographic Questionnaire
  - NASA/TLX Questionnaire
  - Post Questionnaire
  - Counterbalancing in four tasks across two displays