



HAL
open science

Human-computer interaction in 3D object manipulation in virtual environments: A cognitive ergonomics contribution

Sarwan Abbasi

► **To cite this version:**

Sarwan Abbasi. Human-computer interaction in 3D object manipulation in virtual environments: A cognitive ergonomics contribution. Computer Science [cs]. Université Paris Sud - Paris XI, 2010. English. NNT: . tel-00603331

HAL Id: tel-00603331

<https://theses.hal.science/tel-00603331>

Submitted on 24 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Human-computer interaction in 3D object
manipulation in virtual environments:
A cognitive ergonomics contribution**

Doctoral Thesis submitted to the Université de Paris-Sud 11, Orsay
Ecole Doctorale d'Informatique
Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
(LIMSI-CNRS)
26 November 2010

By

Sarwan Abbasi

Jury

Michel Denis (*Directeur de thèse*)

Jean-Marie Burkhardt (*Co-directeur*)

Françoise Détienne (*Rapporteur*)

Stéphane Donikian (*Rapporteur*)

Philippe Tarroux (*Examineur*)

Indira Thouvenin (*Examinatrice*)

Acknowledgements

One of my former professors, M. Ali YUSUF had once told me that if I learnt a new language in a culture different than mine, it would enable me to look at the world in an entirely new way and in fact would open up for me a whole new world for me. I have found what he had said to be entirely true. So this was one of the things which had motivated me to look for a scholarship in a country, where the primary language of communication was not English (or for that matter, any of the other local South-Asian languages that I learnt as a child), and hence I am here today writing my thesis in France. More concretely, this research was made possible through a contract between Higher Education Commission (HEC) and the French Embassy in Pakistan. While HEC funded most of my research for which I am grateful; the French Embassy arranged for us (me and certain other selected scholars) to learn and discover the French language and its culture by means of immersive language courses as well as arranging visits to different parts of the country before the commencement of academics. My interactions with the different people that I have met in France as well as the exploration of its different cities and regions has enabled me to enrich myself personally, academically, and culturally.

This research came into being as a result of a meeting that was held between me and Michel DENIS, who has become my thesis director at LIMSI-CNRS at the Université Paris-Sud 11 (Orsay). Since that time and up until now, it would not have been possible to carry on, had it not been for the help provided to me by him on a continual basis. He has always been there whenever I have needed him for his intervention in the capacity of my aid, be it administrative or academic. He has gone beyond the normal limits, in helping me out with progressing farther, even when sometimes I had thought that it was beyond my limits, though his aid was not limited just to encouraging me only. He responded to my emails most promptly, and always gave invaluable, timely, and relevant feedback whenever I had needed it. This was indeed extraordinary. Michel, I really thank you for that.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

I was lucky enough to have gotten a chance to work under the direct supervision of Jean-Marie BURKHARDT, my co-director, and with whom I also shared the same work-place at the Université Paris Descartes. Whether at his office desk or at lunch table, he provided me with invaluable academic feedback on a continual basis and helped me learn and generate new ideas whenever I had needed them, and made me see things from a fresh point of view. I cannot thank him enough, for other than giving me useful advises throughout my research, he also ensured that I had enough resources for my personal sustenance especially towards the end of my thesis.

No research takes place in isolation and without the interaction and/or help from a vast variety of people and my research is not an exception to that. The most critical portion of the technical facilities and assistance was provided by the VENISE team at the LIMSI-CNRS lab at Université Paris-Sud 11 (Orsay). I would primarily like to thank Patrick BOURDOT, who is also the director of VENISE, who kindly accepted to let his virtual reality room and other facilities at his lab be used for my research. I cannot possibly thank him enough for accepting to help me, as a major part of my research would not have been possible without it. In addition to just the place and equipment, I was also helped by various other people of the VENISE team, including Pierre MARTIN, Damien TOURAINE and Jean-Marc VEZIEN, without whose help also, the experiments in the virtual setting would not have materialised. In particular, Pierre MARTIN helped me extraordinarily, to make my underlying ideas about the virtual puzzle (especially the interactions part) come to life. I cannot thank him enough for the devoted and untiring effort that he put in for me.

I am grateful to Françoise DETIENNE and Stéphane DONIKAN for agreeing to be reviewers of my thesis. They provided me feedback and suggestions which proved very valuable for me and which are reflected in the current version of the thesis. I am also thankful to Philippe TARROUX and Indira THOUVENIN who graciously accepted to be the examiners of my thesis.

There are many other people that I would like to thank, who have helped me in one way or the other, and I will probably not be able to name them all here. They appear here in no

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

particular order: my former colleague Rami AJAJ (researcher in graphics), a former student and an internee at Université Paris Descartes, Gosia TARASEWICZ (ergonomist), internee Adrien SCHWARTZ (ergonomics student) and Karin HEIDLMEYER (psychology student). I am thankful also to many friends and family members, particularly Safdar Abbas KHAN, Sohail IQBAL, Issa ABBASI, Tjaša NABERGOJ, Marie-Laure LE GUEN, Asim ARIF and Taj KHAN, who helped me at different times and in different ways, whether their help was in the form of brain-storming, typing, proof-reading, giving ideas when I got stuck, or in the form of giving helpful advices and encouragement.

I thank also all the people who graciously accepted to participate in my experiments and without whom this research would also not have been possible.

Last but not the least, I would like to thank my colleagues Julien NELSON, Cécile GIRARD, Amel Ben AYED, Sophie CAPO, Jeannine BOURDEAUX, Katarina BOHACOVA and Stanislas COUIX, who helped me at different times and in different ways with my research work, as well as with my personal development.

It is perhaps impossible to name everyone, and I am sure there would still be a lot of people whose names I have forgotten, and for which I sincerely apologise in advance.

Executive Summary

It is proposed to investigate the cognitive processes involved in assembly/disassembly tasks, and then to apply the findings to the design of 3D virtual environments (VEs). Virtual Environments are interactive systems that enable one or more users to interact with the simulation of objects and scenes usually in three dimensions, in a realistic fashion, by means of a set of computational techniques covering one or more sensory modalities (vision, touch, haptic, hearing, etc.). Often described as the ultimate direct manipulation interface, this technology seeks to make the interface eventually 'disappear' in order to provide users with a 'natural' mode of interaction. Virtual reality (VR) is the experience of being within a VE. One objective of the VR technology is indeed to exploit natural human behaviour without requiring any learning from their users [Fuchs2003], [Bowman2005].

Moreover, VEs are a stimulating field of research because they involve perceptually and cognitively novel situations [Burkhardt2003]. VEs also offer a large potential of innovative solutions to existing application problems. Among others, assembly tasks are a major focus for VEs [Boud2000], [Brooks1999], [Lok2003-a], [Lok2003-b], due to their numerous potential applications, such as assembly/disassembly of objects, scientific research (e.g., molecular docking [Ferey2009] etc.). The common feature in VEs is the use of representations and devices to support the users in handling and arranging several distinct elements in a three dimensional (3D) space under specific constraints.

Most of the current devices and interaction techniques have focused on providing users with high-fidelity sensory stimulations, rather than targeting real-life or task-centred functions associated with the corresponding interfaces.

While many contributions have been made to the field of VR, there are only few empirical data that have been published. We believe that it is very unlikely that more adapted VEs and assistance to users' task – in the specific context of assembly tasks – will follow either just by chance [Brooks1999], by making repeated trials, by tuning what we already have at hand, or

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

by more realistic sensory renderings, without any reference to the 'specific properties of the tasks' including its cognitive dimension. Consequently, a clear picture of the cognitive processes and constraints in real tasks involving spatial manipulation should lead to a significant enhancement of the users' interactions with VEs. This enhancement can be made by creating better or new guidance mechanisms (e.g., video feedback, object collision detection, or avoidance mechanisms) adapted to the users' goals and strategies. This project thus involves work both from the cognitive side and its implications on 3D interactions in industrial VEs.

The objective of this doctoral work is to contribute to a better understanding of human factors (HF) – including performance and cognitive processes – related to assisting spatial 3D manipulation and problem-solving in assembly/disassembly tasks in VEs. For that purpose, we compared performance and strategies of subjects while they solve a simplified spatial task requiring them to assemble pieces to form a specified shape in various conditions of interfacing actions in real and virtual environments. The assembly task chosen was neither very easy such as put peg-in-a-hole type task, as in [Zhang2005], [Pettinaro1999], or [Unger2001], nor highly complex and specific, such as performing open heart or liver surgery [Torkington2001] (one whose results could be applied only to that specific kind of task). The chosen task was semi-complex, in which the users were required to construct a 3D cube using seven rectangular blocks of different sizes and shapes.

The methodology used had two tiers: real and virtual. For the chosen assembly task, a study was first conducted in real settings, which was to provide inspiration, input, and insight for the main experiment to follow. The main experiment that followed was similar in design, but the difference was that it was conducted in virtual settings. The experiment in virtual settings was conducted in three modalities – the classical keyboard-mouse, the gestural modality, and the vocal modality.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

The thesis is divided into two parts. Part I discusses the motivations for our problem and problem domain: puzzles. It then focuses on theoretical aspects underlying puzzles – more specifically spatial puzzles.

Chapter 1 discusses the motivation behind our work and gives a broad introduction to our problem domain, the state of the art in VR and VEs, particularly with reference to the works supporting assembly tasks. It highlights the importance of the HFs related to assembly tasks in virtual and real environments.

Chapter 2 focuses on problems and problem solving, and the relationship of problem solving with puzzles. It discusses the categorisation of different problems as well as the role and nature of constraints while solving problems. This chapter also discusses the role puzzles have played in the popular and scientific culture of problem solving.

Chapter 3 is dedicated to a special category of puzzles, the spatial puzzles, because our chosen problem (and experiment) is also categorised as a type of spatial puzzle. This chapter briefly discusses some of the popular classical spatial puzzles of historical importance, and also proposes to categorise the different kinds of spatial puzzles on the basis of their physical and structural aspects.

Chapter 4 discusses the issues related to interfaces and highlights the importance of their characteristics, and the role they can play (both in real and virtual settings) in rendering any problem easy or difficult.

Part II is dedicated to our experiments designed to studying 3D objects manipulation tasks in real and virtual environments.

Chapter 5 lays the groundwork for the two experiments (one in the real setting, and the other in the virtual one) to follow. It explains the idea and motivation behind the experiment

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

which is in fact a spatial puzzle. The chapter also discusses the methodological aspects regarding the experiments, including physical and technical aspects of the puzzle.

Chapter 6 is dedicated to the experiment that was conducted in the real setting. It discusses the experiment design and methodology used, followed by the result of the experiment and analysis.

Chapter 7 is dedicated to the experiment designed and conducted in the virtual settings. The experiment was in fact based on the same puzzle which was discussed in chapters 5 and 6. This chapter thus discusses the aspects that are specific to the virtual settings, including the structure (hardware and software components) of the experiment and the different modalities used. It discusses the existing related work. Different modalities used and the method employed for the experiment are presented followed by the results of the experiment. Finally it summarises, compares and discusses the results of the experiment in different settings followed by analysis, observations and our commentary/explanations on them.

Lastly, chapter 8 is the concluding chapter. It gives the quick overview of our work. It draws general conclusions followed by our suggestions and recommendations. Two sections provide distinct sets of conclusions regarding the contribution of our study to knowledge in ergonomics and its contribution to informatics and human-computer interactions. In the end, we discuss applications for our work and future possibilities.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

CONTENTS

PART I	5
INTRODUCTION AND THEORETICAL FRAMEWORK	5
1 CHAPTER 1: INTRODUCTION	6
1.1 VR IN SUPPORTING ASSEMBLY TASKS	7
1.2 HUMAN FACTORS RELATED TO ASSEMBLY TASKS IN REAL AND VIRTUAL ENVIRONMENTS	9
1.2.1 Assembly/disassembly and handling of objects: Ergonomic approaches in real situations	10
1.2.2 Assembly and manipulation in virtual environments: Current work and results from an ergonomic point of view	10
1.3 AN ALTERNATIVE FRAMEWORK: SPATIAL COGNITION, PROBLEM SOLVING AND 3D PUZZLES	12
1.4 CONCLUSION: THE OVERALL APPROACH TAKEN IN THIS WORK	15
2 CHAPTER 2: PROBLEM AND PROBLEM SOLVING: BACK TO BASICS	17
2.1 DEFINITIONS	18
2.2 THE NOTION OF PROBLEM SPACE: MODEL FOR ANALYSIS VS. REPRESENTATION ELABORATED BY THE SUBJECTS	20
2.3 ROLE AND NATURE OF CONSTRAINTS	24
2.4 TYPES OF PROBLEMS	24
2.5 DIFFERENT WAYS OF PROGRESS/REDEFINITION	27
2.6 OTHER APPROACHES TO PROBLEM SOLVING	29
2.7 CONCLUSIONS: PUZZLES AND PROBLEM SOLVING	30
3 CHAPTER 3: SPATIAL PUZZLES: AN OVERVIEW, ISSUES INVOLVED, AND NEW PERSPECTIVES	32
3.1 INTRODUCTION TO SPATIAL PUZZLES	33
3.1.1 Dissection puzzles	36
3.1.2 Mechanical puzzles	42
3.1.3 Mazes	44
3.2 DISCLAIMER TO FINALITY OF CATEGORISATION	45
3.3 COMPUTERISATION CHANGING THE NATURE OF PUZZLES	45
3.4 BUTLER'S CATEGORISATION OF SPATIAL PUZZLES	47
3.5 OTHER APPROACHES TO 3D PROBLEM SOLVING IN RELATED DOMAINS	49
3.6 CONCLUSIONS: TOWARDS MORE COHERENT CATEGORISATION OF PUZZLES	50
4 CHAPTER 4. INTERFACE, PRESENTATION MEDIUM AND CONSTRAINTS	52
4.1 PRESENTATION MEDIUM, INTERFACE, CONTROLS	53
4.2 RULES AND CONSTRAINTS	54

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

4.3	INTERFACE-RELATED CHARACTERISTICS OF SPATIAL PUZZLES	57
4.4	VISUALISATION	58
4.5	CONCLUSIONS: THE CRUCIAL ROLE OF PRESENTATION AND INTERACTION	60
PART II		61
EXPERIMENTAL CONTRIBUTIONS TO THE STUDY OF 3D OBJECTS MANIPULATION TASK IN REAL AND VIRTUAL ENVIRONMENTS		61
5	CHAPTER 5: THE PUZZLE EXPERIMENTS	62
5.1	THE IDEA AND MOTIVATIONS BEHIND THE EXPERIMENT	63
5.1.1	Questions raised in our experiments	64
5.1.2	The task chosen	64
5.2	MATERIALS, PHYSICAL AND TECHNICAL ASPECTS OF THE PUZZLE	65
5.2.1	Materials	66
5.2.2	Conventions used for measurement of the dimensions of blocks	68
5.2.3	Comparison of the blocks used in our experiments with the Soma Cube	69
5.2.4	The complexity of the blocks	70
5.2.5	Formal description of the cube including the different orientations and the positioning of the blocks	74
6	CHAPTER 6. THE BLOCKS MANIPULATION TASK: THE EXPERIMENT IN REAL SETTINGS	83
6.1	PHYSICAL SETTING	84
6.2	DESIGN AND METHOD	86
6.2.1	Participants	86
6.2.2	Procedure	86
6.3	DATA COLLECTION	88
6.4	METHOD OF ANALYSIS	88
6.5	NOTATIONS OF MOVES	91
6.5.1	The idea of attempt as the container of moves	93
6.6	RESULTS	93
6.6.1	Overall performance	93
6.6.2	Moves and attempts	97
6.6.3	Types of moves	104
6.6.4	Styles of solving and classification of blocks	105
6.6.5	Verbalisations	107
6.6.6	Post-experimental comments	109
6.6.7	Conclusions	110

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

7	CHAPTER 7: THE VIRTUAL 3D-PUZZLE	111
7.1	HARDWARE/SOFTWARE	113
7.2	BIG PICTURE: THE OVERALL SCENE AND THE SETUP	113
7.3	LITTLE PICTURE: THE COMPONENTS OF THE SCENE AND THE BEHAVIOURAL ASPECTS COMMON TO ALL MODALITIES	114
7.3.1	The convention for axes	115
7.3.2	The blocks	115
7.3.3	Selected and deselected blocks	115
7.3.4	Arrow emanating from the moving objects denoting the axis of movement	116
7.3.5	Casting of the shadows	116
7.3.6	Collisions	117
7.3.7	Occlusion	117
7.3.8	Six degrees of freedom (6-DOF)	117
7.3.9	The view-ball	118
7.3.10	Virtual levitation (lack of gravity)	119
7.4	EXISTING RESEARCH ON MODALITIES AND INTERACTION TECHNIQUES AND THE CHOICES MADE FOR THE VIRTUAL PUZZLE	119
7.4.1	Our chosen modalities	124
7.4.2	Other related issues	125
7.5	DESCRIPTION OF MODALITIES FOR THE EXPERIMENT IN THE VIRTUAL SETTINGS	127
7.5.1	Keyboard/mouse: The classical modality	127
7.5.2	Gestural modality	130
7.5.3	Vocal modality	138
7.5.4	Multimodality (G+V)	139
7.5.5	Salient features of the interface	145
7.6	METHOD EMPLOYED FOR THE RESEARCH	145
7.6.1	THE OVERALL PROTOCOL	145
7.7	RESULTS	151
7.7.1	Participants	151
7.7.2	Overall performance	153
7.7.3	Partial success distribution for G and V	154
7.7.4	Mean times for solution	156
7.7.5	Comparison of moves in different modalities	156
7.7.6	Solution as a function of number of attempts	161
7.8	COMPARISONS AND DISCUSSION	166
7.9	THE MULTIMODAL CONDITION (WITH BOTH G AND V)	172
7.9.1	Success rate	172
7.9.2	Average time for solving the puzzle	173

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

7.9.3	Average number of moves	173
7.9.4	Number of moves per attempt	174
7.9.5	Number of attempts	174
7.10	CONCLUSIONS	174
8	CHAPTER 8: DISCUSSION, CONCLUSIONS, AND PERSPECTIVES	176
8.1	OVERVIEW	177
8.2	CONTRIBUTION TO ERGONOMICS KNOWLEDGE	178
8.3	CONTRIBUTIONS TO INFORMATICS AND HUMAN-COMPUTER INTERACTION	180
8.3.1	Difficulties to map the knowledge from real setting onto the interactions commands in a virtual environment	180
8.3.2	Performance with multimodal inputs and suggestions for improvement	180
8.4	PERSPECTIVES	182
8.4.1	Research perspectives	182
8.4.2	Applications (of manipulation of objects in 3D space)	184

Part I

Introduction and theoretical framework

1 Chapter 1: Introduction

This chapter discusses the motivation and gives a broad introduction to our problem domain, presents the state of the art in virtual reality and virtual environments, particularly with reference to the works supporting assembly tasks. It highlights the importance of the human factors related to assembly tasks in real and virtual environments. It touches upon the ergonomics of assembly tasks in real life situations, and how an improvement in that area can lead to improved efficiency. It concludes by analysing different approaches taken to study human problems, one of them being the use of spatial puzzles for studying human spatial cognition. Finally, it briefly explains the experiments that we conducted in real and virtual settings using various interaction modalities.

Assembly tasks are a major focus for virtual environments (VEs) [Boud2000], [Brooks1999], [Lok2003-b], since their potential applications are enormous [Fuchs2006]. VEs use various representations, devices and interaction techniques, either generic or tailored to specific devices, to support their users in handling and arranging several distinct elements in a three-dimensional (3D) space under specific constraints. For example, in desktop computing, a mouse is a generic device which uses certain generic interaction techniques, such as the pointer which represents the placement of the mouse on desktop, and moving the mouse moves the mouse pointer over the desktop. To select an object on the desktop, the user may first move the mouse so that the mouse pointer is over the desired object, and then click the mouse key to select the object over which the mouse pointer was placed. Likewise, the keyboard is another such standard input device for desktop computers. In VEs, there are no such omnipresent devices which are accepted by the whole industry [Brooks1999]. Improvements in interfaces that resemble any kind of "realistic" or real-life like situation would in turn help advance the field of VEs itself.

1.1 VR in supporting assembly tasks

What Poupyrev said back in 1997 ("manipulation of objects in VEs is often awkward and inconvenient") still seems to be quite relevant, acknowledging that tiny factors, such as lack of tactile feedback, or poor design of interaction techniques, can turn a reasonably simple task (i.e. simple in the real-world) of grabbing and moving an object into a frustrating experience in the virtual world [Poupyrev1997].

[Wang1998] deals with grasping (or prehension) and makes a comparison of performances in real and virtual tasks. MacKenzie and Iberall [MacKenzie1994] discuss the versatilities and the complexities of the human hand, and talk about the complexities involved in grasping and the task of 'manipulation' in general.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

More recent works related to assembly/disassembly or manual (hand related) tasks generally involve comparative experiments in which the users perform the task in “real” and “virtual” settings. See, for example, the work of Axelsson and colleagues [Axelsson2001], who tests performance in real and virtual settings (though their focus is rather geared towards factors such as presence, co-presence, collaboration and even leadership shown among participants). [Wideström2000] focuses more on the collaboration, and not the interactions or the interface.

A variety of tasks are used for this purpose, and performance is compared in real and virtual environments. It is reported that performances are indeed different under the two settings, but this does not help answer the critical reasons behind those differences, or identify the shortcomings which could point one in the right direction, eventually helping build VEs which could reduce those performance gaps.

We believe that the focus right from the start should be to try to focus on the interface and interaction techniques and to try to identify (or shortlist) the critical differences in interaction techniques which could play the most critical role in users' performance, and then to work on those interaction techniques.

[Froehlich2000] developed a new interesting input device which was in the form of a cube shaped box, with rods projected out of the six sides of cube (with in fact three perpendicular rods passing through the box) and buttons at the top (Figure 1.1). The rods could be pushed and pulled and they represented the X, Y and Z axes, which is promising for certain applications due to its natural mappings with the real 3D world.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

We are constantly reminded that the right VE interface depends highly on the unique set of requirements of a given application: scale of distances, required form of movements and operations [Bowman2005] and other requirements such as power and precision [MacKenzie1994]



Figure 1.1: The Cubic Mouse Device

1.2 Human factors related to assembly tasks in real and virtual environments

There is no denying of the fact that the human hand is indeed a very versatile and complex tool [MacKenzie1994]. We use it for a vast variety of complex tasks ranging from writing to hammering nails, from turning the steering wheel of the car to changing its gear, from playing golf to pushing/pulling heavy objects. All these tasks have different power or precision requirements, and necessitate different hand postures and grasping techniques to perform each of them. MacKenzie and Iberall have treated the subject in great detail in their book 'The Grasping Hand' [MacKenzie1994].

1.2.1 Assembly/disassembly and handling of objects: Ergonomic approaches in real situations

Much of the work in ergonomics related to assembly tasks by operators in "real" situations is yet to be developed. The domain lacks a definition of guidelines for the design of physical machines [Wartenberg2004]. The focus is on the physical and physiological (workload, gesture, posture) and rarely consider the detailed features of the task (for example tolerance to precision, visual and manipulation requirements), which affect the task itself (even if one ignores other factors like 'poor work posture' which have their own costs in the long run [Li1999]).

It has been shown (for example in [Wartenberg2004]) that the presence of constraints such as increase in the precision requirements greatly increases the assembly times (sometimes by over 100%) and induces greater numbers of shorter cycles of movement, more control, verification and "immobility" (i.e. a strong restriction of space and a reduction in variation) of body movement and posture. The assembling tasks have also been addressed in ergonomics literature on procedural instructions and on learning process (for example in [Ganier2004]). This second approach is often focused on the effect of the instructions on assembly performance and their characteristics rather than on the processes of spatial problem-solving, like reasoning, and the effect of physical properties of materials, and manipulation. These studies provide data on 'cognitive processes' as follow-up procedural instructions. They suggest principles and potential solutions on improvement and assistance by means of instructions and documents.

1.2.2 Assembly and manipulation in virtual environments: Current work and results from an ergonomic point of view

Current research on using assembly tasks have been conducted in the context of designing VEs. For a long time, such tasks were considered uniquely from the scientific and technical view-point [Pettinaro1999] and more recently as a major axis for VR and Augmented Reality

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

(AR) [Boud2000]. Indeed, many real applications such as assembly/disassembly of objects in the industry with goals pertaining to training or maintenance, evaluation of the conceived product [Fuchs2006], scientific research (for example molecular docking [Férey2009]), and also with the objectives of training, result in evaluation or assistance in finding solutions, or using problem situations for conceptual learning [Roussou2006].

The technical literature is filled with strong hypotheses, about how VEs can contribute to assembling. Yet one rarely finds empirical evidence published on the utility or usability of VEs conceived for using different situations and specific tasks [Abbasi2009-b]. A finding based on expert opinion and confirmed by some empirical studies [Boud2000] shows that this type of task is difficult in VEs and the manipulation is quite often unintuitive, even for "experts" of the domain. Several explanations have been put forward, such as mismatch and the difference in scale between physical and digital manipulation spaces, or the lack of comprehensive information feedback during the direct manipulation of objects.

To improve the ergonomics of these systems, researchers are committed to making the design of the interaction techniques more effective, employing three different approaches (sometimes combined). The first approach involves the development and evaluation of interaction metaphors based on an inventory of actions of the objects in 3D space. In this approach, the assembly is seen as a task of manipulating objects consisting of three generic actions: the selection, the positioning and the rotation of objects (see for example [Bowman2005]). The second approach tries to evaluate the contribution of high-fidelity sensory information, with the performance in VEs. These studies focus for example on the visual modality, in particular stereoscopy [Werkhoven1998]. Recently a lot of work has focused on tactile (touch) and haptics (force feedback) [Boud2000], [Unger2001], [Lécuyer2002]. The third approach deals with "augmentation" or "substitution" of sensory information presented to subjects, for improving their performance in VEs. These studies compare and assess the contribution of different combinations of information -

complementary or redundant, according to the varied formats - of visual, auditory and haptic information, for example [Lécuyer2002] and [Petzold2004].

These studies converge, amongst others, on four main results. (1) The time to complete an assembly task in "virtual" environment takes much longer (twice as long or more), than in "real" environment [Boud2000], [Unger2001]. (2) The visual information, according to its format and function (realistic, simplified, depth information, type of avatar etc.) strongly affects the user performance [Werkhoven1998], [Petzold2004], [Kadri2007]. (3) The addition of haptic information generally facilitates the assembly of objects in space [Boud2000], [Lécuyer2002], [Petzold2004]. And (4), the provision of information on more than one modality has mixed effects, as it may either assist or confuse the subjects with information overload and thus induce lower performance.

The dimensions of spatial-reasoning and problem-solving are missing in these studies, which poses a limitation in that it is difficult to say how their findings apply to tasks involving complex reasoning (for example tasks that require spatial reasoning and planning for solving them), like in 3D assembly task. Moreover, these works are based on extremely simple experimental or "atomic" situations, for example inserting a coin in a hole, stacking boxes, etc. The assembly situations and puzzle solving found in some studies are more of a pretext to create interactions between users in collaborative virtual environments [Heldal2005].

1.3 An alternative framework: Spatial cognition, problem solving and 3D puzzles

There are some works that have made use of assembly and manipulation tasks in their research to understand different human aspects, like problem solving. [Svendsen1991] used object manipulation tasks to study the different approaches in problem solving, where differences in cognitive load impact strategies and learning modes. In particular, he showed

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

that command interfaces more likely induce a learning mode based on insight, whereas direct manipulation tends to induce trial and error learning. Interestingly, when people use command interfaces, they are more able to verbalise the principles governing the solution of the problem than when using a direct manipulation interface.

There is some work done in spatial visualisation and mechanical reasoning but it tends to focus on issues other than those directly related to cognitive reasoning. For example, [Keehner2006] discusses spatial visualisation and mechanical reasoning in the context of the differences in performance of spatial skills of people earlier in their careers; versus when they have acquired enough experience to the extent that their actions become more automatic and less attention-demanding. That is to say that it tests whether spatial ability contributes to performance, and how the contributions of the abilities change with greater practice and familiarisation.

We believe that before going farther in finding technology oriented applications to spatial-cognitive problems, it seems more reasonable to invest resources in studying such problems in real-life settings from different aspects, and to know more about them. However, even in these contexts where spatial dimensions and manipulation are involved in the problem to be solved, we have not found any investigation addressing methods and issues about how physical manipulation and environmental constraints might interplay with the cognitive aspects of the solving process implied in the situation.

[Hinckley1994] surveys the works focusing on design issues of free-space input devices, and complains that there is a big void of works whose findings could be applied to different circumstances. The ones which do exist are either about single applications built for users with specialised tasks, or there are other formal studies, but which analyse individual phenomena in isolation. There are other studies based on the differences between the

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

performance in the real world and virtual or augmented environments (for example [Kabbash1993], [Wang1997]).

We would rather emphasise that more research is needed on such realistic problems which require object manipulation, but where the users are also requested to perform some problem resolution. So there is great value for research to focus on how users solve 3D object manipulation tasks which require not only mechanical skills (that is, tasks with just the manipulation aspect associated to it), but also require problem resolution skills.

In order to be usable, the technology (for humans) ought to be designed in such a way that one takes into account the way humans interact with everyday objects. Spatial cognition approach ought to take into account concepts such as *affordance* and other human tendencies such as following patterns based on *Gestalt principles*. *Affordance* (or Don Norman's preferred term *perceived affordance*) is a visual clue to the function of an object [Norman1988]. The designer cares more about the perceived affordance (the affordance which also takes into account the 'context', 'culture', 'instinct' and 'mental model'), or in other words what actions the user perceives to be possible rather than what is true. For example, in the context of a website, the perceived affordance of underlined text is that it indicates a hyperlink; by clicking, the user will link to information that relates to the hyperlinked word or phrase. The Gestalt theory attempts to describe how people organise visual elements into groups or unified wholes (based on certain principles such as similarity, continuation, closure, and proximity).

In addition, there could be other principles such as capitalising on well-established conventions (for example, it would be hard today to introduce new keyboards other than the prevailing *Qwerty/Azerty* ones) and making use of principles learnt from previous experiences (for example, when a rigid body comes in contact with another, they collide

rather than merge). The human dimensions of the interface can become of increasingly greater importance. For example, if one is to compare two systems, both of which are slow and one of them responds to a certain command after a few seconds while the other shows a rotating hour-glass and responds after the same time, the users show a preference for the second one [Nielsen1994], [Lautenbach1999]. Firstly, this is so because in the first system, in the absence of feedback, the user thinks that the systems is simply not responding, gets frustrated and tries to give another command. On the other hand, if the second system shows an hour-glass sign, it has the effect of simply reassuring the user that the system is performing some action in response to the user's command. This sole fact induces the user to be more patient. In some cases, even when the waiting time in both cases is equal, the user's perceived sense of waiting time is reduced in the latter case.

Thus, as one can see in the light of this experiment, human dimensions (especially human perceptions) are important in designing interfaces. Various principles have been introduced and have evolved as good practices in creating interfaces that are perceived to be more usable by humans.

1.4 Conclusion: The overall approach taken in this work

In the current chapter, we discussed that while Assembly tasks are increasingly becoming a major focus for VEs due to their enormous potential applications, their usability still continues to be an issue due to two things: firstly, due to the lack of standardisation of devices and interaction techniques (unlike the traditional desktop computing where the devices and interactions have reached a certain maturity already); and secondly, the potential possibilities applications of VE is also much vast and diversified in comparison to the traditional desktop computing. We also emphasised the importance that one approach to make VEs more usable is to not only consider the current task at hand in the VE, but an even better approach is to start even a step before, by first understanding the task at hand and considering human factors related to it in the real world.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

In our research, we used 3D-object manipulation *puzzles*. Puzzles can be 2D or 3D. Encyclopaedia Britannica defines a *puzzle* as a problem difficult to solve, or a mental challenge. According to Oxford Advanced Learner's Dictionary (OALD) the term puzzle is described as “a game that one has to think about carefully in order to answer it or to do it”. Apart from this primary meaning, there are two other alternative descriptions proposed by OALD. According to one of them, puzzles specifically refer to "jigsaws" (in which users have to assemble a 2D figure from various parts). According to the other, more general definition, a puzzle is "something that is difficult to understand or explain". It may be worth mentioning here that the term puzzle in French is also generally used in the sense of jigsaw, or more precisely a game of patience composed of many fragments which must be assembled to conform to a figure (for example in Larousse dictionary the term puzzle is defined as "Jeu de patience, composé d'un grand nombre de fragments découpés qu'il faut rassembler pour reproduire un sujet complet" [Larousse1991]). However, the definition that we intend to use in this text is the first one, according to which a puzzle is “a game or a mental challenge that one has to think about carefully in order to answer or solve it”.

Two separate experiments were conducted in the scope of this work, which were based on solving 3D-object manipulation puzzles. One of these experiments was conducted in real settings (using real 3D blocks) while the other one was conducted in virtual settings using various interaction techniques. In the virtual experiment, the users operated and manipulated the blocks by means of the classical keyboard-mouse modality, as well as by gesture and voice-based modalities.

The next chapter introduces and discusses the notion of *problems*, the different kinds of (simple and complex) problems that humans encounter, how they solve them, the different ways in which the problems can be solved as well as the role of *constraints* under which problems have to be solved.

2 Chapter 2: Problem and Problem Solving: Back to

Basics

Humans encounter problems every day. However, to find a solution to a given problem, it is indispensable to define a framework where the problem can be expressed clearly. For the problem formulation, this chapter presents and discusses the notions of problem space and solution space, as well as the role of constraints in handling problems, and how they might affect and make a problem easier or more difficult. It points to the fact that puzzles have a special place in the history of problem solving, in that humans invent them for challenging others, for fun and entertainment, and even for studying problem solving per se in a controlled fashion.

2.1 Definitions

The nature of human problem-solving activity has been studied by psychologists over the past hundred years. In the cognitive literature, problem-solving has been defined as a higher-order cognitive process that requires the modulation and control of more routine or fundamental skills.

According to Newell and Simon [Newell1972], problem solving is illustrated by every situation where from the outset an agent does not have an obvious already available procedure to reach the final objective. In other words, a person is said to be confronted with a problem when he/she wants something and does not know immediately what series of actions he/she can perform to get it. This definition implies that situations where the procedure to reach the final objective is already available may not be categorised as problem solving.

Problem solving is the cognitive processing directed at finding solutions and performing a sequence of operations. The actions involved in obtaining the desired object could be specific or general, physical or abstract, and may include physical actions, or perceptual or mental activities.

To have a problem implies that at least certain information is given to the problem solver: information about what is desired, under what conditions, by means of what tools and operations, starting with what initial information, and with access to what resources (i.e. goal, resources, side conditions, etc.).

Newell and Simon [Newell1972] define a problem as follows: A person is confronted with a problem when he/she wants something and does not know immediately what series of

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

actions he/she can perform to get it. The desired objective can be tangible (an apple to eat) or abstract (remembering something or proving a theorem); it may be specific (for example eating an apple) or general (eating something to appease hunger). The actions involved in obtaining the desired objective (whether physical or abstract) can include physical actions (walking, writing), perceptual activities (looking, listening), or mental activities (judging the similarity of two symbols, remembering a scene, etc.).

Depending upon 'what is the task at hand', 'who is solving it', and 'the expertise of that person to solve the given task', a given "problem" at hand could either be a 'situation of elaboration' of procedure, or a 'situation of execution' in which the solution is obtained by simply following the known procedures. In other words, what defines a 'problem' (or a 'situation of execution') is not only the situation, but also the relationship between the task and the competence of the subject. A certain task that may be a 'problem' for a certain subject may just be 'situation of execution' for the other. (See figure 2.1.)

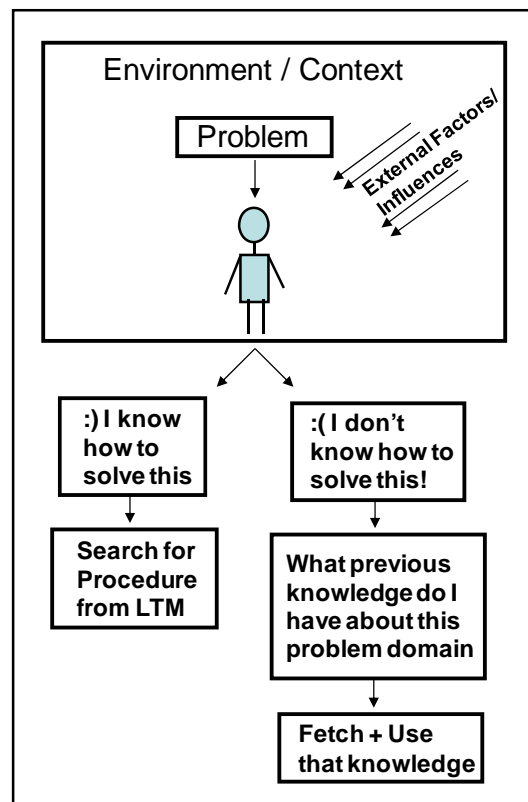


Figure 2.1: Traditional information processing point of view of problem solving

2.2 The notion of problem space: Model for analysis vs. representation elaborated by the subjects

We (humans) are generally said to be using our problem-solving skills to proceed from a given state of a problem to a desired goal state. To elaborate, one can face a problem in two ways:

- one possesses the knowledge applicable at the situation (that allows one to activate a certain procedure that may be applied in such a situation); or
- one does not possess the knowledge applicable that permits/allows one to elaborate what actions to take (i.e. it doesn't allow us to activate a certain procedure that may be applied in such a situation). The situation is thus perceived as a 'problem', and thus one must furthermore construct an interpretation of problem.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

The notion of “problem space” is initially a generic model to analyse the solving process of a problem in an information-processing framework. It is based on the idea that solving a given problem has three main components: a given state (or initial-state); a goal state (or final-state); and a set of operators which one applies while one is actively solving the problem [Newell1972]. The set of all possible operations that can be performed in an attempt to reach a solution is called the “problem space”; while solution space is defined as the displacement in search-space that may be represented by the graph of which the nodes represent the states that can take the situation based on (or followed by) the actions taken by the subject [Newell1972].

In the classical view, the problem difficulty can be attributed to two major sources:

- the amount of knowledge required to solve the problem; and
- the exponential growth of the search-space as problem length and complexity increases

It is worth noting that a difference should be made clear between the concrete realisation of the problem and the abstract description than can be made of its problem and solution space.

The notion of problem space is at the very core of the construction of this interpretation, called “representation of the problem”. Representation of the problem means constructing an interpretation of the situation that allows one to define the “search space”. This interpretation comprises three components of the problem situation: the interpretation of the initial situation or state, interpretation of the final state, and interpretation of the allowed (or legal) actions or operations that one can perform.

At certain times the operators that one would be using, and how will they be used to reach the goal state, may be fairly straightforward, while at other times they may not be.

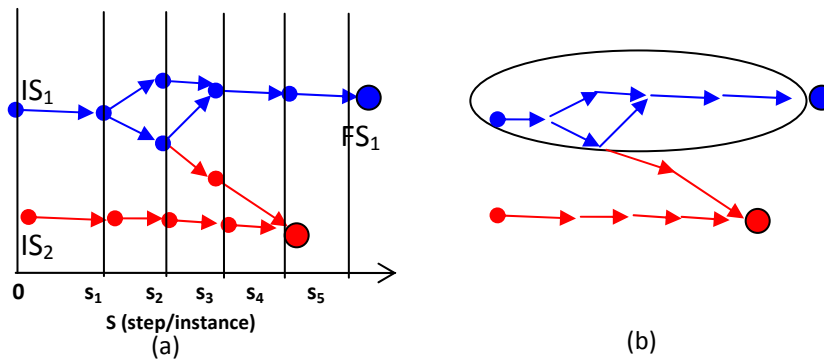


Figure 2.2: Problem Space (represented by the whole diagram), Solution Space (represented in blue), Initial States (IS_1 and IS_2) and the Final State (FS_1)

Figure 2.2 can help better visualise the idea. In this figure, the presence of an arc between two states indicates that one can pass from one state to the other, meaning that there exists a valid (or legal) action that permits to pass from one state to another. The initial state (IS) and the final state (FS) are two different nodes on the graph (the start-node and the end-node), while the solution is a path connecting the start and the end nodes.

It can be the case that the search space doesn't contain a solution state. In that case, it is not possible to achieve the solution in that search-space and the problem is insolvable. The subject must thus revise certain aspects of the interpretation of the problem, notably the ones concerning legal actions, for example by eliminating too restrictive interpretations. In figure 2.2 above, there are two possible initial states and only one of them leads to the final state. Note that first of all, it is only possible to reach the final state if one starts from the first of the two initial states (i.e. IS_1 and not IS_2). Moreover, it is possible to reach the solution at any given time, while one is in one of the states shown inside the oval (ellipse) in figure 2.2(b). But if one reaches to a state outside the ellipse, then in the current problem, it is no longer possible to reach the solution.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Of course, every problem can have its own graph with its own unique properties. Furthermore, as stated above, one finds solutions to problems based on the given circumstances and within the given parameters of the problem at hand. Before solving the problem, one must first know what it is that wishes to solve, under what conditions, under which constraints, and what available resources are at hand. While for some tasks, one may know this information explicitly, while for the others, this information may be hidden or implicit, and one may have to discover it as one goes along (the exploratory phase). When a child is asked to move a big stone from one place to the other, he/she might not know that it is too heavy for him/her to lift. The child might realise this by actually trying to lift it up, and realising that even a reasonably small stone may have an enormous mass (and weight). Similarly, knowing that a sponge is a very light material might make one think that it may be very easy to displace a very large piece of sponge (say 5 x 5 meters), until one finds by trying it out for the first time that it is in fact too heavy to be lifted unaided and bare-handed.

To achieve a certain objective, when proceeding from the initial state to the goal-state (under certain conditions), the user almost invariably passes through several intermediary steps. The 'initial-state' - the state in which the problem is presented - is sometimes visual in nature, be it on paper or as a physical object in three dimensions, depending upon the problem at hand. Real life, however, is more complex than a jigsaw puzzle, and so what exactly constitutes a solution can also change from the solvers' view point. In real life, problem solving also entails finding satisfactory solutions based on a personally defined solution-space including the goal-state, the resources available, and the constraints at hand. To explain this idea, the term *satisficing* was coined [Newell1972]. The concept is defined as setting an acceptance level or aspiration level as final criterion and taking an action based on that acceptance level or aspiration level (and which may not necessarily be the best or most optimised action in the given circumstances). What this means is that in most of the real world tasks, one sets for oneself a task (final-state or goal-state). When realising that there is a constraint or a limitation of a certain resource, one can decide to alter or modify the goal-state. In other words, this could translate to downplaying one's expectations.

2.3 Role and nature of constraints

Everything operates under certain constraints. We and other living things are all constrained on earth. The earth's gravity constrains us, and thus we cannot escape off into space. Any given problem also operates under certain constraints, which can be of different kinds, tangible, intangible, physical, etc. Physical properties of the matter may impose certain constraints on how they may be handled or manipulated. A particular nut goes only with a certain kinds of compatible bolts, for example a bolt thicker in diameter than the inner hole or a certain nut, would not pass through that nut in normal circumstances.

Constraints can also take the form of verbal rules, for example in case of the physical incarnation of Tower of Hanoi, the user cannot move the disks horizontally (sideways) directly from one bar to the other because of physical constraints; however, the other constraints come only in the form of verbal rule, such as 'moving only one disk at a time' or 'not stacking a bigger over a smaller one'. Although there is no physical hindrance and one can move two disks at a time or place a larger disk over a smaller one, nevertheless such a move is disallowed and is against the declared rules [Richard1993].

2.4 Types of problems

As per [Greeno1978], problems may be distinguished into three types, namely:

- Problems of *induction of structure*: Problems that require identifying the "relationship" characterising the group of given elements for example to find the rule governing a series of numbers and predict the next one in the series or simply to find a rule that it characterises. Here is an example of a problem of this type:
1, 2, 4, 8,

The goal here is to identify the relationship between the successive numbers and to predict the next number in the series. In the current case, the solver has to identify that

each number is obtained by multiplying the value of its immediate neighbour to the left by 2 (thus 1 multiplied by 2 yields the next number in the series i.e. 2, similarly 2 multiplied by 2 yields the next number in the series i.e. 4, and so on and so forth). Thus the correct response to the problem above is 16.

- *Transformation-of-state problems:* A transformation-of-state problem has an initial state, and one must arrive at the final state, by utilising the permitted operators that allow one to transform the given or current situation. Some examples of puzzles that lie in this category include Chinese Rings Puzzle (or other wire or entanglement puzzles), Tower of Hanoi, missionaries and cannibals, while some real-life examples include preparation of eggs (or any other meal), or making finished end products in a factory using raw materials (say, fabricating silicon-chips from silicon).
- *Arrangement problems:* In arrangement problems, one has a group of elements arranged in a certain manner at the beginning, and one must find one or more other arrangements that satisfy a certain criterion or criteria. For example when a child arranges his toys in a big basket or bucket so that they all fit inside, or when a fruit and vegetable seller does an arrangement of his fruits vegetables. Note the objective of the fruit-seller could be either compactness (to maximise space), or it could just be to arrange his things aesthetically, or both. Other examples are anagram or cryptarithmic, where anagram is a type of *word play*, the result of rearranging the letters of a word or phrase to produce a new word or phrase, using all the original letters exactly once; while cryptarithmic aka “Verbal-arithmetic” is a type of mathematical *equation* with unknown *numbers*, whose *digits* are represented by *letters*. The goal is to identify the value of each letter.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

It can be argued that situations where objects change their given form in a more permanent (or an almost irreversible manner) like in chemical reactions or where the given object is torn or broken into two are the only situations that should be qualified as transformation of state. If one accepts this definition, then making a chair from raw wood qualifies as well, since it undergoes transformation since one has to cut and carve it out some wood, and since this transformation is fairly permanent and irreversible; however, building a house from Lego bricks does not qualify as transformation-of-state problem.

We however prefer the less strict definition, whereby all those situations are to be qualified as transformation-of-state problems, in which the given object changes its form, regardless of whether the transformation is chemical or physical, or reversible or irreversible. Thus for example, if the aim is to form a complicated object (like a house or a car) out of Lego bricks, it is considered as a transformation-of-state problem.

To clarify a bit more, whenever the transformation is of the form where the goal or aim is inherently focused on the transformation of state, it is considered as a transformation-of-state problem, but in situations where the focus is not specifically on the transformation-of-state per se, but rather in the way things are arranged with respect to each other, then it is considered as an arrangement problem. The missionaries and cannibals problem is an example of this type. The problem states that: there are three missionaries and three cannibals on one side of a river, along with a boat that can hold one or two people; and one is asked to find a way to get everyone to the other side of the river, without ever leaving more cannibals than missionaries on any side at any given time. Notice that the main focus is not to transform the objects at hand but to arrange them physically; or temporally in a certain order (manner). Notice also that in this problem, it doesn't really matter in which order the cannibals cross the river, i.e. their order or their inter-arrangement with each other is not of any importance.

To recap, it really depends how one defines transformation, but despite of that it could in fact be a very slippery slope. Since the two definitions have in fact arisen by studying some examples and naming and creating a category by seeing some sort of a resemblance amongst them based on certain generalisable characteristics, and not the other way round.

However they had not been devised in a fashion that guaranteed mutual exclusion. Due to this reason, some problems may qualify under both headings. For example if we were to consider the Tower of Hanoi, its objective is to arrange the disks on the three pegs or towers in a certain fashion (thus qualifying it as an arrangement problem), however, in the process, the towers made up of the pegs and disks transforms from one form to the other (thus one could also argue that this problem qualifies as a transformation-of-state problem as well). Similarly, Rubik's cube can possibly be considered as an instance of both as well, depending upon how one looks at the problem. (Rubik's cube is described and discussed in more detail in chapter 3.)

In summary, there are many examples of spatial puzzles and games that either do not precisely lie in either of the above mentioned categories, or do not uniquely lie in one category, but in more than one at the same time.

2.5 Different ways of progress/Redefinition

While one is in the process of solving a problem, or in the event of failure or an impasse, one undergoes the process of redefinition. This may be done by means of:

- *Redefinition of the task objectives* (which could mean identifying subtasks and solving those subtasks, and trying to achieve local victories, as a stepping stone to reach the overall bigger goal); or

- *redefinition of available resources* (which could mean gathering more tools, for example buying a spanner to open up a box to further probe and explore what might have gone wrong inside of the box; or thinking out of the box and rediscovering that the possibilities, tools or operators available at hand are more than one thought, or that there exist less constraints than one had at first thought).

In the well-known case of the nine dots problem, the objective is to cross out all the nine dots by drawing four straight lines, without lifting the pencil in between (Figure 2.3).

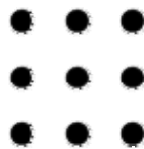


Figure 2.3 : The nine dots problem

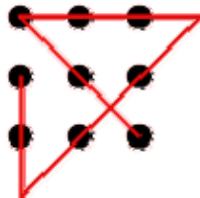


Figure 2.4 : One possible solution to the nine dots problem

One of the techniques of solving some problems is to realise that some operations are too restrictive, and that one needs to rethink (reassess) if one is implicitly placing more restrictive conditions than has been explicitly mentioned. In the nine dots problem, the reassessment/redefinition of available resources translates into a modification of the inferred rules of action or reassessing the constraints imposed upon oneself. One possible solution to the problem is shown in figure 2.4. The solution to the problem becomes easier when one realises that one has not been told that the straight lines cannot go beyond the boundaries of the imaginary square that many people see.

2.6 Other approaches to problem solving

In contrast to the classical approach to problem solving, there are other models too. For example, according to the Rasmussen's model of 'problem resolution', when confronted with a problem, the human first collects the data available at hand. This is followed by the identification of the problem. This includes diagnosing the situation, precise identification of the problem, and its description. In other words, in this step, one identifies and formalises the problem. This is then followed by building schematic solutions taking in view, the constraints of the system, envisaging and defining goals, and probably even sub-goals. And then finally comes the actual implementation of the solution [Guiost2008], [Rasmussen1980].

Rasmussen's model distinguishes three different categories of human behaviour as per the skills of the person involved and the problem at hand. The three categories of human behaviour are:

1. Skill-based behaviour
2. Rule-based behaviour
3. Knowledge-based behaviour

Skill-based behaviour represents sensorimotor performance during acts that take place spontaneously without conscious control as smooth automated and highly integrated patterns of behaviour.

The *rule-based behaviour* is applied by humans as a sequence of subroutines in a familiar work situation, typically consciously controlled by a stored rule or preset procedure that may have been memorised beforehand, derived empirically during previous occasions or

communicated from other persons' know-how as an instruction or prepared on occasion by conscious problem solving and planning.

Lastly, the *knowledge-based behaviour* is applied in unfamiliar situations, when humans are confronted with an unknown situation and they use their knowledge and its experiment to invent a new adequate solution [Rasmussen1983].

2.7 Conclusions: Puzzles and problem solving

The underlying assumption for using puzzles for studying problem-solving is that they are relatively simple tasks, but they capture some interesting properties of "real world" problems, and the cognitive processes underlying participants' attempts to solve those simple problems are representative of the processes engaged in when solving "real world" problems.

Just like any other problem, since a puzzle is something that needs to be solved, the first idea is obviously that solving a puzzle is a problem solving activity too. According to [Butler1994], a puzzle is essentially a "problem or enigma that challenges ingenuity". Puzzles are often contrived as a form of entertainment, but they can also stem from serious research problems. Solutions to puzzles may require recognising patterns and creating a particular order. In a typical puzzle, one may be required to piece together objects in a logical way in order to come up with the desired shape, picture or solution.

While the present chapter has introduced the idea that puzzles are kinds of problems, and that puzzle solving is a problem solving activity, the next chapter discusses a special category of puzzles – the spatial puzzles. It discusses many types of puzzles that lie under this category since the 3D object manipulation task that we chose for our experiments also lied

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

under this category of puzzles. The chapter also proposes a categorisation of these spatial puzzles. Our interest is to show a categorisation of puzzles based on their physical and spatial aspects. The physical structure and form of the components of a puzzle are of prime importance from the point of view of the interaction, since the way the users would interact with the objects depend upon the way they perceive them, and the way they perceive them depends upon the shape and form of the objects in the puzzle. The chapter will also consider the fact that computerisation is changing the very nature of puzzles, not only by increasing their varieties and possibilities, but also by renewing the issues of their interface design.

3 Chapter 3: Spatial Puzzles: An Overview, Issues

Involved, and New Perspectives

This chapter discusses a special category of puzzles known as 'spatial puzzles'. Spatial puzzles themselves form quite a vast domain, and in the existing literature, there is a dearth of a complete and coherent review. This chapter, thus, in addition to giving a broad overview, makes an attempt to present spatial puzzles in a coherent manner, and proposes a taxonomy of these puzzles. This is, to the best of our knowledge, the most elaborate and consolidated one, from the point of view of puzzles' physical features and their interfaces. Physical puzzles of historical importance are the major focus. Since the problem that we used for our experiments is a kind of spatial puzzle too, thus our puzzle (strictly speaking a puzzle that is very well known and very similar to our puzzle) is discussed along with other spatial puzzles. We can thus see where our particular puzzle lies in the broader picture.

3.1 Introduction to spatial puzzles

Spatial puzzles are a kind of puzzles, that consist of objects or pieces that must be fitted into a specified configuration [Butler1994]. This definition fits well with our earlier definition of puzzles (in general), a puzzle is essentially “a game or a mental challenge that one has to think about carefully in order to answer or solve it” and solving a puzzle is a problem solving activity as well. Thus all spatial puzzles are also puzzles but not vice-versa. Spatial puzzles are in fact puzzles that have a spatial aspect associated with them. Famous examples of spatial puzzles include: Jigsaw, Wire Puzzles, Rubik’s Cube, Tetris, and Sliding Puzzles, etc.

There are two broad classes of space-related problems: those involving large-scale environments and those involving the perception and manipulation of small-scale configurations of objects. Both of these are studied separately. Spatial problem solving may generally refer to any of the two following categories:

- a) Spatial problem solving in large scale environments, which includes movement in large spaces, for example small and large-scale navigation or wayfinding. (Spatial problems involving large-scale environments are studied from the perspective of wayfinding as discussed above, or from the perspective of navigation of environments, for example [Denis2007], which may further be classified into open spaces and closed spaces).
- b) Spatial problem solving in small-scale (or fixated) environments, which refers to manipulating the configurations of objects in one’s immediate space, for example, being seated in one place and manipulating objects within one’s immediate reach or surroundings. (See [Butler1994], [Abbasi2009-a] for more on such type of spatial problems. In contrast, to the previous category, spatial problem solving in smaller scale environments involves situations where individual elements must be combined to form larger configurations [Butler1994]. It is precisely under this second category that most of the assembly tasks are found.)

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

There is a wide variety of spatial puzzles. In an effort to better visualise and organise the enormous range of spatial puzzles that exist, we propose that they can be organised across three broad categories, namely Dissection Puzzles, Mechanical Puzzles, and Mazes. Thus the following is a further sub-categorisation of the second kind of spatial problems as discussed above, namely 'Spatial problem solving in small scale (or fixated) environments'.

Table 3.1 categorises the various famous spatial puzzles. (The first column contains the super-categories; the second one, the categories; and the third one, the concrete puzzles). This categorisation is based on the puzzles according to their physical and interaction properties. As was mentioned in the previous chapter, we think that such a classification, i.e. one considering the physical dimension of the puzzles, is a pre-requisite to understanding how the users would perceive or interact with them.

Table 3.1 : Table of spatial puzzles

SUPER-CATEGORY		GENERIC (TERMS)	CONCRETE (ILLUSTRATIONS)	
Dissection Puzzles	2D Dissection Puzzles <i>Related concepts: Polyominoes; Polyamonds</i>	Tangrams 2D →		2 D Puzzles
		Jigsaw (2D) →		
		Coloured Tiles <i>[are Tiling Problems]</i>	MacMahon's Coloured Tiles	
			Tetris	
	3D Dissection Puzzles	Box-Packing Puzzles	Soma Cube	3 D Puzzles
			Mikusinski's Cube	
			Bedlam Cube	
		Coloured Cubes	Instant Insanity	
	Burr Puzzles <i>[are interlocking puzzles]</i>			
Mechanical Puzzles	Sliding Puzzles	14-15 Puzzle (aka the 15 Puzzle)	3 D Puzzles	
		Rubik's Cube		
		Rubik's Magic		
		Rubik's World (like Rubik's 2x2 mini cube in the world's shape)		
		Twobik's cube (mini rubik's 2x2 cube)		
		Pyraminx Puzzle (like rubik's 2x2 mini cube in the pyramid's shape)		
	Wire and other disentanglement Puzzles (similar: Ring-String-Ball Puzzle)	Chinese Rings Puzzle		
		Heart and Arrow Puzzle		
		Double-bow		
		Loony Loop		
	Other Abstract Puzzles	Peg Solitaire		
		Rubik's Snake (folding puzzles)		
	Yoshi's Cube (folding puzzles)			
Mazes	Mazes 2D-single screen	PacMan	2 D Puzzles	
	Mazes 2D-Multi room-screen	Rogue (Hack): dungeon crawling PC game – [2D Maze]		
		Go off the Earth puzzle [2D abstract/sliding puzzle]		
	Mazes Virtual 3D ()	Doom		
	Counter Strike			

We are yet to find a categorisation of puzzles on these two criteria combined. We believe that categorising puzzles from this perspective is in line with our goal: analysis of puzzles from the perspective of their interfaces in an effort to improve (or enhance) users' performances for a given puzzle. This could be either for 'improving their success in a given task', or for 'minimising the difference in their intentions and their actions or error rate, i.e. were they able to perform a certain operation as they desired to').

3.1.1 Dissection puzzles

Dissection puzzles are basically tiling puzzles where a solver is given a set of pieces that must be assembled in a certain way to produce a desired shape. Tetris is a famous example of a computer-based dissection problem, the goal of which is to fill the given field without spaces with the help of various figures composed of four square blocks. Other examples of dissection puzzles include *box-packing puzzles* (for example Soma Cube and Bedlham cube), *tangrams*, *jigsaw puzzles*, *coloured-tiles*, and *burr puzzles* etc. (See below for a more detailed description of these puzzles).

To dissect means literally to cut into pieces. From an abstract viewpoint, most of them have been characterised as *polyforms*, i.e. forms that arise from edge-to-edge gluing of several copies of a simple shape, such as a square, a cube, an equilateral triangle, or an equilateral triangle cut in half. Polyform is the general term. The specific term *polyominoes* refers to 2-dimensional (2D) square-based forms whereas the term *polycubes* refers to as 3-dimensional square-based (3D) objects. The same logic as the one presented afterward can be applied to other shapes. For example, a diamond consists of two joined equilateral triangles and adding more triangles produces higher order sets of *polyamonds*.

3.1.1.1 Polyominoes: 2D square-based forms dissection puzzles

A *polyomino* is a figure lying in the same plane built by joining together identical square-based forms, edge-to-edge. The name polyominoes is composed of two parts, namely *poly* and *ominoes*. It could contain any number of squares, and the structure could be called *n-ominoes* depending upon the number of squares that it contains. For example *dominoes* are built of two squares which form a 1 X 2 tile and there is only one possible form for dominoes (Figure 3.1). *Triominoes* (aka Trominos) are polyominoes made of three squares that can take two distinct forms, a 1 x 3 rectangle and an L-shape (Figure 3.2). *Tetrominoes* (aka tetraminoes) are polyominoes composed of four squares, and could have five unique forms. These sets get more versatile as their base number increases, for example 12 unique forms for *pentominoes* (which are polyominoes based on five squares), 35 for *hexominoes*, and 108 for *heptominoes*.

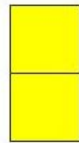


Figure 3.1 : Domino

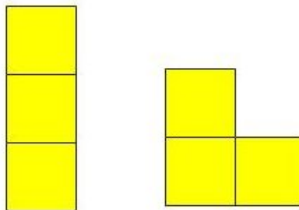


Figure 3.2 : Two different possible forms for triominoes

Although strictly speaking, tetrominoes could have only five unique forms (it could be claimed that they have 7 different forms) when rotation and reflection operations on

the given forms are not allowed like it is the case in the game of *Tetris* (where the flip operation is unavailable) (Figure 3.3)



Figure 3.3 : The seven different forms of tetrominoes (comprising of four unit squares joined together) used in Tetris

3.1.1.2 Most current examples of polyominoes

Tangrams are dissection puzzles consisting of seven pieces (five right-angled triangles, one square, and one parallelogram) called tans. The objective is to form a specific shape with the seven pieces in a non overlapping way.

Jigsaw puzzles are perhaps the most popular form of dissection puzzles. A large image or picture is cut into various smaller pieces of various shapes of different forms and is presented to the user in a scrambled form, and the objective is to assemble them back to form the original image. In most cases, the jigsaw puzzles have a guiding image to help the user to assemble them back in the original form [Demaine2007].

Edge-matching puzzles (that first appeared in the 1890s) are similar to jigsaw puzzles – i.e. they are similar in spirit to jigsaw puzzles, but there are some important differences in details. The goal of an edge-matching puzzle is to arrange a given collection of several identically shaped but differently patterned tiles (typically squares) so that the patterns match up along the edges of adjacent tiles. Typical patterns range from salamanders to frogs to insects, but in their simplest form each edge of a tile is simply coloured one of several

colours, and adjacent tiles must be coloured identically along their common edge [Demaine2007]. Figure 3.4 shows a concrete example of such a puzzle with insects.

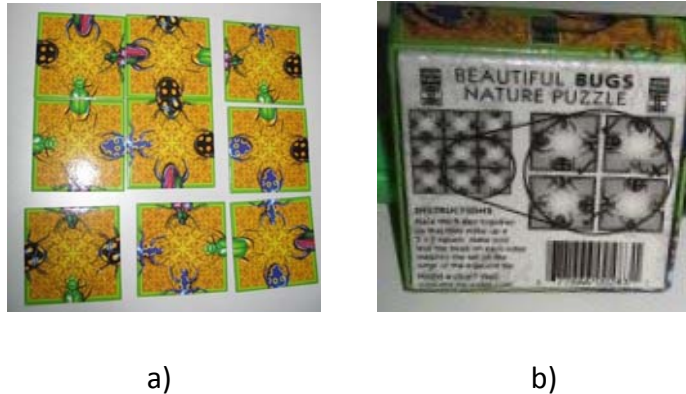


Figure 3.4 : An edge-matching puzzle with insects

Edge-matching puzzles are more challenging compared to standard jigsaw puzzles. Firstly, there is no global image to guide the puzzler, but more importantly, the fact that the pieces do not have an irregular shape or border gives no additional cue (or clue) to the solver, as to which two pieces are likely to fit each other. Although small and subtle, such differences play an important role in determining the level of difficulty of this puzzle.

In the case of jigsaw puzzles, sometimes, irregular projection/shape of the individual pieces guide the user towards matching and fixing the right piece with its corresponding ones because there are only a handful – and at times, just one – piece(s) that complementarily fit with the other. At other times, the unique match of the shape is one of the factors that helps reinforce the user's confidence that the two pieces should be put together.

Coloured-tiles consist of tiles or pieces required to be assembled in a certain specified form. In the case of coloured-tiles, they may be assembled into a number of shapes following the

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

rule that adjoining edges must be the same colour. Coloured-cubes uses cubes, as is evident from their name, instead of tiles, while the basic principle remains the same – that of assembling the cubes into specified shapes while following certain rules.

Tetris is a computerised game involving randomly selected shapes falling from the top of the playfield, one at a time. These shapes are comprised of 7 different patterns of tetrominoes. (See Figure 3.3) The goal is to prevent those shapes from piling up to the top and ending the game. To do this, the player must move and rotate the shapes as they fall, attempting to fit them together. Whenever a row of blocks is completely filled with coloured squares, it is removed and all blocks above it drop down to fill the empty row. If the player is unable to fill lines completely, the tetrominoes (or, as they are called in the game, tetrads) will stack up and eventually reach the top of the playing field. The game ends when a new tetrad that is placed at the top of the playing field is unable to drop at all due to the filled blocks.

Tetris has been subject to many changes throughout releases since the 1980s. Thus it is difficult to place a standard on the game. In the standard the PC-based game-puzzles, there is a playfield measuring ten spaces across by twenty spaces down. The player can rotate the falling tetromino ninety degrees at a time within the plane of the playfield by pressing the counter-clockwise or clockwise rotation buttons, provided the piece has room to rotate.

The player can shift the falling tetromino sideways one space at a time by pressing the left or right arrow, provided the piece has room to move. Pieces cannot shift through walls or other blocks. In the classic version of the game-puzzle which exists on the PC, which has been tried by the author, the controls are as follows: the left and right arrow keys permit to rotate, the down key makes the block fall down (faster or instantly, both modes are considered to be fairly standard). (The official website of Tetris is: <http://www.tetris.com/>)

3.1.1.3 Polycubes: Some examples of 3D cube-based dissection puzzles

Polycube is a polyform with a cube as its base form, rather than squares as is the case in polyominoes. Also just like polyominoes, polycubes too can also be categorised as *tricubes* (structures containing 3 minicubes), *tetracubes* (4 minicubes), *pentacubes* (5 minicubes) and so on and so forth. The most current examples of polycubes are the Soma Cube and the Bedlam cube. They have been also called *box-packing puzzles*, in that they are 3D blocks that need to be assembled according to a specified given form. The form is usually solid (i.e. without holes in it).

[Demaine2007] considers polyform packing puzzles as ultimate forms of jigsaw puzzles. By ultimate, he is probably implying that these puzzles are 'hard' or 'difficult', since he backs up his claim by reasoning that not only is there no guiding image and two pieces fitting together say nothing about whether they are together in the final solution, but also two pieces can fit together in several different ways. This difference in interface makes another major difference i.e. only completing the entire solution guarantees correctness of any local part of the solution.

The *Soma Cube* is a well-known box-packing puzzle invented in 1936. It consists of 7 different blocks or components, each of a different shape. The goal of the puzzle is to combine those 7 blocks to form a solid regular cube, with each side equalling 3 units. The original version of the Soma Cube contains seven unique blocks, one of which is a tricube, while the rest are tetra-cubes, totalling $((3 \times 1) + (4 \times 6))$ or $(3 + 24) = 27$ mini-cubes (see figure 3.5). As per [WolframMathW] and [Gardner2008], there are 240 standard solutions to the standard Soma Cube (if one does not include rotations or mirror images).

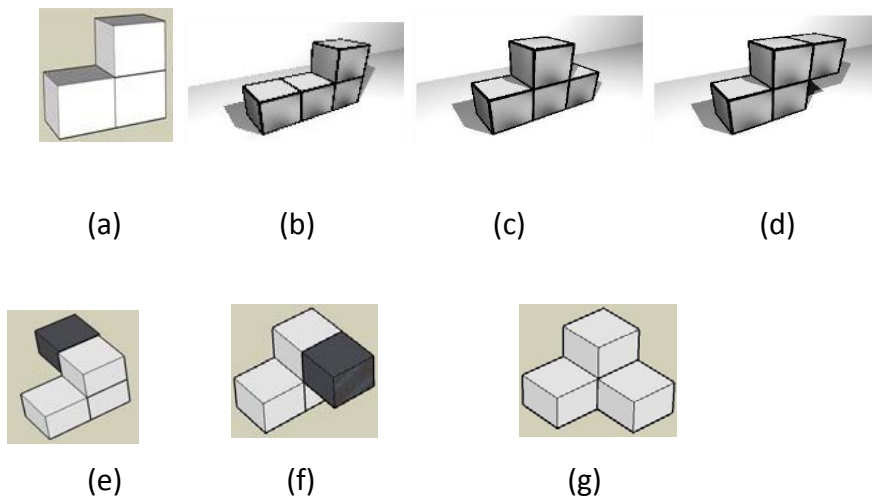


Figure 3.5 : The seven blocks in the Soma Cube (that are used to build up a 3x3x3 cube)

There is another puzzle that is quite similar to the Soma Cube that goes by the commercial name Bedlam Cube. It contains 13 pieces, and transforms into a 4x4x4 cube. It has 19,186 distinct solutions, excluding rotations and reflections.

3.1.2 Mechanical puzzles

Mechanical puzzles are generally hand-held objects that must be 'manipulated' to achieve a specific goal [Butler1994]. They have a manipulation and a "hands-on" mechanical aspect associated with them. Please note that *mechanical puzzles*, which are defined here as a subcategory of spatial puzzles, must not be confused with the broader term *mechanical problem-solving*. The definition for mechanical problem solving that we prefer to use is the one used in [Hegarty2004] which includes all forms of *physical* problem solving where the focus is on spatial or movement aspects, as well as those kinds of problem solving that can mentally be conceived in the mind's eye as having a physical form and possible movement. Thus this definition of 'mechanical problem solving' is much broader than that of 'mechanical puzzles' in that it encompasses any activity, whether physical or mental, that may have a mechanical aspect associated with it within its domain, for example as mentioned before, when a car mechanic is trying to mentally figure out what might be

wrong with the brake system of a car, then that is also considered to be a mechanical problem solving activity. When you are asked to imagine three gears of the same size put next to each other, and are asked to simulate them moving in your mind, this too is an example of a mechanical problem solving activity. As per this definition, an abstract mathematical equation or algebraic problem solving would thus *not* be considered as mechanical problem solving. Following are some examples of mechanical puzzles:

- i) *Sliding piece puzzles* consists of a set of tiles, held in a frame, that are set to an initial configuration or are randomised and then must be moved, one at a time, to a new arrangement.

- ii) *Sam Loyd's "14-15" puzzle* or The Fifteen Puzzle is a sliding pieces puzzle made of square blocks, labelled with the numbers 1 through 15 (figure 3.6). The blocks are placed in a flat container with room for 16 squares, which leaves one cell empty. Without lifting a block out of the container, the player is to rearrange the blocks into a different pattern.

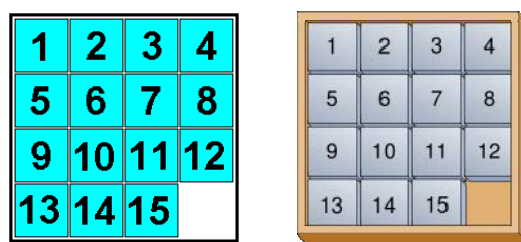


Figure 3.6 : Sliding Puzzle

- iii) *The Rubik's Cube* is a 3D puzzle which was invented in 1974 by Hungarian sculptor and professor of architecture Ernő Rubik. In a typical cube, each face is covered by nine small squares of different colours. When the puzzle is solved, each face or side of the Cube is to have a distinct colour (figure 3.7).

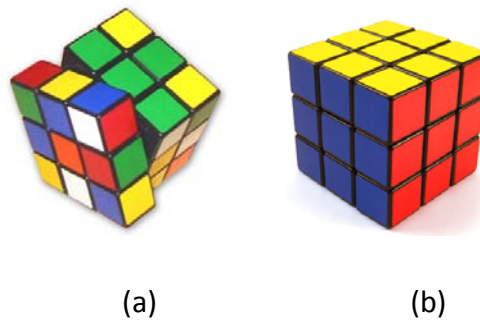


Figure 3.7 : Rubik's Cube

Note that the examples of box-packing puzzles (in section 3.1.1.) also possess the characteristics of mechanical puzzles as well.

All of the above examples of mechanical puzzles belong to the sub-category of *sliding puzzles*. However, there are other kinds of mechanical puzzles that require manipulation, although they do not fall under the dissection puzzles category, such as those that lie under the sub-category “wire and disentanglement puzzles” amongst certain other abstract puzzles, of which *Peg Solitaire* is one example.

iv) *Peg Solitaire* is a board game for one player involving movement of pegs on a board with holes with the entire board with pegs except for the central hole. The objective is to empty, making valid moves, the entire board except for a solitary peg in the central hole.

3.1.3 Mazes

Some people consider *mazes* as a special category of puzzles in their own right. They are also a comparatively new phenomenon. First of all, they are generally PC based, and second of all, they are almost always interactive in nature. This is unlike some puzzles where the subject solves it by means of an instant insight, rather than exploring the maze. The goal of mazes (or maze-type puzzles) is to find the correct route to some hidden treasure or

something similar. Examples of maze-type puzzles include Dove, the Atari game ET, Doom, Rogue (Hack), etc. (See also sections 3.3 and 3.4).

As it turns out, all the puzzles that lie under this category are essentially computer games, and the computer gaming industry is currently one of the fastest growing industries in the world today. Due to this fact, a number of new puzzles under this category continue to be introduced all the time. Their thorough review is beyond the scope of this thesis.

3.2 Disclaimer to finality of categorisation

We do not claim this categorisation to be a perfect one. It will surely not be able to accommodate all puzzles for all times to come. Some puzzles will lie in more than one category (multi-inheritance), while others may not lie in either. And one obvious and simple reason for this is that no one creates a puzzle keeping in mind a categorisation. One of the factors which have already made it possible to create the choices of the puzzles much richer and creative is the advent of computer technology. It is a topic of discussion in its own right and we will explain it in more detail under a separate heading.

3.3 Computerisation changing the nature of puzzles

The advent and proliferation of computers allowed one to replicate and create virtually unlimited variations and forms of verbal and visual puzzles. In fact, in recent times, puzzles have taken other forms that were never before possible.

As you will see from the examples below, the way puzzles are conceived can potentially fundamentally be altered through computerisation at one extreme, or they can remain unaltered on the other extreme.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Computerisation can be used at different levels from simple replicas of paper-based puzzles (i.e. scanning the old paper-based tests and presenting them on the computer as closely as they existed on paper), to others that are impossible to be conceived on paper. The computerised puzzles may range from those having command line interface to ones having visually rich interface; from completely passive to fully interactive; from 2D to pseudo or quasi 3D to real 3D; and from utterly bland and detached from reality to one that is fully immersive.

As just mentioned, taking puzzles on computer take the form which is not very different in essence from the type of problems that one solves in IQ tests taken with paper and pencil (figure 3.8).

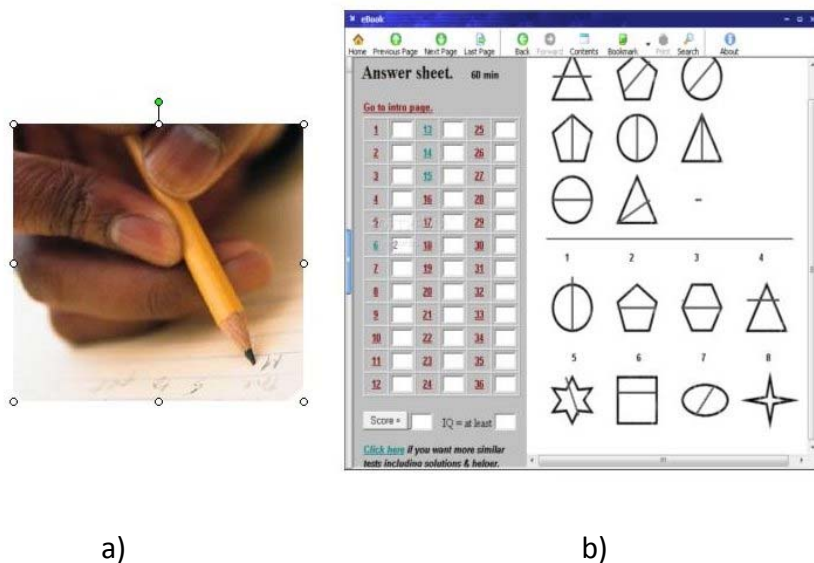


Figure 3.8 : Examples of both paper/pencil and PC screen versions of same/similar tasks. a) Shows a test being taken with paper and pencil; b) Shows a similar test being conducted on a PC monitor

However, with the advancement in computer technology, the most important development which has taken place is what could be called interactive puzzles, as opposed to static or offline puzzles. In fact, many of these puzzles cannot be crisply categorised in any of the traditional problem solving categories. For example, the maze puzzles are spatial and interactive in nature, i.e. there are issues of placement, movement and/or timing involved.

But these puzzles can have situations and modus operandi that is unlike any of their predecessors.

One main reason for this is the fact that earlier on, access-to/interaction with dynamic puzzles (puzzles containing dynamic objects) was only possible by means of physical puzzles. This fact also meant that one always remained in touch with the physical world and its related constraints (and never lost touch with it). That is, even while one was solving puzzles in the physical world (whether in the real world environment, or in laboratory environment), one was required to operate within the physical and mechanical constraints of the real world. Thus if an independent object was thrown up in space, it always first decelerated (due to gravity) and then came down.

In fact, computerised puzzles form a category in their own right, and their characteristics are very different from any of the earlier defined categories.

3.4 Butler's categorisation of spatial puzzles

Probably the most quoted work related to Spatial Puzzles is the review by [Butler1994] where he proposes to categorise spatial puzzles. Butler defines spatial puzzles as ones consisting of objects or pieces that must be fitted into a specific spatial configuration. Butler has discussed 22 concrete examples of spatial puzzles and placed them under several headings, which include various concrete examples of Tangrams, Jigsaw, Box-Packing and other dissection puzzles; and Sliding and other mechanical puzzles, etc.

From this first review of Butler's list of abstract and concrete spatial puzzles, we extracted a classification (please refer to table 3.1 above for our proposed classification) in an attempt

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

to come up with a hierarchical and a more complete list of other puzzles and categories than the ones introduced in Butler's review.

Butler's presentation of puzzles is more descriptive than analytical. It appears that Butler has categorised all the puzzles that he has discussed either by means of their appearance or goal, or by finding one common characteristic between them while ignoring the others and examining them holistically and from different viewpoints. For example, under the heading 'mechanical puzzles', he has placed together Tower of Hanoi, wire puzzles (for example Chinese rings puzzle) and other transformation-of-state puzzles; whereas under the heading 'box-packing puzzles', he has placed Soma Cube, Mikusinski's Cube, and other arrangement puzzles – apparently those having a free-space aspect associated to it, and their distinct 3D characteristics. We think that these puzzles mentioned under the heading of box-packing puzzles also qualify as mechanical puzzles.

A typical Rubik's cube consists of six faces, and each of those faces is made up of nine coloured squares. When the puzzle is solved, each face of the cube consists of a different solid colour (for example green on one side, red on the second etc.). This puzzle too has a classification problem in our view. [Butler1994] has placed the Rubik's cube under the heading of sliding puzzles, although this is not the only aspect associated with that puzzle.

Furthermore, instead of making an attempt to introduce any coherent framework or a hierarchical roadmap to describe the puzzles, Butler has chosen to group together similar puzzles, and presents each of them, one by one. The author has also not always been very clear as to whether all the categories are mutually exclusive; or whether there are certain categories of puzzles that are broader than others and contain certain other categories within them; or whether a given puzzle can be classified under more than one category, which in our view should be the case.

Thus while this review is to the author's knowledge the most complete and updated review of its kind, it still leaves the reader with a lot to be desired. Our proposed table (Table 3.1) above is in fact an attempt to sort all the puzzle descriptions and explanations in a more lucid and organised way. In our categorisation, any given specific puzzle is placed either under one and just one category, and never under two. This, we think, would aid the reader see all the puzzles in the bigger perspective more easily. As it turns out, the puzzles have been traditionally classified in a rather free-styled manner (for example, names of many categorisations are listed at [PuzzlesCat]), and hence confusion prevails about which puzzle is to be placed under which category in case of conflict. In our case, whenever there seemed to be a conflict and a puzzle seemed to qualify under two categories, a decision was made as per the best judgement possible based on the reviewed resources. Of course, one needs to start from somewhere, which serves as a starting ground to later on improve upon it.

3.5 Other approaches to 3D problem solving in related domains

While this work focuses predominantly on 3D puzzles and puzzle solving to investigate problem-solving processes and mechanisms, other works have approached some facets of the problem-solving related issues in slightly different directions. They include those who have solely used paper-based tests for measuring components of intelligence. Psychological studies on intelligence have exploited representations of spatial puzzle to measure some of its components. For example, [Shepard1971], [Shepard1982] has shown that to solve a problem involving identifying different objects on paper, humans mentally rotate those objects in their minds. Other works focus on measuring human performance in different spatial paper-based tests. For example, the time it takes to identify if a picture matches another one correctly, or whether it is its mirror image, is directly proportional to the angle of rotation of the visual object in question. Some results also indicate that there is a difference in performance in such tasks involving mental rotation between males and females (see [Shepard1971], [Vandenberg1978], [Zacks2005]). Some have used 3D blocks and figures in their experiments, for example [Eigler1930] who used Kohs cubes.

While some have focused on spatial abilities, from the aspect of differences at different scales, or individual differences [Hegarty2004], [Kato2003]. The focus has varied from an investigation of whether there exists a correlation between spatial and mathematical abilities and human performance in 2D and 3D tasks [Ho2006], to spatial visualisation and mechanical reasoning [Keehner2006], and to wayfinding (the ability of humans to find their way in a 3D space [Hegarty2006], [Kato2003], which could be very close to some of the maze puzzles discussed above). Still others touch upon interface and design related issues, for example [Hinckley1994], who focuses on free space 3D user interfaces.

We, on the other hand, are interested in using puzzles as a complex and multidimensional tasks. In our view, a spatial puzzle is an assembly task. According to Rasmussen's model, it would be a skill-based behaviour for those who already know its solution, and thus the most complex processes at work for this assembly task involve the user's sensorimotor skills. On the other hand, for those who do not know how to solve it, the need to apply reasoning requires either rule-based or knowledge-based behaviour. Since our objective was to examine subjects who have not already solved the puzzle presented to them, thus the task would mainly be for them a cognitive task.

3.6 Conclusions: Towards more coherent categorisation of puzzles

Interestingly, most works on puzzles (including Butler's work) give consideration to the *users* of puzzles. However, the listing of the puzzles is not always presented in a consistent or coherent manner, nor is there any attempt to list all the puzzles that are discussed in an overall master table which would clarify the position of and ideally serve as a comparison or reference point for all new and upcoming puzzles in the category of spatial puzzles.

So on our part, while an attempt has been made to put forward coherent and consistent categorisations and sub-categorisations, this categorisation is also based on the spatio-

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

physical aspects of puzzles, which in our view is the pre-requisite in determining the users' perceived ease or difficulty level of a given puzzle. While in the current chapter, many of the well-known spatial puzzles have been discussed, and we have tried to categorise them with respect to their physical (spatial) and interactive aspects, the next chapter explores and discusses more specifically the interface-related issues at a deeper level.

4 Chapter 4. Interface, Presentation Medium and Constraints

This chapter discusses one particular aspect of spatial puzzles which we think plays an important role in terms of their usability and in making any puzzle easier or more difficult. The chapter shows how the same problem can be presented in different ways, for example verbally, on a paper, on a computer screen, in a virtual environment, or physically in real-life, etc. Various factors can play a role in the user's performance. One of them is the different ways the same problem is presented, and the impact it has on the user's performance. A problem perceived differently by the user implies different strategies in solving it. Furthermore, the objective is to understand how rules and constraints play a role in making a given problem easier or more difficult. We also try to make the case that the very logic of categorisation of puzzles can fall apart, given a few changes in the interface. And finally, we emphasise the importance of interfaces. For example, providing an interface that aids better visualisation, which in return provides cues or insights to the user, may help attaining the correct solution.

In the previous chapter, we defined and categorised puzzles and placed them in distinct categories. In this chapter, we will try to analyse a level further, by discussing what sets a given puzzle apart from the others in terms of difficulty, and what are the human factors that make them easier or difficult to solve? In our view, there are two kinds of factors involved. The first of these factors has to do with the problem-space and the solution space. The second factor includes details of the implementation and the interface of the problem, i.e. the presentation medium (verbally, on paper, etc), the operators (i.e. what kind of operators/operations are available, for example move up, move down, rotate, etc.), and the controls (for example, pressing the left key moves the given object to the left, etc.). This chapter discusses and builds upon all these concepts in some detail.

4.1 Presentation medium, interface, controls

A factor that helps distinguishing one puzzle from others (or classifying them) is the *presentation medium* of the problem. The problem can be presented verbally, on paper, on a computer screen, in a VE, or physically in real-life, etc.

Before we talk about the various aspects of the presentation medium and interface and controls, and how they might have an impact on the perception and difficulty of the problem, we would first like to discuss a few puzzles. Once these puzzles have been discussed and defined, we will then be in a better position to communicate our point in the context of those concrete examples.

Communicating the final state to the user can take many forms: It may either be communicated verbally, by means of picture(s) or photo(s). Another instance of the puzzle in its solved-state (or final-state) or any other physical metaphor may be used, or any combination of these, as discussed in Chapter 2. Of course, this is not an exhaustive list, and one could think of other ways of helping the user visualise the final state, like making a

projection of the final state on a computer screen, a wall, or a 3D image in a virtual environment (which may be either completely static or fully interactive).

The various interface-related factors associated with puzzles will now be discussed. One shall first discuss puzzles like Tetris, after which we would follow our discussion on the factors of interface in general and presentation medium in particular, in some detail.

4.2 Rules and constraints

There are certain 'rules' and 'constraints' that must also be followed to solve the problem and to reach the solution. Almost every (physical) problem has a given set of rules that comes with it. For example, puzzles usually come with a set of instructions. These instructions may contain the final goal (with or without the help of a diagram) and a set of rules that must be followed while solving the puzzle.

Rules are essentially conditions expressed verbally (or in written form), that must be observed or followed while solving the puzzle. For example, some instructions may define the overall goal, while others essentially consist of the list of the rules that must be observed while executing the task.

Notice that in the case of rules, the user has to him/herself remember to abide by them, and that it is possible to violate them in principle if one wishes to do so! In fact that is the exact reason why the rules have to be explicitly stated – because it is possible to violate them.

Constraints, in contrast, are something that need not be explicated, but they just exist, for example in the form of physical barriers, and under normal circumstances it is very difficult

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

to violate them. The user is forced (or strongly guided) to execute his/her movements in accordance with these constraints.

Degree of freedom is another oft-used concept to describe the "constraint-ness" or the freedom available in a given situation. In the real world, in the absence of any constraints, we are said to possess six degrees-of-freedom (6 DOF) in terms of a movement of a dot (or a rigid body) within a 3D space. In 6 DOF, there are three degrees of freedom for making longitudinal displacement or movement (aka translation); and three orthogonal directions for rotations. DOF, as described by [MacKenzie1994], may be phrased as possibilities of movement of a certain type, while all other kinds of movements are restricted. By this definition, a human hand has 27 degrees of freedom.

As another example, Rubik's cube gives 3 degrees of rotational freedom (i.e. the rotation along the top slice of the cube, the right slice of the cube, and the front slice of the cube) for each of the its 26/27 visible pieces. One may argue that other translational degrees of freedom also exist, as one can, in principle, move the whole cube in all three directions; but since moving the whole cube does not contribute towards the solution of the cube, and which is why that would not count.

The rules and constraints may be used as a control or tuning factors. Thus, in simple terms, they make (or may be used to make) problems easier or harder to solve. They can impact the level of difficulty of a puzzle in two ways: Firstly, that of increasing or decreasing the level of difficulty of the problem (this attribute is mostly associated to rules); and secondly, that of making it sometimes more inconvenient or in some cases, guiding the user towards the solutions (mostly associated to constraints). This impact can be also on different dimensions of the problem solving processes.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

On the one hand, the constraint in Rubik's Cube is that it is always intact (meaning that its component cubes cannot be taken apart), and it is that what makes it difficult to solve. If all of its 27 cubes could be taken apart and if it were possible to put each one of them back in place one by one, then the Rubik's Cube would become a trivial problem to solve. On the other hand, decreasing the constraints to a certain degree may not always necessarily increase the difficulty level. For example, imagine that in the Rubik's Cube, a longitudinal slice of the cube could be rotated in isolation. This would allow the cube to be deformed, and it could make it more difficult to put it back in the cubical-form again.

The rules and constraints are important factors in determining the level of difficulty of the problem. It depends upon where (the given situation) and how they are applied. It is their combination with the problem space and interface that really defines the level of difficulty of a problem.

For example, quite different scenarios can be used in the description of a spatial problem solving task [Simon1976]. However, having the same solution-space does not mean that the different scenarios are entirely equivalent at the cognitive level. Depending upon the variations, they may be perceived differently, and subjects may exhibit strong performance differences as well. Another way of saying that is to state that different presentations of the same problem lead to different perceptions, performances, and solving strategies. In some cases, essentially, they can even be treated like entirely different problems altogether at the cognitive level.

[Kotovsky1990] shows that different isomorphs of the "Chinese Ring Puzzle" with exactly the same search-space lead to strong differences in mean solution times depending upon the nature of activities, ranging from an average of 10 minutes in the case of a certain digital version to more than two hours in the case of a certain analog isomorphic version! This leads

to the obvious deduction that search-space is not the only critical factor for the difficulty level of the problem (see also [Kotovsky1985], [Clément1997]). We tend to believe that the interface of the puzzle certainly plays a critical role in not only how the problem is perceived, but also how it is approached and solved.

4.3 Interface-related characteristics of spatial puzzles

As the previous example shows, the interface-related aspects of puzzles are of pivotal importance. This one compound-factor can fundamentally alter the way puzzles are categorised, and completely transform the way problems are viewed, and attacked by the user.

Sam Loyd's "14-15" sliding pieces puzzle (a puzzle which requires sliding square pieces in essentially two dimensions) and Rubik's Cube, which requires rotating 27 mini-cubes that form up a bigger 3 x 3 sized cube, are also two entirely different puzzles, from the interface (and hence also the cognitive problem-solving) view-point. See 3.1.2 for introduction to the two puzzles.

One is of the view that, of course, both puzzles do have a seemingly common feature, in that, one has to physically move certain pieces or blocks to solve the puzzle, but fundamentally, at the cognitive level, there exist more differences than similarities.

In the "14-15" sliding puzzle, which is a 2D puzzle, it is impossible to rotate a sliding square piece because of its given interface and thus there is no concept involved of the orientation of each block (i.e. either rotation on the same plane, or making an orthogonal movement of a piece with respect to a given plane); whereas the Rubik's Cube allows both the rotation of

pieces and movement of pieces in different planes, rendering it to be an entirely different kind of a problem altogether.

By comparing the puzzles mentioned above, the point which one would like to make is that Butler's paper is a wonderful resource and a good starting point for finding descriptions and explanations of many spatial puzzles together in one paper. However, it does not provide any critical cognitive-based analysis for example, on interface-related aspects, and their impact on the way problems are viewed or solved. Nor does it offer any coherent hierarchical categorisation of spatial puzzles. We believe that filling this gap will lead to a better understanding of the nature of activities associated with problem-solving of spatial puzzles.

4.4 Visualisation

Visualisation has been an important cognitive resource in human discovery and invention and visualisation techniques are powerful problem-solving tools [Rieber1995].

Visualisation is defined as representations of information consisting of spatial and non-arbitrary characteristics (i.e. "picture-like" qualities resembling actual objects or events). [Rieber1995] includes both internal (for example, mental imagery) and external representations (for example, real objects, printed pictures and graphs, video, film, animation). This is unlike verbal representations which are arbitrary (for example there is no natural reason why the word "boat" should be used to represent the real object)

Visualisation is helpful in problem solving and can give important cues that can lead to the right solution [Kaufmann1979]. There could be instances where it can lead astray as well. Consider the following examples:

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

A man had four chains, each three links long. He wanted to join the four chains into a single, closed chain. Having a link opened cost 2 cents and having a link closed cost 3 cents. The man had his chains joined into a closed chain for 15 cents. How did he do it?

While the reader can try to solve the problem (of course before reading ahead for the solution), try to also reflect and take note of the strategies that you use to solve the problem.

Some people might find it easier to solve it by first drawing the four chains on a piece of paper which actually aids them to visualise (or to construct a visual representation of) the problem's entry conditions.

The obvious solution of four links opened and closed would cost of twenty cents. But after working through opening and closing links with a visual model, one discovers that the solution rests in opening all three links of one of the chains. Then those three links can be used to join the other three chains. When the problem is converted into visual form, the solution is easy to derive and/or follow.

As another example, three people were presented with an everyday problem of fixing food while they had a recipe designed for four people. When the recipe called for two-thirds cup of cottage cheese, one of them solved this problem of "three-fourths of two-thirds" by measuring out two-thirds of a cup onto a table, patting it into a circle, and marking a cross on it. One excess quarter of it was then removed to get the correct portion [Reiber1995]. This these example also shows how everyday people use spatial and concrete reasoning abilities to grapple with problems often expressed in abstract form in traditional mathematics.

Indeed many might have encountered people describing that after having grasping a solution instantaneously and as a whole, they have (or had) trouble to putting the idea, already completely conceived, into an appropriate verbal form to share with others (or put it in writing for a scientific paper). Visualisation is thus a powerful cognitive strategy for all people, and researchers who study the problem-solving process have long recognised visualisation as an essential strategy [Rieber1995].

4.5 Conclusions: The crucial role of presentation and interaction

In this chapter, we talked about the different presentation and interaction related characteristics that play a role not only in the users' perception of the problem as being easy or difficult, but also have an impact on users' performance (success or failure) in the given task. In the chapters that follow (that is, in Part II of the thesis), we will talk about the two experiments that we conducted (one in real settings and the second in virtual settings) while keeping in mind the presentation and interaction aspects, and analyse and compare them. Chapter 5 first talks about the task at hand, while Chapters 6 and 7 are dedicated to the two experiments, one each in the real and virtual environments respectively. Finally, Chapter 8 will offer conclusions and whatever we learnt about the different interaction techniques we used and their performance and suggestions for their improvements.

Part II

Experimental contributions to the study of 3D objects manipulation task in real and virtual environments

5 Chapter 5: The Puzzle Experiments

This chapter discusses all the details, including the physical and geometrical characteristics, of the puzzle which was used in our experiments in the real and the virtual settings. It discusses in particular the aspects that were common to the puzzle in both experiments. It explains the main idea and motivations behind the experiments and certain facts about the task selected.

This chapter is an introduction to the two experiments that were conducted. Prior to conducting an experiment in a virtual setting, an experiment was conducted in a real setting. The objectives of the first experiment were to explore the relationships between action and cognition related to 3D problem solving by direct manipulation.

The main purpose of this initial experiment was to get a deeper insight into how humans solve real-world problems that involve physical manipulation together with cognitive processes related to 3D spatial problem solving. Through this experiment, we also wished to identify if there were any patterns or general solving rules, at the macro or micro level. For this, we designed a task in which the users were required to solve a 3D puzzle, one which required acumen at both the mental level as well as the sensory-motor level.

5.1 The idea and motivations behind the experiment

A 3D blocks based task was conceived and conducted to study the processes and skills involved in 3D-spatial problem-solving. The task was in fact a 3D puzzle which may be categorised as a box-packing puzzle (see “3D Dissection Puzzles” in chapter 3). Conducted in the real and virtual settings, the task was called the *Blocks Manipulation Task* (BMT) in the real settings, and *Virtual 3D Cube Puzzle* (or simply *Virtual Puzzle*) in the virtual settings. The goal was first to try to find out by empirically testing in real settings how exactly humans solve such problems where one has to employ a variety of skills like reasoning, mechanical inference (3D mental rotation, etc.) and spatial manipulation.

We were also interested in investigating the various factors that can influence the subjects’ performance and strategies. Thus the purpose was to observe the processes involved in solving problems calling for assembling and disassembling manipulations, and to determine what could be the local techniques involved (at the micro level), and which could be the more global techniques or reasoning used (at the macro level) by humans.

5.1.1 Questions raised in our experiments

We tried to seek answers to the following specific questions:

- What are the factors that influence the performance and/or strategy in solving spatial puzzles by humans? Could there be any different traits or approaches, or other pointers, which could differentiate the BMT solvers from non solvers?
- Are there factors that influence the performance or the strategy? And if there are, then what could be the different factors and strategies in solving this type of spatial puzzles by humans? And could it be possible to identify or categorise them in some way?
- Are all solutions to the puzzle equally likely to occur, or are there certain kinds of preferred solutions that are more likely to occur than the others? And in case there are, could one suggest why it is so?
- Are there basic visuo-spatial capacities required to perform the BMT? To provide a first answer to this question, we used two classic psychometric instruments that are commonly used to measure individual visuo-spatial abilities (the MPFB and the MRT).
- How could the knowledge acquired in the experiment in the real settings be used in developing the Virtual Puzzle, and how can the interface for the Virtual Puzzle be designed making use of not only the insights gained but also mixing it with the currently known virtual environment interaction techniques?

5.1.2 The task chosen

The chosen task was of average difficulty. It was neither too difficult nor trivial. It contained seven blocks from which the users were to build a cube, and the users had to explore and work their way towards a solution. Pre-tests revealed that it was certainly not an impossible task (as some people were able to solve it, though not instantly). However, the solution was not apparent to everyone, at least not right at the very beginning.

The problem could be solved using different approaches. Moreover, the chosen task was readily replicable in a virtual environment involving manipulations.

The task involved 3D problem solving, spatial reasoning, solving by direct manipulation, reasoning, visualisation of solution, planning and strategy-building, and manual dexterity. Apparently, most of these abilities interplay in many real life situations, for example:

- Mobile phones that require their batteries to be put inside the phones first to make them operational
- Do-it-yourself card-board file and book holders that have to be folded and assembled first before they become useful
- Assembling of the ever growing (and increasing by the day) variety of modular furniture, which most of us arrive at assembling
- Everyday items like electrical appliances like café machines (how do we know how to operate one even if we have never ever operated that particular machine before? One has to answer questions like, what is the purpose of each cavity provided, one has to where to put the water)
- Repairing a motor vehicle requires assembling/disassembling.

5.2 Materials, physical and technical aspects of the puzzle

Before describing the actual design, methodology and the protocol of the experiment, let us discuss the blocks that were used in the BMT. It is more appropriate to discuss these details here rather than in Chapter 3 along with other puzzles because firstly, it merited a separate mention and a more detailed description; and secondly, it is better that this description just precedes the details of the experiment. Nevertheless, comparisons with puzzles discussed previously in Chapter 3 are made freely wherever required and appropriate.

5.2.1 Materials

The BMT consisted of 7 coloured wooden blocks (figure 5.1). Note that the blocks neither are all of different colours, nor the entire same colour. Three of these blocks have a unique form but have the same yellow colour (the three central blocks in figure 5.1), while there are four blocks having similar shape but different colours (yellow, red, green, blue). In other words, those blocks which have the same shape have a different colour, and those with the same colour have a different shape. Thus none of the two blocks have both the same shape and colour and each block can be uniquely identified.

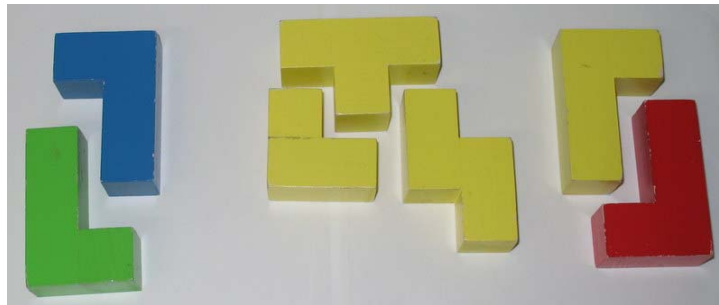


Figure 5.1 : The seven blocks given to the subjects during BMT - the experiment in the real setting

Figure 5.2 shows the *virtual blocks* i.e. the blocks used in the virtual puzzle, while Figure 5.3 shows the two alternative equivalent views which may help better visualise the blocks, with 5.3-a depicting the blocks in 3D, and 5.3-b giving a functional view of the blocks. All the blocks in fact may be visualised as being composed of multiple mini-cubes (figure 5.3-b).

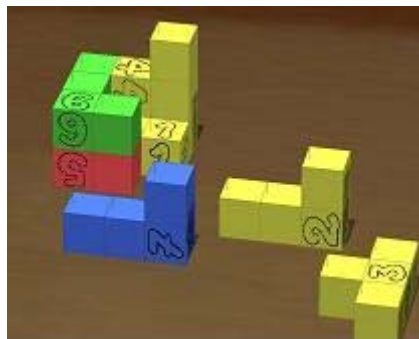
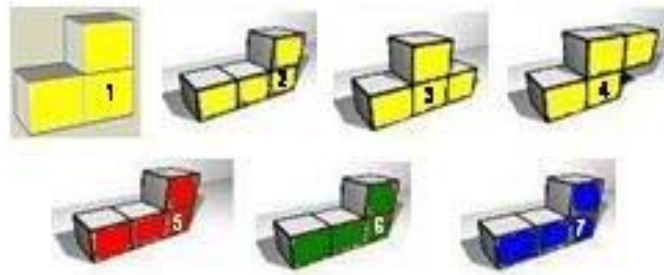
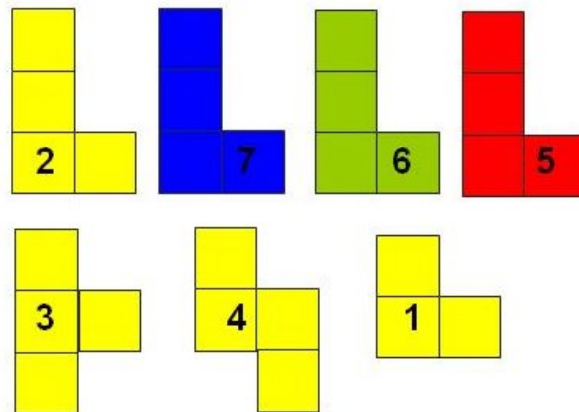


Figure 5.2: The blocks as in the Virtual Puzzle – the experiment in virtual setting



a)



b)

Figure 5.3 : a) The blocks as visualised in three dimensions, and b) Functional diagram of the seven blocks, each of which could be thought of to be decomposed into 3 or 4 mini-blocks

There are four distinct shapes of the blocks. The first is the longer 'L-shaped' block (as mentioned before, there are four blocks like this one), the second one is 'T-shaped' block, the third one is 'S-shaped' (or Z-shaped), and finally the 'small L-shaped' block.

5.2.2 Conventions used for measurement of the dimensions of blocks

For the purpose of uniformity, when measuring the dimensions of the blocks, we will use the following conventions for the blocks' dimensions (illustration at figure 5.4) :

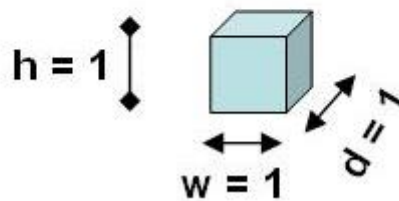
- i) as much of the blocks' base should be in contact with (or touching) the table as possible;
- ii) the blocks shall be placed in the flattest possible manner (i.e. with minimum height from the table); and
- iii) the longest side in front shall extend from left to right (and shall be measured as the block's width)

Legends

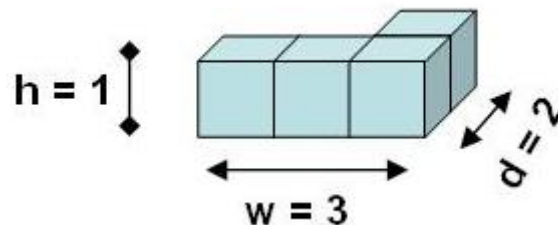
w : *width*

d : *depth*

h : *height*



a)



b)

Figure 5.4 a) Visualising 1x1x1 block; b) Visualising L-shaped block

In the example figure 5.4-b above, the block has the width of '3', depth of '2', and height of '1' unit. Notice that this height of '1' unit is constant for all the seven blocks. (Notice that the shape of the block in the example figure above corresponds with the four similar L-shaped blocks used in BMT and the Virtual Puzzle (figures 5.1-5.3)).

Throughout the text, the terms 'unit-cube' and 'mini-cube' will be used interchangeably, as well as the terms 'blocks' and 'objects' interchangeably.

5.2.3 Comparison of the blocks used in our experiments with the Soma Cube

As the reader might have already remarked, there is indeed a striking resemblance in the form of the individual blocks as well as the overall goal of the puzzle used in our experiments and the Soma Cube Puzzle, also known as the *Soma blocks*. But while their external form and the goal of the two is the same, the shape of the blocks in our experiments was nevertheless not exactly identical to that of the Soma blocks. (The reader may verify this by comparing the shape of blocks in figure 5.3 with figure 3.5.)

Among the more notable differences of BMT with respect to the Soma Cube, in the case of BMT, not all the seven shapes in our puzzle were unique (some of them were in fact of the same size and shape, except for the fact that they were of a different colour), and were thus interchangeable, which is not so in the case of the Soma Cube. Thus our puzzle was somewhat easier to solve than the Soma-cube, in principle; since in our case, for a given solution, two blocks could inter-replace each other, which is not true in the case of the Soma Cube. Another way of saying this is that in the Soma Cube, it is not possible from a given solution to obtain a second one by just interchanging two pieces, whereas it is possible to do so in the case of BMT. Another notable difference in their physical characteristics was that all our blocks could be placed flatly onto the table (as in figures 5.1, 5.3-b and 5.4-b), unlike the Soma blocks (for example, the Soma blocks in figures 3.5-e, 3.5-f, and 3.5-g cannot possibly

be placed flatly on the table). In other words, the height of all the blocks used in the BMT is one; which is not true for certain blocks of the Soma Cube (compare figures 5.1 and 3.5).

The subjects, when asked, did not find that colours played any part in their strategy or solution of the cube, and it was just the shape of the objects that determined their solutions/strategies.

5.2.4 The complexity of the blocks

We approximate the intrinsic complexity of the block in the BMT, (i.e. what made them more or less complex) by assessing the complexity of the shape of the blocks. Figure 5.3 shows, (and as discussed before as well), that the block numbered 2, 5, 6 and 7 have the same form.

The formula proposed for calculating the complexity of the blocks takes into account the lengths of various dimensions of the blocks as well as their forms. But before discussing the complexity formula *per se*, we shall first discuss a few conventions and assumptions that we shall make (for example how each block is to be placed, and how the dimensions of the blocks are to be measured, etc.), for easier description and to ensure the uniformity of calculation in the measurement of the complexity of the blocks.

The cube shown in figure 5.5 has three sides: the front side (in green), the top side (the top of the cube), and the right side (to the *right* of green, *right* from the reader's perspective). Unless otherwise stated, the perspective of the blocks would always be as such when calculating its complexity.



Figure 5.5 : Basic 1x1x1 block

For uniformity of calculation and to avoid calculation ambiguities, the first thing is to place the block in the manner such as explained in section 5.2.2, which essentially states that *as much of the blocks base should be in contact with the table as possible and in the flattest possible manner; and finally the longest side in front shall extend from left to right (and will be considered as the block's width).*

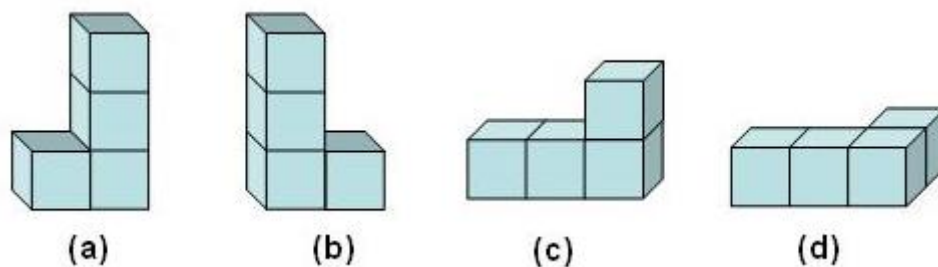


Figure 5.6 : Different shapes in which the L-shaped block can be positioned

Figure 5.6 shows that the same L-shaped block placed in four different ways. In this figure, the configurations 'a' and 'b' and 'c' above do not satisfy the conditions whereas the configuration 'd' does, and so this is indeed the configuration that will be considered. The conventions for the dimensions of width, height and depth are illustrated in figure 5.4 above. (There are in fact 24 different ways of placing this block. The following section shows all those 24 possibilities).

Refer to figure 5.7 to see how the angles for each block are calculated. Imagine that you are moving your finger in a straight line over the block shown in figure 5.7. When the finger takes a turn, it is counted as an angle. Figure 5.7(b) shows a close-up view of it, with arrows showing the two different sides of the block (perpendicular to each other), and the red dot is precisely where they make the angle of 90° . The block shown in this sample figure contains six angles depicted by the six tiny red dots (figure 5.7-c).

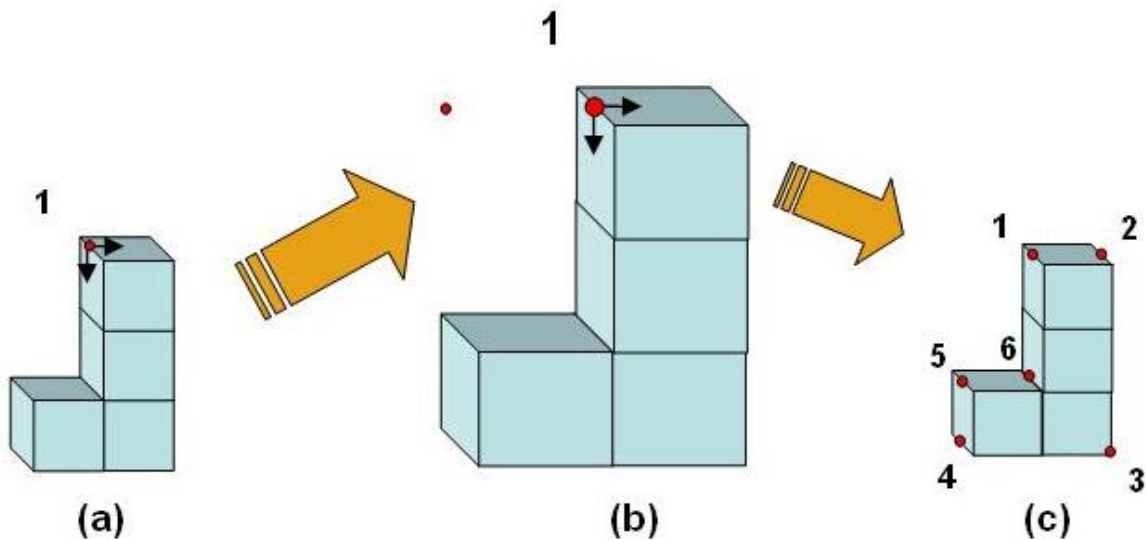


Figure 5.7 : Counting angles in the block

We can now describe the formula proposed for the complexity of blocks:

Formula: $[(w + a) \times h] + d - K;$

where

$w = width;$ -> (real calculated values for blocks used in the our puzzle range from: 2-3)

$a = angles;$ -> (real calculated values for blocks used in the puzzle range from: 6-8)

$h = height;$ -> (real calculated values for blocks used in the puzzle range from: 1)

$d = depth;$ -> (real calculated values for blocks used in the puzzle range from: 2)

K is a constant having value 9.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

For the blocks used in our puzzles (i.e. the BMT and the Virtual Puzzle), the values for width 'w' ranged from 2 to 3, the number of angles 'a' ranged from 6 to 8, while the depth 'd' of all blocks was 2, and the height 'h' of all blocks 1. Note that without subtracting the constant: a 1x1x1 cube (figures 5.4-a and 5.5) has the value 6; whereas among the blocks used in our puzzles (see figures 5.2 and 5.3), a small L (block#1) has value 10; the long Ls (blocks# 2, 5, 6, 7) have value 11. Thus the final values of the blocks used in the experiment, after this subtraction of the constant K, come out to be as follows:

Block # 1 has complexity index = 1

Block # 2, 5, 6, and 7 has complexity index = 2

Block # 3 has complexity index = 4

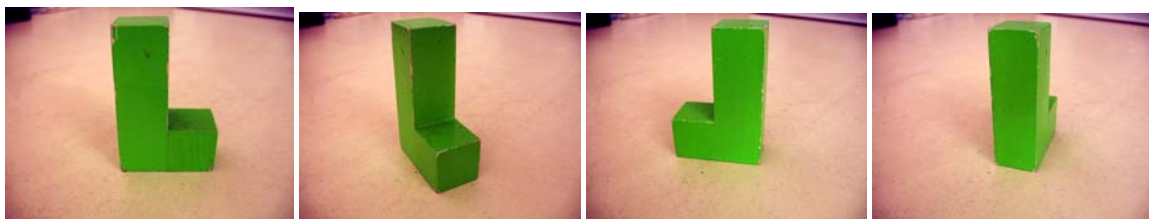
Block # 4 has complexity index = 4

Note that we have discussed the complexity under this section as it concerns the material, physical and technical aspects of the blocks used in our experiments. The complexity of blocks is a relevant feature since it is expected to shape the representation of the task for the participants and to elicit specific strategies. For instance, if the participants spontaneously classify the seven blocks in terms of complexity, one may expect that they will use some of them as "grounding" pieces during their search for a solution. In line with a number of basic mechanisms of human perception (including those promoted by the Gestalt theory), the blocks which are the most complex ones are likely to be used as starting pieces to which other (less complex) pieces are added. The reverse procedure (adding complex blocks to simple ones) would be unexpected.

5.2.5 Formal description of the cube including the different orientations and the positioning of the blocks

In this section, we will show the different orientations in which the blocks may be placed as well as their different possible starting positions.

Earlier in this chapter, we claimed that each block may be placed in 24 different *orientations*. We shall first discuss and illustrate those 24 different orientations (that is, how each of the given blocks may be placed in 24 different ways). Look at the four images O-01, O-02, O-03 and O-04 below (short for, Orientation-01, Orientation-02, Orientation-03 and Orientation-04 respectively). The green block is placed on the table with the small projected part of the L-shaped block pointing towards the right in the image titled 'O-01'. When the block in image one is turned 90° clockwise, we get the orientation of the block as shown in image 'O-02'. If we continue to rotate it by 90° yet again in the same (clockwise) direction, we get block as shown in image 'O-03', and after yet another similar rotation, we can get the block as shown in image 'O-04'.



O-01

O-02

O-03

O-04

To help the user visualise better, the first four blocks above are shown again from a different angle. Let us call them 'O-01-from-above', 'O-02-from-above', 'O-03-from-above' and 'O-04-from-above' respectively. These four blocks are in the same corresponding orientations, except for the fact that we are now viewing them from another angle (i.e. from the top of the blocks).

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI



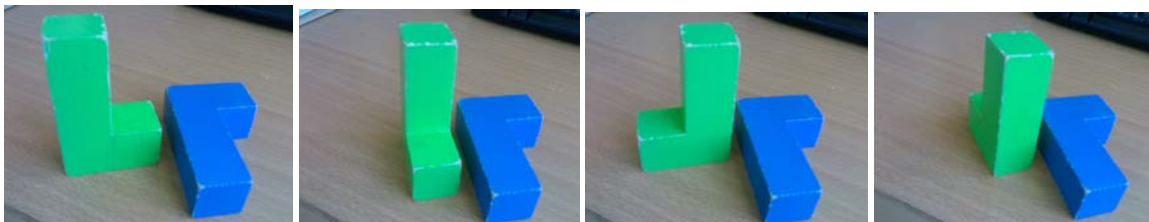
O-01-from-above

O-02-from-above

O-03-from-above

O-04-from-above

And the same four orientations are shown yet again, from yet another angle and with another block besides it.



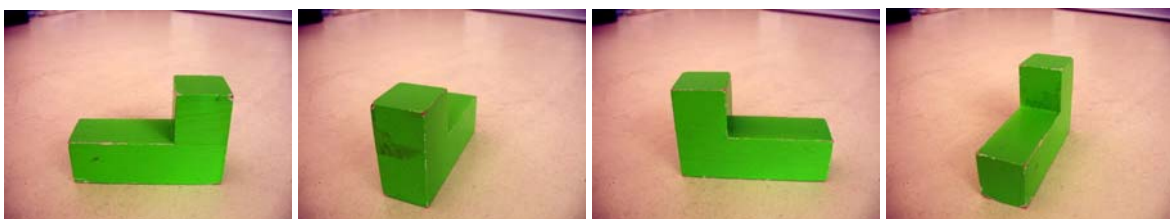
O-01-with-block

O-02-with-block

O-03-with-block

O-04-with-block

Now, the block is placed in a different way, and then rotated four times to get four different orientations. The principle remains always the same - the successive images in each line are 90° apart. The block position shown in the first of the four images in the line, when rotated by 90°, yields the block in the second image, and so on and so forth.



O-05

O-06

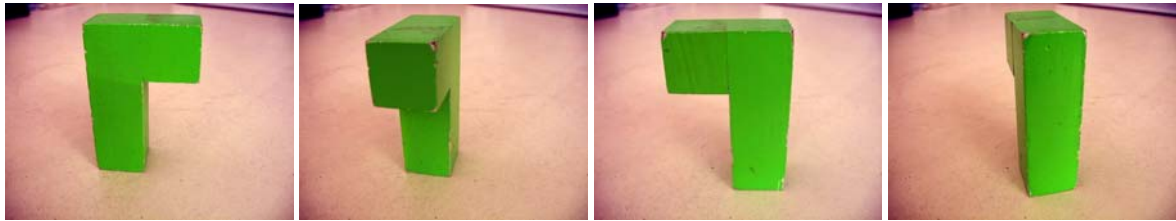
O-07

O-08

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

And we proceed in the same way.

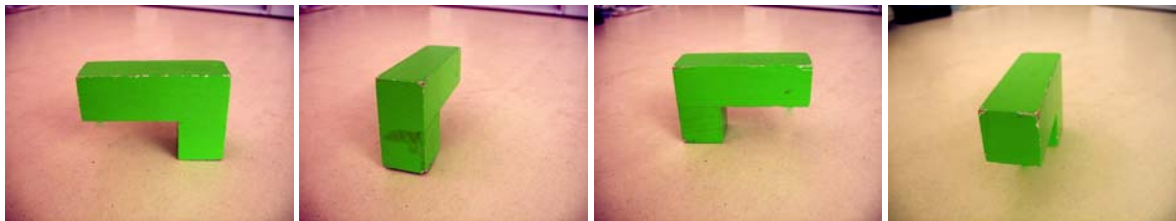


O-09

O-10

O-11

O-12

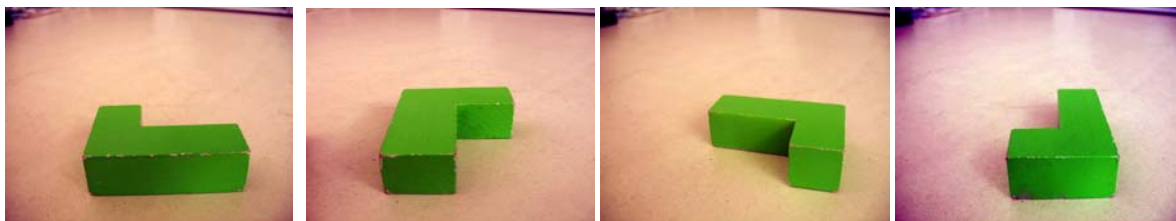


O-13

O-14

O-15

O-16

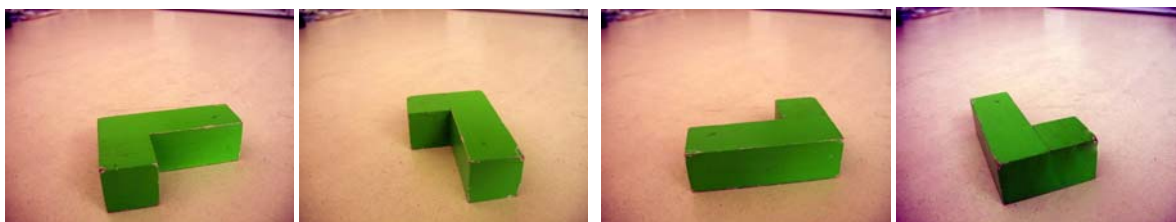


O-17

O-18

O-19

O-20



O-21

O-22

O-23

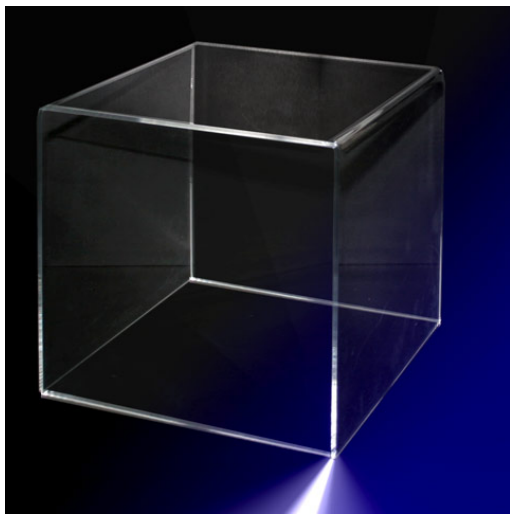
O-24

Thus we have the 24 different orientations (or ways) in which each block can be placed, independently with respect to the other blocks.

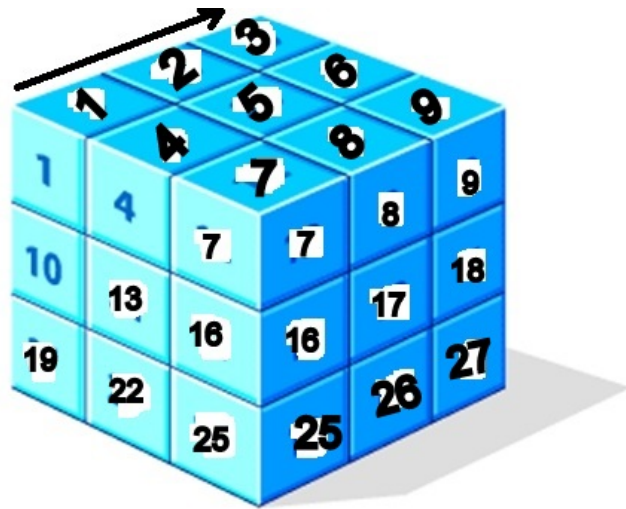
Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

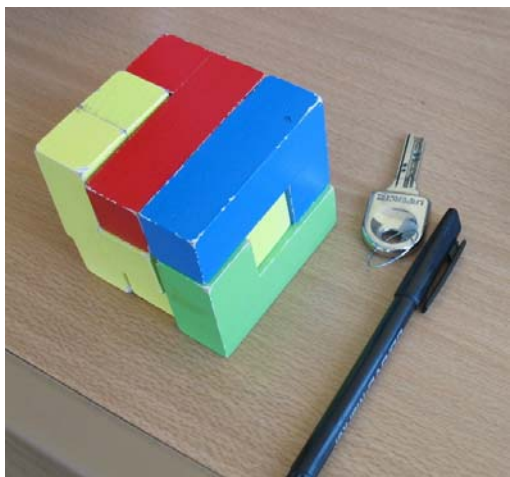
We would like to introduce the concept of valid or invalid (or legal and illegal) combination of orientations and placement of blocks 'within the workspace', but it is first necessary to first define what we exactly mean by *workspace*. We define workspace as 'the imaginary 3 x 3 x 3 cube-shaped space within which each block ought to lie, in order to solve the puzzle'. The workspace can be imagined to be composed of 3 x 3 x 3 unit cubes, or a hollow space containing a total of 27 unit-cubes (see figure 5.8). To solve the puzzle all the blocks ought to be placed within this workspace of 3x3x3.



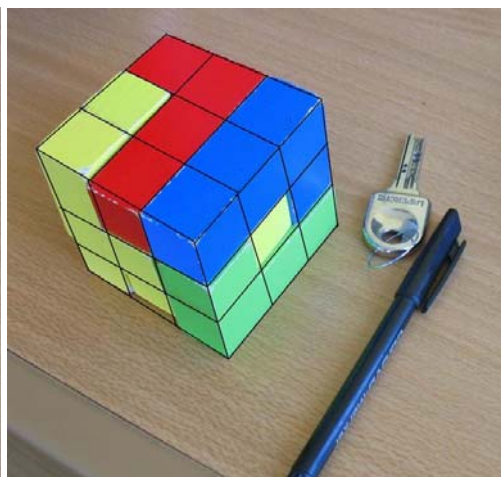
(a)



(b)



(c)



(d)

Figure 5.8 : Workspace concept depicted in various ways

Figure 5.8 shows different ways in which the workspace may be visualised. A workspace may either be visualised as a glass box (figure 5.8-a); or as a 3x3x3 cube resulting in 27 mini-cubes (see figure 5.8-b), which have been numbered here with the top slice occupying numbers from 1-9, the middle slice from 10-18, and the bottom slice from 19-27. Figure 5.8-c shows how the cube looks when solved using the wooden coloured blocks, i.e. where all the 27-minicubes of the workspace are occupied by the blocks. Figure 5.8-d is the same as figure

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

5.8-c, except for the fact that it has added imaginary gridlines, to help visualise the 3x3x3 cube workspace.

Given that there are 27 positions of the workspace, and provided that there are 7 blocks in our puzzle, the user could thus place those 7 blocks starting from any of those 27 positions of the workspace, and in any of the (earlier discussed) 24 different orientations. Thus this gives us a total of potentially: 27 (starting positions of each block) times 24 (orientations of placing each block) times 7 (total number of blocks), or $27 \times 24 \times 7 = 4536$ ways of independently placing the seven blocks. However, we will see that not all the combinations of positions and orientations are valid ones.

Depending upon which of those 27 positions that we start from, and which of the 24 orientations that we select from above, some of them may extend out of the workspace, the 3x3x3 space – as shown in figure 5.9. Let us call this phenomenon when the block moves out of the 3x3x3 space, *invalid* or *illegal placement* of the block. So in other words, we could say that ‘a certain block when placed at a certain starting position (1..27), and in a certain orientation (1..24), would either lie entirely within the workspace (legal placement), or partly outside of it (illegal placement). Which of the orientations are legal and which ones are not depends upon the size and shape of that block.

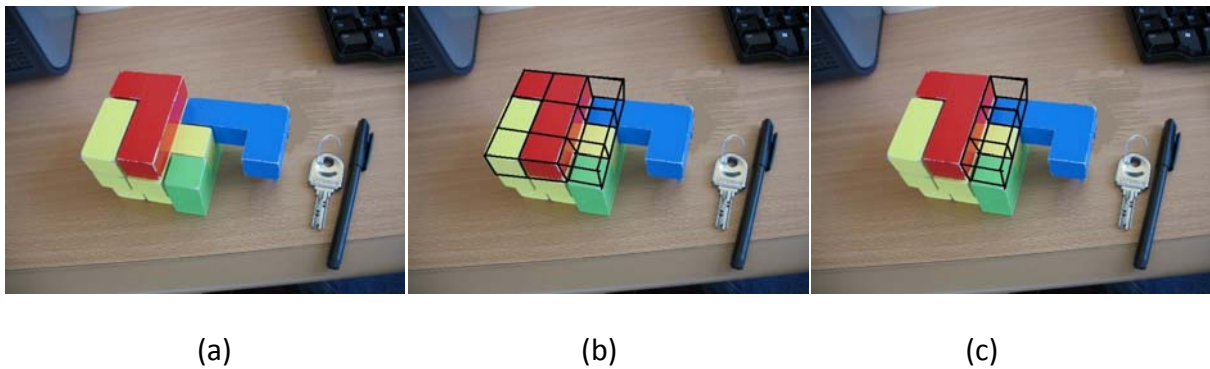


Figure 5.9 : Blocks being placed in the workspace

In figure 5.9-a, one can observe that the blue L-shaped block is placed outside the workspace. Figure 5.9-b shows the same image with the top layer (slice) of the workspace shown with gridlines. Notice that the left part of this workspace top-layer contains blocks, while the right part of this is empty and does not contain any blocks. Figure 5.9-c highlights only the empty part of the workspace, i.e. the one which is hollow. The blue block ought to cover this hollow in the space, in order to solve the cube, and not be projected out part as it currently is in figure 5.9.

One can observe in figure 5.9 that the blue block is occupying the 18th position in the workspace (see figure 5.8-b for the numbering of workspace), and it is placed in the 19th orientation as described above (see O-19 above). We can clearly see in figure 5.9 that the blue block in its given orientation is out of the workspace (and thus its placement is illegal). In other words, one can notice that the blue block has three unit-cubes which lie outside the workspace. In order to solve the puzzle, one must devise a way to somehow either find a way to rearrange the blue block so that those three unit-cubes of the blue block occupy the same space as the three hollow unit-cubes of the workspace.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Now, had there been no other blocks placed before this one, then this current problem could have been resolved by rotating the blue block clockwise, so that its orientation corresponds to that of O-20, and thus this block would no longer have lied outside the workspace, and occupied the workspace positions 18, 17, 16 and 13 (again, see figure 5.8-b for the numbering of positions in workspace). However, notice that in the current case, this would not be possible since there are other blocks occupying the workspace positions, 16, 17 and 13. It is a physical constraint that two blocks cannot occupy the same positions in the workspace.

A solution, however, does exist, whereby one can reach the solution by rearranging the blue block alone (i.e. without necessarily rearranging any of the other blocks) in a different manner. The reader might have already remarked that to solve the cube, the blue block must be placed in the orientation O-16 (whereby, the blue block occupies the workspace positions 7, 8, 9 and 18).

Thus the important constraints that can be observed in this problem are: a) there are certain positions and orientation combinations which must be avoided where the block would eventually lie outside this workspace; and b) since the blocks are placed one after the other, and since no two blocks can occupy the same space at any given time, so provided a block has been placed at a certain position, it potentially reduces the number of placement combinations (of positions and orientations) for the subsequent blocks. In fact, each of the blocks placed, which constitutes part of the solution, also reduces the possibilities where the subsequent blocks may be placed.

The cube in its final solved form can be imagined to be composed of 27 units (dimension: 3 x 3 x 3). Thus, each of the seven blocks (figures 5.2 and 5.3) maybe placed in any of those 27

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

different positions (ignoring all the constraints for the moment). Furthermore, each block can be placed in 24 different orientations at each of these 27 positions.

In this chapter we discussed about our 3D puzzle, its physical and geometrical characteristics and the various dimensions of our puzzle. In the following two chapters, we will report the experiments conducted in the real and virtual settings, both based on the manipulation of the same 3D puzzle.

6 Chapter 6. The Blocks Manipulation Task: The Experiment in Real Settings

This chapter reports and discusses the experiment which was conducted in the real settings. It discusses the details for the experiment. An analysis of the task, as well as the materials used, the methodology and the protocol for this experiment, are presented. The chapter then describes how the subjects' actions were interpreted and converted into data. The method of analysis used for the experiment is discussed and notation of moves performed by the subjects to complete the task is introduced. Finally, the results observed from the experiment are presented, followed by some discussion on them.

Chapter 5 discussed the main characteristics of our puzzle, in particular, the characteristics of the blocks in terms of their form, colour, etc. It was also put into perspective where our puzzle lied under the broader categorisation of puzzles and more specifically spatial puzzles. We also proposed a formula for calculating the complexity of the blocks and the concept of workspace and the different orientations in which the blocks may be placed. All these aspects are common for the 3D puzzle in the physical and virtual environments. In this chapter, we will discuss the aspects more specific to the experiment conducted in the real setting.

6.1 Physical setting

The 3D-puzzle task consisted of constructing a regular cube by placing seven wooden blocks in their right positions and orientations. There were several solutions to the puzzle, however this information was not revealed to the participants beforehand. Figure 6.1 shows the setup where the experiments were conducted while figure 6.2 shows a closer view of the blocks. The participants sat on a chair with a table in front of them. A lamp which shed light on the blocks and that lay behind the participants was turned on at the time of the experiment. The experiments were recorded with a video camera, which was installed on a table in front of the participants, and captured the movements and manipulation of the blocks as well as the participants' hands.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI



Figure 6.1 : Laboratory environment for the tests in real settings

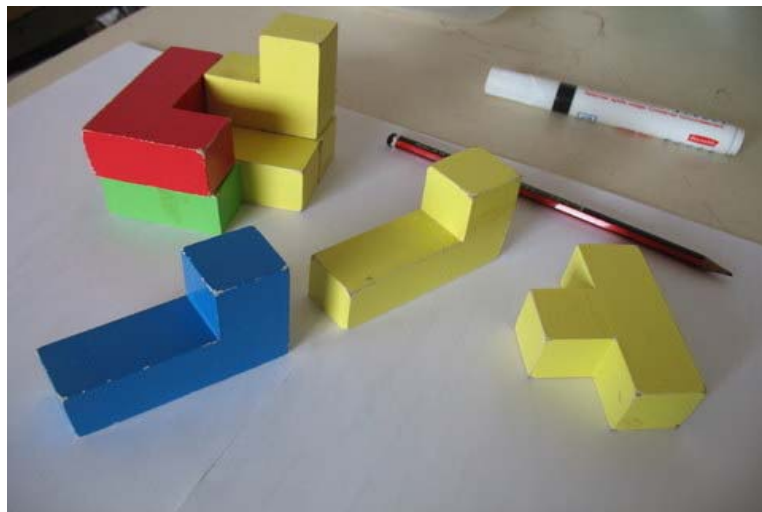


Figure 6.2 : Close up view of the blocks

Figure 6.3 shows the subjects attempting to solve the BMT.

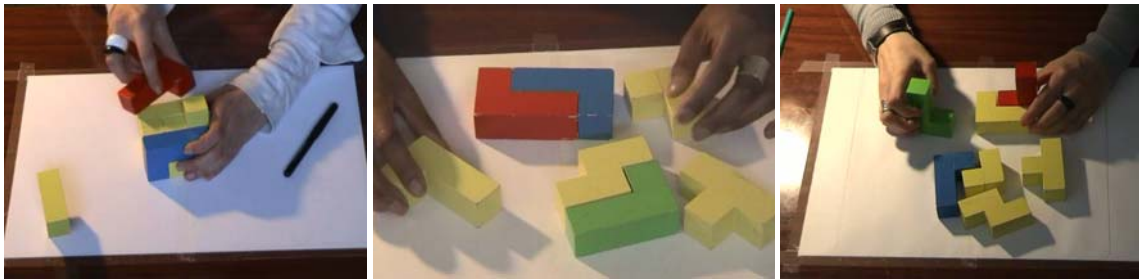


Figure 6.3 : Subjects attempting to solve the puzzle during the BMT

6.2 Design and Method

6.2.1 Participants

Twenty-four participants (10F, 14M) aged between 23 and 39 years ($m = 27$ y, $sd = 4$) participated in the study. Five participants were left-handed, while 19 were right-handed. All but one participant were university level (Master or PhD) students. Nine out of the 24 participants were from the psychology domain (with quite diversified sub-domains and specialisations) while others were from computer science (with diversified sub-domains: graphics, VR, signal processing, audio, linguistics, etc.).

6.2.2 Procedure

All the participants sat for the Blocks Manipulation Task. In the BMT, the subjects were shown 7 (seven) blocks having different shapes and colours as shown in figure 5.1. Their task was to make a regular cube out of those blocks. The participants were required to limit their area of manipulation to an A3-paper sized demarcated area. The subjects were encouraged to verbalise freely their strategies and ideas that came to their minds to solve the blocks puzzle; they were also free to pose any questions, or share any comments that they had about the task.

They were periodically reminded to verbalise, especially when they did not speak up for a while. From time to time, a conversation was started for example by encouraging them to

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

explain a particular move, or a question was posed to them based on some observation made on the fly (for example when they totally changed their starting strategy, etc.).

The subjects filled in two questionnaires, one before and one after the test, and were interviewed at the end. In addition, a few questions were posed to them while they solved the puzzle (questions posed included 'whether they were finding this puzzle easy or difficult'; and 'how many solutions they thought existed to the problem').

The participants were simply asked to solve the problem at their own pace. However, only those who solved it within 10 minutes were categorised as solvers, otherwise as non-solvers. They were nevertheless allowed to continue for longer if they wished so, but their actions after the lapse of 10 minutes were not taken into account.

They were made to solve the puzzle at a leisurely pace, that is, as they would have had done in real life, and so as not to create a condition of competition. The intention was not to push them into solving it fast, or to test their performance under time-pressure conditions, but rather to extract the information and cues, by analysing different solution approaches and strategies, and looking for any common solving patterns that may be found, while the problem was solved.

The main idea was to see if it was possible to extract any verbal or non-verbal cues or any other information that might give an insight into the solution of the problem, for example, whether there were any patterns to be found amongst all participants, and whether any generalisations could be made based on that; whether there were any preferred solutions or approaches to the problem; whether there were any differences in strategies of those who

solved the puzzle and those who didn't; whether there was a preference for placing certain blocks together, and ideally, whether any cogent explanation could be found for that.

After the task ended (i.e. either after they had solved the puzzle, or after they had given up on solving the puzzle, or after they were stopped), they were given another questionnaire to fill in.

The participants were also invited to complete two visuo-spatial tests, the Minnesota Paper Form Board (MPFB) and the Mental Rotations Test (MRT). We were interested in knowing if the visuo-spatial capacities measured by these tests were predictive of the capacity of solving the 3D puzzle.

6.3 Data collection

To document the task – the BMT – two digital cameras (a video camera, and a photo camera) and two tripod stands were used. The BMT as well as the interview sessions were video-filmed.

For coding the moves for the BMT, a coding scheme was developed for noting down and analysing the actions performed by the subjects for solving the given task. The following section describes how the notation works.

6.4 Method of analysis

One has to make a decision as to how and at what level one wishes to observe and to describe the actions because there could be various levels (and ways) of describing the same

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

action, for example an action may be described at the physical level, the logical level and at the cognitive level (see figure 6.4).

The physical level is the description at the lowest and the most literal level, while the cognitive is the highest, most semantic (and contextual) level, and at which the actions could be observed. The logical level description lies between these two, i.e. it is a level of description higher than that of the physical level, and lower than that of the cognitive one. This is the level at which we examined and analysed the actions of the subjects in our experiments. This level of description is discussed in greater detail below with contrasting examples.

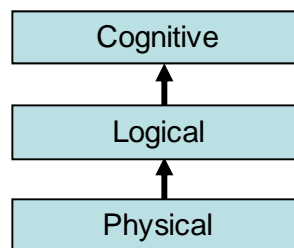


Figure 6.4 : Hierarchy of description of actions

Though of course there could be inter-linkages (and overlaps) between these levels, here is how one could broadly differentiate between these levels:

As mentioned above, the *physical level* description may be seen as the raw and literal description of the actions. Imagine as if a machine, unaware of the context or the meaning of the human actions or intentions, was describing the actions. An example describing actions at this level could be something close to the following description: “*Block 2 picked; moved 5 cm upwards; rotated clockwise by 90°; rotated anti-clockwise by 90°; moved left 15 cm, moved down 3 cm*”. This could be a description of picking up a block from the corner of

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

the table and placing it on the top of another block already placed at the centre of the table, where the participant was building his/her solution.

The description in the above example was what we call the *lowest level* or *physical* description; but we decided to describe the actions at a level that would be somewhat more meaningful to the humans. For example, for the same actions described in the example above, the following description could be easier to communicate and prove more helpful to the interlocutor to understand and visualise the action, as well as contextualise it better. The *logical* description would thus be something like: *Add Block 2*.

What we are more interested in knowing is the fact that a certain block has been added to the solution. Whether the block was moved 5 cm or 7 cm upwards before it was brought down and put on top of another block may not be something of great interest to us. Notice also that the description in the second example above also communicates the contextual information about the action performed. This is the context in which we developed a vocabulary to describe all the actions performed at this level (see below).

The idea behind the logical level description is that only one broad mental action should count for any given move conceived in the mind, regardless of whether the physical actions entailing it were longer (and more complex), or shorter (and less complex). To illustrate the point, consider the following examples. In the first instance, imagine the subject picks up a block from the *table* and holds it onto his/her hand: we say that he/she has performed a '*Pick*' operation. (Note: In the current context, we use the terms "operation" and "move" interchangeably). After waiting for a long time, the subject *adds* that already held block on to the *solution*: we say that he/she has performed an '*Add*' operation.

In another example, suppose that this subject wishes to add another block onto the solution and just picks up a block from the table and adds it onto the solution. Note that this time, we consider it as just one move '*Add*' (and not two separate ones, as was the case in the previous example, where the subject did a *Pick* followed by *Add* – just because the subject had to go through those two actions physically); since the assumed intention of the participant was just to add a block on top of another.

6.5 Notations of moves

A complete set of notation was introduced to record the moves made by the participants while performing the task. Before one could discuss the moves, definitions of a few terms would be in place, as they will be frequently used to describe the moves that follow.

The reader might have noticed that in the previous section (6.4), the terms *solution* and *table* were used. It would be in place to define what we mean by them (also particularly since all the moves will be defined in terms of these.)

- '*Solution area*' (or simply *solution*) is defined as the subject's active area of focus, i.e., where he/she is attempting to build the cube.
- '*Table*' is defined as any area other than the *solution*.

The list of *moves* that were identified and recorded as basic individual moves included the following (note that the following table contains the list of atomic moves performed on individual blocks):

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Table 6.1 : Notations of moves

Name of Move	Description
Add	The subject picks up a block from the table, and adds it to the solution, <i>in the same continuation of action</i> , and then leaves hand contact with that block. Note: To be counted as an <i>Add</i> operation, the subject either leaves contact with block after its placement, or the block continues to be placed on the solution for at least 3 seconds.
Remove	The subject picks up a block from the solution and places it on the table in the same continuation of action, and then leaves hand contact with that block. Note: To be counted as a <i>Remove</i> operation, the subject either leaves contact with block after its placement on the table, or the block continues to be placed on the table for at least 3 seconds.
Rearrange	The subject performs a rearrange operation when he/she grabs a block from the solution, and places it back on to the solution at the same or some other position, without leaving hand contact with that block.
Try	The subject picks up a block from table; places it onto the solution; and without taking the hand off the block, picks it up back again and places it back on the table; and then finally leaves contact with that block. In other words, the subject examines how the form of the solution looks like when a certain block is placed in a certain orientation (i.e. he/she adds it, examines for a short while, just to take it away again after the brief examination).
Try/Hold	The subject picks up the block from the table, tries it on to the solution, and without taking the hand off the block, retrieves it back from the solution and holds on to it (without putting it back on the table). Note: The move that follows this one could be an <i>Add</i> , <i>Try</i> or <i>Leave</i> move (or even another <i>Try/Hold</i> move).
Pick/Hold	The focus is at the table, i.e. starts with the manipulation of the block at the table. The subject picks up a block from the table and just holds on to it without placing it onto the solution. Note: This move may be followed by an <i>Add</i> , <i>Try</i> , <i>Try/Hold</i> or <i>Leave</i> move over this block.
Remove/Hold	The subject picks up a block up from the solution and holds onto it, without placing it either on the table or back on to the solution. Note: This move may be followed by an <i>Add</i> , <i>Try</i> , <i>Try/Hold</i> or <i>Leave</i> move over this block.
Leave	The subject leaves hand contact with a block (that is already held in the hand) and places it on the table. Note that the <i>Leave</i> operation may only be recorded when the subject puts back on the table an already selected block. This move may have been preceded by any of the <i>Hold</i> moves (such as <i>Pick/Hold</i> , <i>Remove/Hold</i> or <i>Try/Hold</i>).
Pick/Leave	The subject picks a block from the table and puts it back on to the table without placing it onto the solution.

The blocks were numbered from 1 through 7, thus for example "Add 5" means that block#5 (which would be the red block in figure 5.3) is added to the solution.

All the moves in the table above are *atomic* moves performed on individual blocks. However, there were certain other moves, like “Rotate” and “Disassemble”, which are performed over a group of blocks.

Table 6.2 : Notation for moves applied over a group of blocks

<i>Rotate</i>		The subject rotates the whole set of blocks.
<i>Disassemble (all)</i>		When all (or all but one) blocks have been taken off the solution in a rapid succession. Note: The Disassemble all is normally the last move of the current attempt. See below for a description of <i>attempt</i> .

6.5.1 The idea of attempt as the container of moves

There is one other concept that we would like to introduce here. During the experiment, the subjects make multiple *attempts* to solve the cube. A new attempt is counted whenever the subject starts all over again, and tries for the solution anew. More precisely, a subject makes a new attempt, when he/she has removed either all blocks or all but one block.

Do note that an *attempt* is a kind of a holder that contains a group of moves from the list above, i.e. every subject makes one or more attempts to solve the puzzle, and each attempt can contain any number of individual moves. Notice also that the subject always starts with the ‘1st Attempt’, and every experiment contains at least one attempt.

6.6 Results

6.6.1 Overall performance

The following acronyms shall be used in the results below: For *Solvers(S)*; for *Non-Solvers (NS)*.

Table 6.3 summarises the frequency of S and NS subjects as a function of gender. Among the 24 participants (14 males, 10 females), 17 (7F, 10M) solved the puzzle, while 7 (3F, 4M) did

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

not. Amongst solvers, 14 were right-handed (5F, 9M), while 3 were left handed (2F, 1M). Amongst the NS, 5 were right-handed (2F, 3M), while 2 were left handed (1F, 1M). On the one hand, male and female subjects did succeed quite similarly in the puzzle solving task (Fisher's exact test, $p=1$). On the other hand, right- and left-handed subjects didn't differ significantly in terms of being successful while solving the puzzle (Fisher's exact test, $p=0.608$).

Table 6.3 : Frequency of solvers and non-solvers as a function of gender

	SOLVER			NON-SOLVER			GRAND TOTAL
	LEFT	RIGHT	SUBTOTAL	LEFT	RIGHT	SUBTOTAL	
MALE	1	9	10	1	3	4	14
FEMALE	2	5	7	1	2	3	10
TOTAL	3	14	17	2	5	7	24

The totality (17 out of 17) of the successful participants solved the BMT puzzle within 9 minutes, of which 3 solved it in less than a minute, and none of those 17 between the last minute (table 6.4). The 17 participants who solved the puzzle did so in the time taken ranging from 21 seconds for the fastest, up to 514 seconds (just under 9 minutes). Half of solvers took less than 3 ½ minutes to reach the solution (median time = 200 s, average time = 231 s; SD = 168 s).

Table 6.4 : Time taken to complete the task

MINUTES	< 1	1-3	3-5	5-7	7-9	9-10	>10
NUMBER OF SUBJECTS WHO SOLVED THE PUZZLE BETWEEN THESE TIMES	3	4	4	3	3	0	7

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

The proportion of S and NS did not differ significantly between males and females (Fisher exact $p=1$) or along the left or right hand dominance (Fisher exact $p=0.608$). Our sample shows indeed no difference between men and women in the Mental Rotations Test (MRT Kruskal-Wallis, $\chi^2=0.031$, $df=1$, $p=0.86$) and MPFB (Kruskal-Wallis, $\chi^2=0.021$, $df=1$, $p=0.88$) (table 6.5).

Table 6.5 : Scores on tests of spatial cognition skills and MRT MPFB by gender and performance

	Score MPFB			Score MRT		
	Solvers	Non-solvers	Total	Solvers	Non-solvers	Total
Men	19.10 (3.81)	18.00 (6.98)	18.79 (4.64)	11.40 (3.63)	7.75 (3.69)	10.36 (3.89)
Women	21.86 (4.14)	12.00 (3.60)	18.90 (6.08)	10.71 (3.77)	8.00 (0.82)	9.90 (5.11)
Total	20.24 (4.07)	15.43 (6.24)		11.12 (3.59)	7.86 (5.40)	

We observe however a significant difference between S and NS (table 6.5) with respect to the MPFB score (Kruskal-Wallis test, $\chi^2=4801$, $df=1$, $p=0.0284$) and a marginally significant difference regarding the MRT (Kruskal-Wallis, $\chi^2=2727$, $df=1$, $p=0.099$). Both tests are also highly correlated ($r =0.55$). (See figures 6.5 and 6.6)

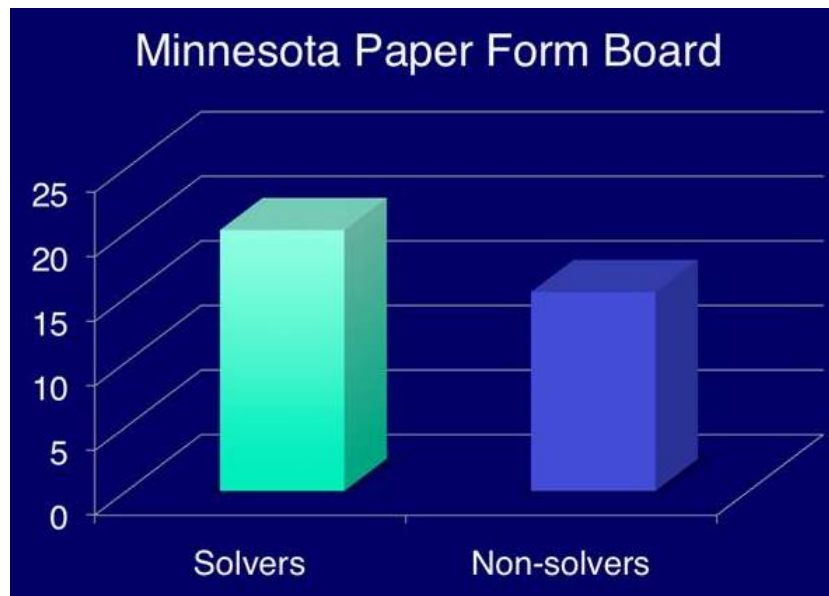


Figure 6.5 : MPFB scores for S and NS participants

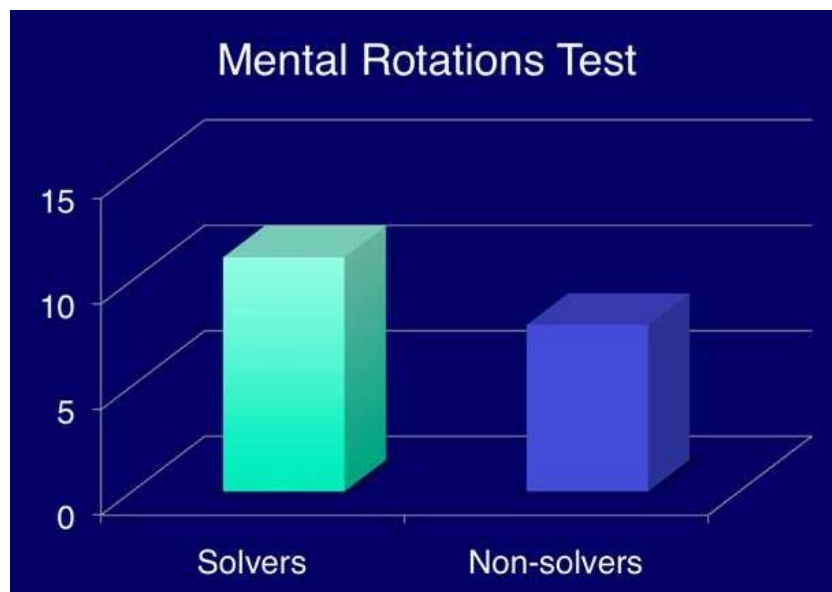


Figure 6.6 : MRT scores for S and NS participants

Figure 6.7 shows the average total times taken by S and NS (231 s and 738 s, respectively). The values differed considerably, although this is explained by the fact that the time was

stopped for S subjects as soon as they had solved the puzzle, while it continued to run till the end for the NS.

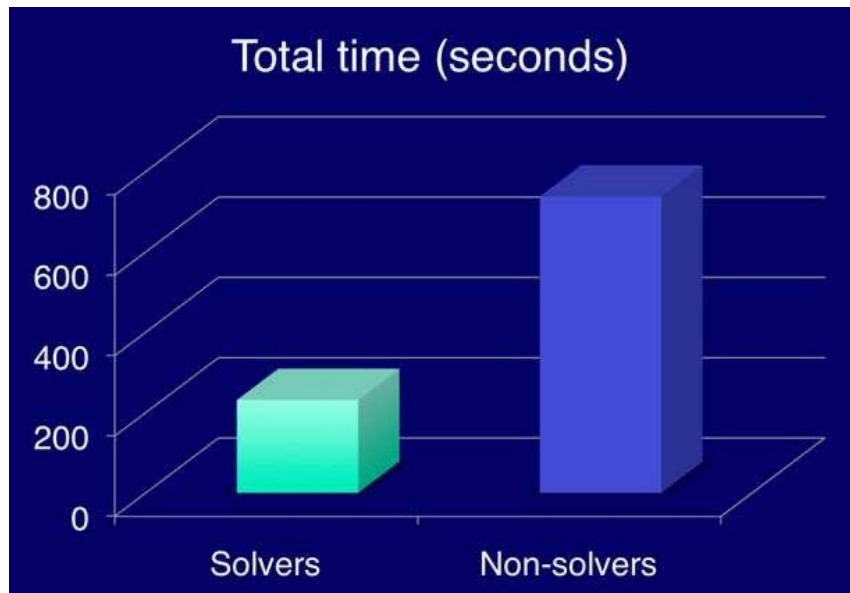


Figure 6.7 : Average of the total time taken by S and NS

6.6.2 Moves and attempts

The total number of *moves* for the total population (regardless of whether they were able to solve the puzzle or not) was quite varied, and ranged from 8 to 246 (Median: 81.5, Mean 90.9). Amongst S, it ranged from 8 to 173 (Mean: 52; median 49), while for NS, the range was 148-246 (mean: 186; median: 195) (see figure 6.8). The difference, in fact, is not unexpected, given that, as noted above, less time was devoted to the task by S than by NS subjects.

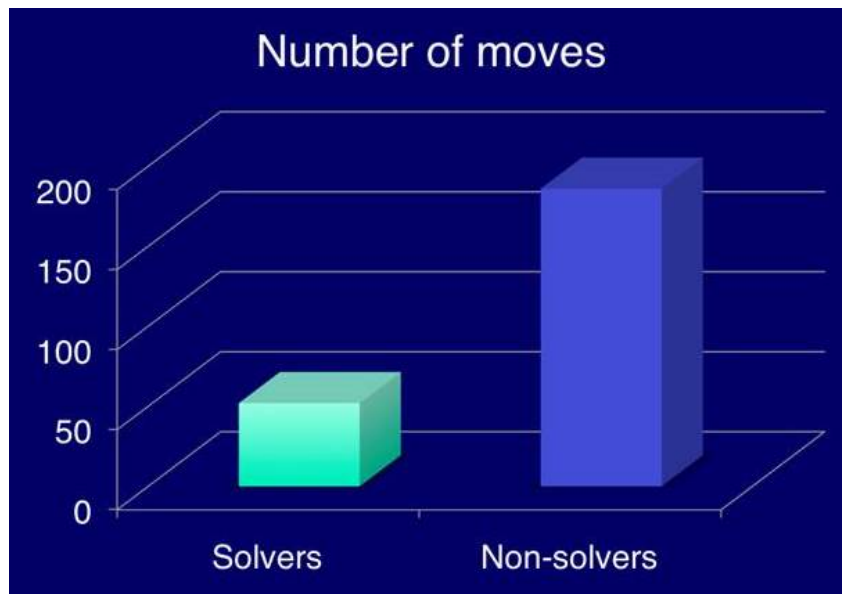


Figure 6.8 : Number of moves by S and NS participants

The number of *attempts* varied greatly between S and NS (4 and 10, respectively) (figure 6.9) but its explanation is the same as that for the number of moves, which is that the S subjects ended as soon as they achieved the solution unlike their NS counterparts.

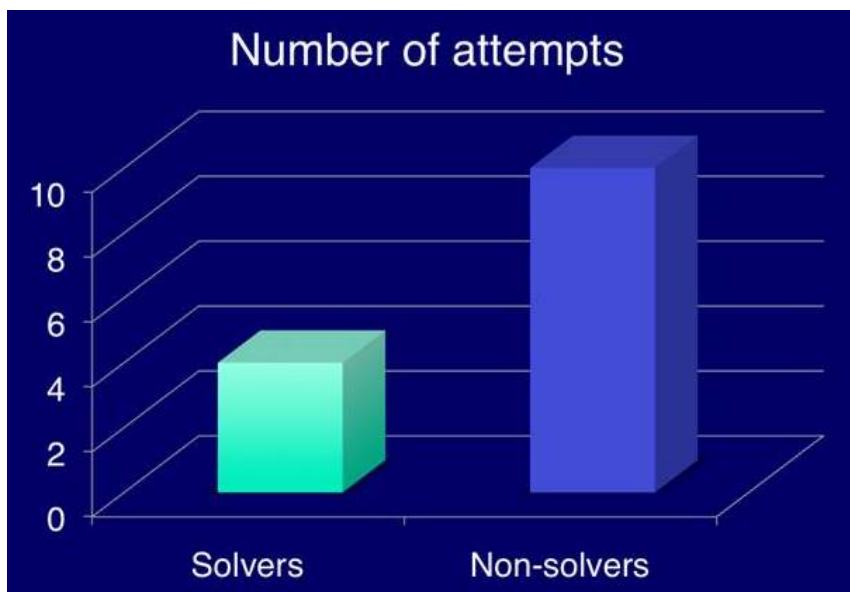


Figure 6.9 : Number of attempts by S and NS

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Of course, what is more interesting to consider is the average number of moves per attempt as well as the average number of moves per minute by S and NS. The former is the measure of the number of actions within the period of time devoted to a particular attempt (figure 6.10); whereas the latter is the measure of the speed with which the subject manipulated the blocks (see figure 6.11). It may be interesting to note here that the S had a (moderately) lower average for the number of moves per attempt as compared to the NS (13 and 14, respectively), and a clearly lower number of moves per minute (13 and 17, respectively). The most likely explanation is that solvers put more thought and analysis in their moves, as compared to the non-solvers who had a more 'hasty' and 'trial-and-error type' approach towards solving the BMT.

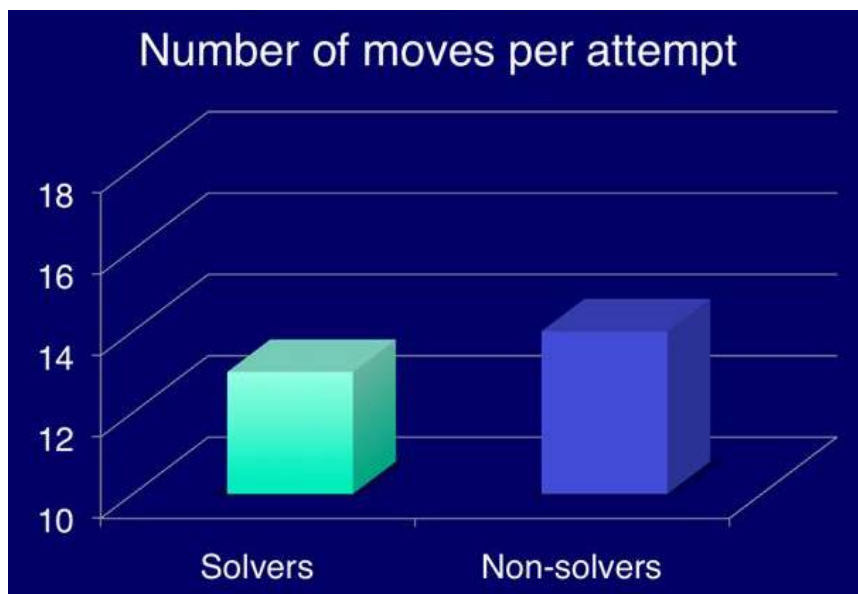


Figure 6.10 : Number of moves per attempt by S and NS participants

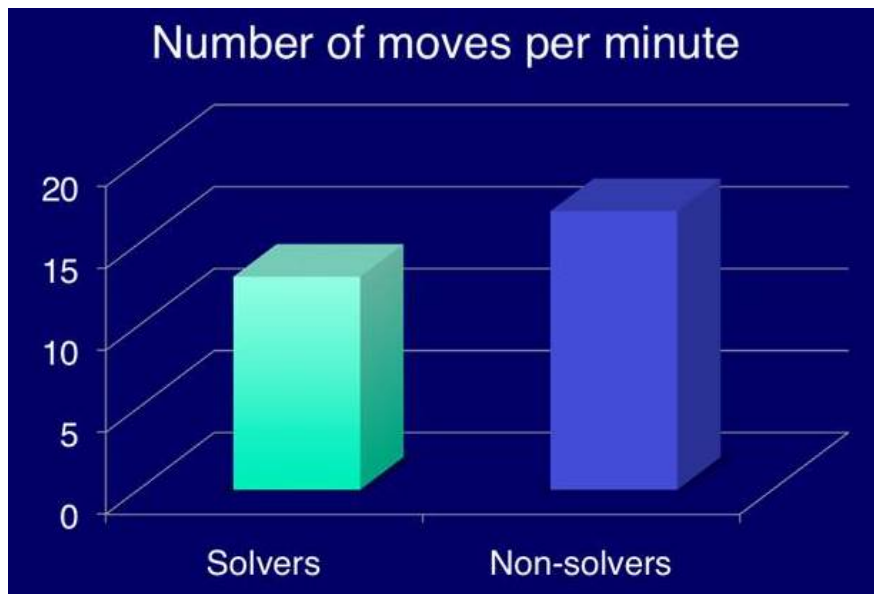


Figure 6.11 : Number of moves per minute by S and NS participants

One subject, who performed the greatest number of moves (246) in the BMT, interestingly had the highest number of moves per minute as well (21.1/minute). This subject was a non-solver. Interestingly compare it with the three subjects with the lowest number of moves per minute (5.4, 5.8 and 6). Incidentally all these were solvers, and the total number of moves that they took to solve the puzzle, were 10, 21 and 12 respectively.

We made a further analysis for solvers by breaking them down into two groups, those who solved the puzzle in less than or equal to 200 seconds (let us call them fast-solvers), and those who took more than 200 seconds to solve the puzzle (slow-solvers).

If we look at the figures of fast solvers, the trends seems to continue, with the contrast (differences in values) being even more marked, for example the average number of moves for the NS and slow-S (only counting the slower-solvers) being 186 and 88 respectively, but in the case of fast-S, if one considers fast solvers only in isolation, it comes out to be 20.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Likewise, the number of attempts made by NS, slow-S and fast-S shows the same trend with their respective average values being 10, 7 and 1, respectively. (See figures 6.12 and 6.13)

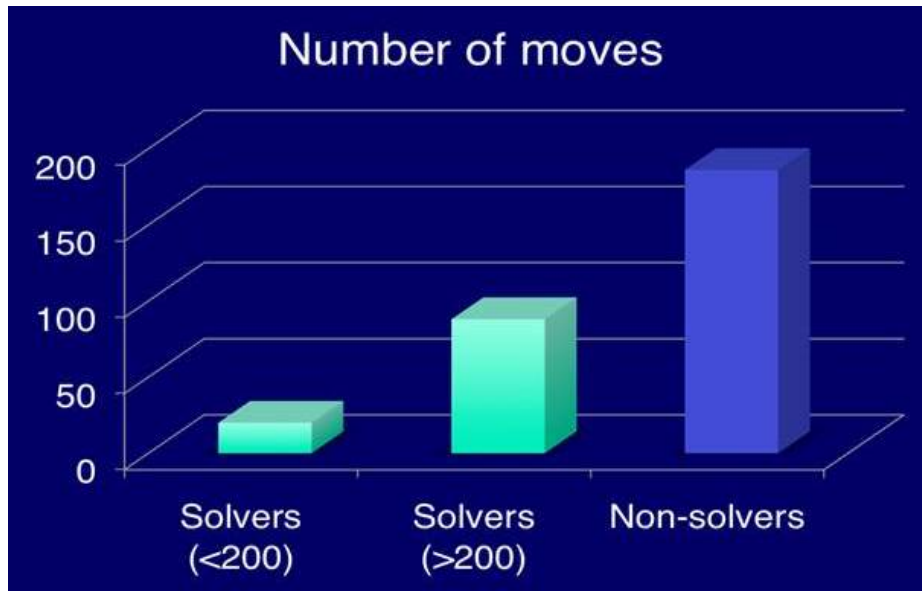


Figure 6.12 : Number of moves by fast-S, slow-S and NS participants

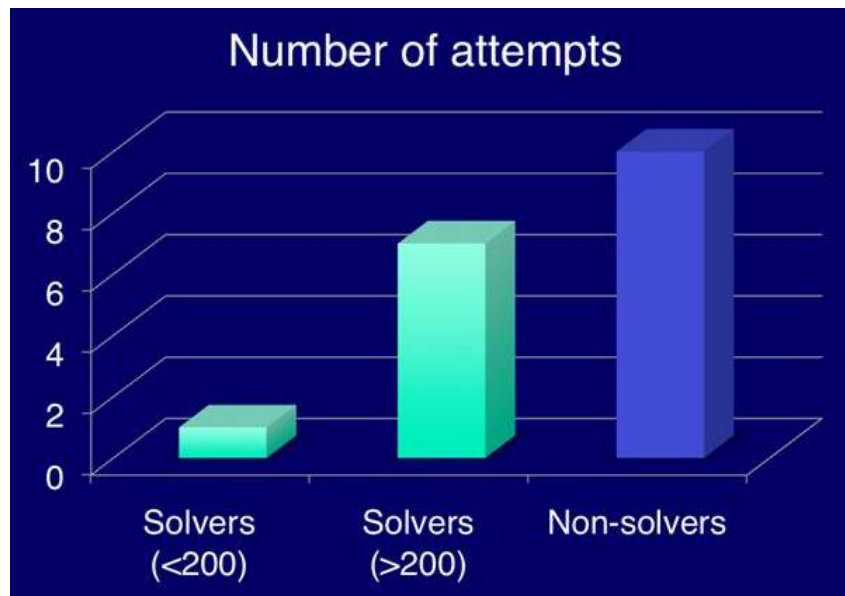


Figure 6.13 : Number of attempts by fast-S, slow-S and NS participants

However, as we said before, the difference in values for number of moves and attempts is inherently coupled with success. The more important differences are to be measured in the ratios for number of moves per attempt and number of moves per minute. And that indeed had some interesting results, for example the number of moves per attempt was not different for NS and slow-S (14), while, more importantly, the fast-S made significantly less moves per attempt (11) (figure 6.14). The fast-S subjects thus seem to be apt at overall executing less moves, but much better selected moves.

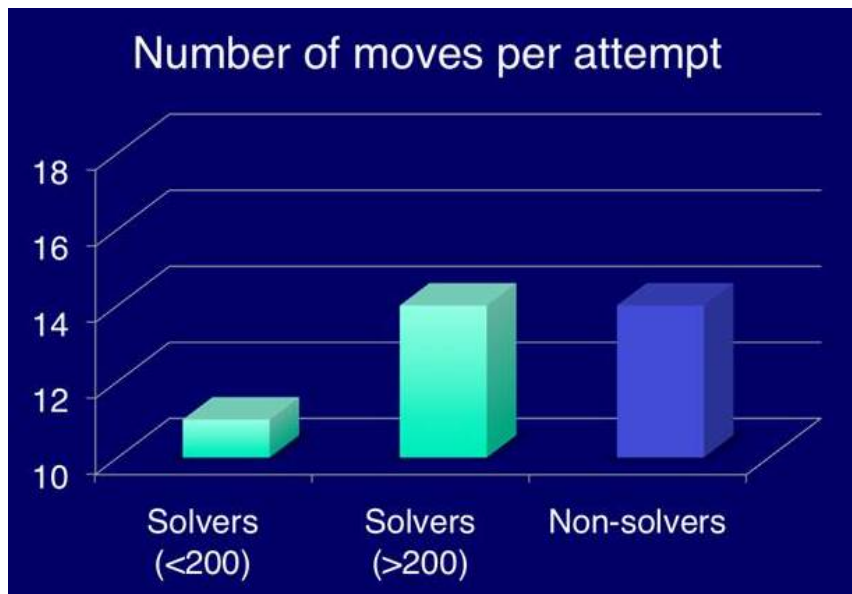


Figure 6.14 : Number of moves/attempt by fast-S, slow-S and NS

And finally the fast-S are in fact the least hasty of them all, in that their average number of moves per minute is the least (11) in comparison to the other two groups (15 and 17) (figure 6.15). This consistency in trend, in fact, suggests that slow-S invest more time in analysing their moves rather than performing an action by trial-and-error.

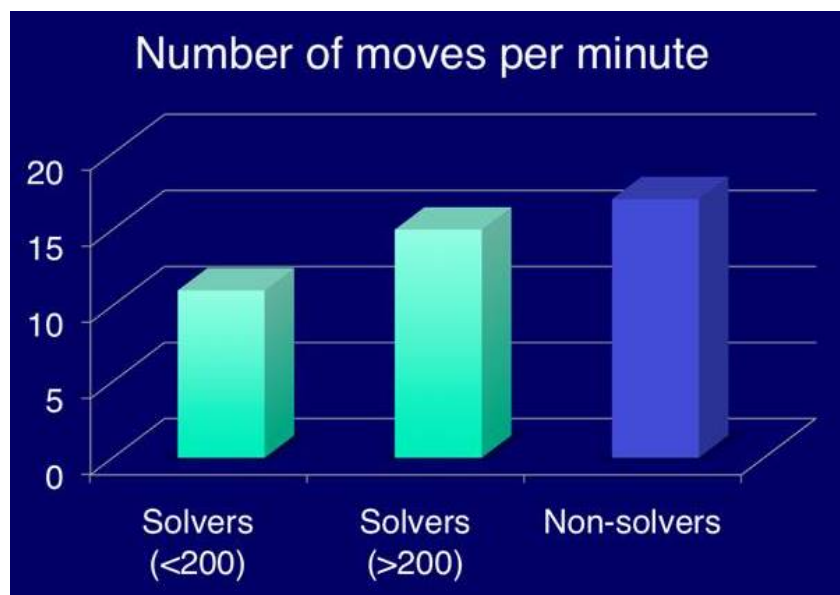


Figure 6.15 : Number of moves/minute by fast-S, slow-S and NS

6.6.3 Types of moves

Amongst all the moves made by the subjects, the most common moves were “Add” and “Remove” as together they constituted more than half of all the moves made (with frequencies of occurrence of 841 and 432, respectively). They were followed by “Pick/Leave” (172), “Rearrange” (139) and “Try” (124) moves. The least frequent were both the compound moves (“Rotate” and “Disassemble (all)” occurring 30 and 32 times respectively).

We observed a weak (but significant) relationship between the fact of being S and NS and the nature of moves performed (Cramer's $V^2=0.031$; $\chi^2=60.0538$, $DoF=10$, $p<0.000$). Relative deviations show that S subjects avoided “Remove” (RD=-0.32) as well as “Try/Hold” (RD=-0.24) moves, whereas they appeared to favour “Pick/Hold” (0.21), “Remove/Hold” (0.35), “Leave” (0.40) and “Rotate” (RD=.40) moves as compared to NS. NS subjects did more frequently “Remove” (RD=0.21). They avoided “Remove/Hold” (-0.23), “Leave” (-0.30) and “Rotate” (-0.20).

But what is more interesting to observe is that NS outnumbered S in all the moves with the exception of two moves: “Leave” and “Rotate”. These moves are both perhaps more related to analysis than just with action. On the other hand, the NS performed almost three times as much “Remove” moves as their S counterparts.

While S performed much less overall moves (766) as compared to that of NS (1178), they performed almost as many number of “Pick/Hold” moves as the NS, which again implies a greater tendency (percentage) of moves which by their very nature imply thinking before putting a block onto the solution. See table 6.6 for the complete frequency list of all the moves.

Table 6.6 : List of actions and their frequencies

MOVE/ACTION	SOLVERS	NON SOLVERS	TOTAL
Add	383	458	841
Remove	116	316	432
Rearrange	45	94	139
Try	43	81	124
Try/Hold	12	28	40
Pick/Hold	20	22	44
Remove/Hold	32	28	60
Leave	18	14	32
Pick/Leave	70	102	172
Rotate	16	14	30
Disassemble (all)	11	21	32
TOTAL	766	1178	1946

6.6.4 Styles of solving and classification of blocks

In the puzzle task, the participants tended to use certain kinds of solutions, but not all the possible solutions were tried. Also, their styles of solving could be placed in two categories broadly. While some subjects did a fair amount of analysis before even attempting to solve the problem (analytical approach), others used a more hands-on and exploratory approach, in which they tested the interactions of blocks while making sense of the problem at hand, and figured out the solution while manipulating the objects (trial-and-error approach).

The participants' comments revealed that they spontaneously differentiated between what they called "easy" and "difficult" blocks. The same types of blocks were consistently judged as "simple" or "complex" by the participants. This classification was correlated with the perceived complexity of their shapes. See section 5.2.4 above for a discussion, and the formula that we have proposed for measuring the complexity of blocks. Five blocks (# 1, 2, 5, 6, 7) were perceived as "simple" and classified as "easy" while the remaining two blocks (# 3, 4) were judged both "complex" and "difficult". See figure 5.3 for the numbering of blocks. Note that the complexity index as proposed by this formula under section 5.2.4 corresponds well with the complexity of the blocks as perceived by the participants. That is to say, blocks

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

referred to as simple or complex by the subjects corresponded with the indices derived by our complexity formula. The blocks used by S to construct the cube was as follows (table 6.7).

Table 6.7 : Number of times where complex blocks (A) and simple blocks (B) were used during each of the positioning actions, for the 17 participants who succeeded

Positioning	Complex blocks (A)	Simple blocks (B)
1st	7	10
2nd	9	8
3rd	8	9
4th	1	16
5th	3	14
6th	5	12
7th (last)	1	16
Total	34	85

Table 6.7 shows which blocks (simple or complex) were placed at each step of the construction, for the successful final attempts of the 17 solvers. For instance, the first line of Column A shows that a block classified as complex (difficult) was used 7 times as the first placed block out of 17. The frequency with which the complex blocks were used during the first three positioning actions was 24 (that is, $7 + 9 + 8$) out of 34 (71%), whereas for the last four positioning actions, the complex blocks were only used 10 times ($1 + 3 + 5 + 1$) out of 34 (29%). This reflects that for S, in line with our expectations, it is a more spontaneous procedure to start problem solving by using the larger and more complex blocks, to which smaller and more simple blocks are then added.

Analysis showed that there is a strong relationship between the level of complexity of the block and the position in which it was used by solvers (Cramer's $V^2 = 0.157093426$). Looking at Relative Deviations superior to 0.20, we observe that S favoured the use of complex blocks at first (RD=+0.44), second (RD=+0.85) and third positions (RD=+0.65) whereas they avoided to use them at fourth and fifth positions (respect. RD=-0.79 and RD=-0.38).

Oppositely, simple blocks are favoured at fourth position ($RD=+0.32$) whereas they are avoided in second ($RD=-0.34$) and third position ($RD=-0.26$).

Another interesting observation was that not even a single participant solved the puzzle task by manipulating the complex blocks at the end (although this type of solution does exist). The reason behind this event is that it is much easier to start with the complex blocks and then to use the simple blocks to fill the gaps. On the other hand, if someone starts by placing the simple blocks first then it becomes quite difficult to find a gap that is suitable to be filled by a complex block.

Similar observations were also made for solving other spatial puzzles. On the basis of such observations, we suggest that solvers, either humans or computer programs, should categorise all the blocks into simple and complex blocks. The rule is to place the complex blocks first and then the simple ones. A more refined partition can be made based on the complexity of the blocks with assigning price to them: the simplest with price 1, and the most complex with a price of 10. Our view is that the solver that uses more expensive blocks at the beginning has a greater probability to complete the task.

6.6.5 Verbalisations

Although the subjects were asked and encouraged to verbalise their actions as well as their strategies, it turned out to be quite difficult to have them translate into words their strategies or to easily find out how exactly they managed to solve the puzzle. For example, when some subjects were reminded that they were supposed to verbalise, some of them did utter phrases that were something like: "OK, so now I do this, and now, I am doing that...". That is to say, that if later on, one heard this comment, it would be impossible to make any sense as to what that subject meant, or was trying to convey to us. Moreover, even if someone saw a video of it, the verbalisation doesn't really offer any substantial added

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

information. The problem was not that they were bad at solving it. Some of them were in fact very good at that. But it was just the fact that they were unable (or not at ease) to verbalise their actions, and/or their strategies. One of the reasons that one could think of is that because we are not required to verbalise such spatial manipulation actions in our everyday lives, so we have not learnt or evolved to do so, and as a result are much less equipped to verbalise ourselves. In other words, we lack the vocabulary for the strategising, although part of that information could be stored in our brains in some form.

There were no more than three subjects who in fact solved the problem in a very calculated manner and what they said could also be more easily comprehended. They first of all counted how many blocks there were in all, they then counted how many mini-blocks each of those seven blocks consisted of, and then added those numbers up. When the total came to 27 mini-blocks, two of them in fact declared that a solution to the problem was in fact feasible. Both these subjects had a rather slow rate of manipulating the blocks and both of them succeeded in constructing the cube well in time.

Nevertheless, the main question that we in fact were hoping to find an answer to, still remained unanswered. So while one could in some way see that they were tackling the problem in a rather analytical way, however, we still couldn't find out from them answers to questions regarding why they put a certain block in that particular orientation and not in any other orientation.

More than three fourths of the subjects placed the initial blocks in a horizontal manner, and some of them even said explicitly that they were doing so, and a couple of subjects also said that they took a horizontal layer-by-layer approach, meaning that they tried to fill up one layer of the puzzle completely before mounting on to the higher one.

Even if more than half of them did not explicitly say it out, later review of videos indeed verified that it was true for most of the solvers. However, even if this approach does seem to work, no one was able to answer the more fundamental questions, for example, giving the reasons why they had chosen such (horizontal) approach in the first place, or what was it (the structure or some other characteristic of the blocks etc.) that made them opt for such an approach to the problem.

Another tip that was communicated by some solvers was to avoid constructing the blocks in a way that creates a 1x1x1 space in isolation. If such a space is created at a corner of top-most layer's workspace (spaces marked 1, 3, 7 and 9 in figure 5.8-b) while the lower levels have already been constructed, this strategy cannot be successful. This is so since none of the blocks available to the subjects was of that dimension (i.e. of size 1x1x1).

One of the other things that we asked the solvers, in the after-experiment interview was: at what point in time did they realise that they have solved the puzzle? More specifically, they were asked, how many blocks before actually finishing the solution were they able to visualise it. We wished to find out if there existed a big contrast in the amount of blocks that different subjects were able to visualise in their heads. As it turned out, most of the subjects saw it just before or after the second last piece, and there was none who, for example, visualised the solution after having placed just three blocks.

6.6.6 Post-experimental comments

All those who couldn't solve the problem rated it as very difficult (some of those who gave it a quantitative score rated it at 8 or more on a difficulty level from 1-10 with lower index depicting ease and greater index depicting difficulty). When asked to comment about the problem, they used words like "stressful", "frustrating", "hard", and "difficult".

In contrast, most of those who solved it termed it as very easy to medium (with three exceptions: 1 medium and 2 difficult). When asked to comment about the problem, they used words/phrases like “interesting”, “entertaining”, “very easy”, “it’s like a game”, “amusing”, “happy”, and “funny”.

6.6.7 Conclusions

The first result of interest reported above is the fact that different strategic patterns with distinct outcomes for those subjects who eventually solved the problem within the ten minutes allocated and those who did not succeed. Solvers, in addition, were found among those people who have the higher visuo-spatial capacities as tested by classic instruments of the psychometric repertoire.

We also found that our a priori analysis of the characteristics of the blocks suggests that they differ in *complexity*. Complexity refers here to the distinction between two types of blocks based on their shape, simple vs. complex ones. This analysis coincided with the perceived complexity of the blocks by the subjects as they verbalised them as “complex” (or difficult) blocks, and “simple” (or easy) blocks. The subjects who solved the puzzle systematically put the more complex blocks in the solution much before the last two steps. At least one of the two complex blocks was systematically placed in one of the initial three positions.

Having looked carefully to how people proceed in manipulating blocks in a real setting, we are now in a good position to examine how they will handle virtual blocks in contexts involving various modalities.

7 Chapter 7: The Virtual 3D-Puzzle

This chapter reports and discusses the virtual puzzle based experiment, which was conducted in the virtual settings. First, the hardware and software used for the experiment as well as the overall setup (for example, the physical settings) for the experiment are discussed. The experiment in the virtual settings was conducted using three input modalities, the Keyboard modality, the Gestural modality, and the Vocal modality. The various aspects of the interface of the various elements are discussed, such as: the differentiation between a given block in a selected state and a deselected state, the different operations that could be performed over the various blocks, such as the translation and rotation in different directions, the feedback given to the users when they interacted with the blocks (that is, what did the users actually see when they translated or rotated a block, such as the shadows of the blocks, or an arrow emanating from the block in the direction of its movement, etc.). Then, the overall protocol adapted for the experiment in virtual setting is discussed, followed by the results and discussion over them.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Based on our initial experiment in the real settings, we designed and conducted the experiment for the virtual environment. The experiment in the virtual environment was based upon the one conducted in the laboratory environment. It was tried to make this VE experiment as close to the one in RE as possible. We shall call it the 'Virtual Puzzle'.

Like the real world experiment, the virtual puzzle contained seven independently manipulable blocks. The experiment tested not just the participants' spatial manipulation skills, but also the mental component of problem resolution. Images were projected on a large (larger than life) screen on one of two screens in a CAVE-like environment.

The overall task was the same, i.e. to build a cube from the seven blocks of the same size and shapes as were used in the experiment in the real settings. Other fundamental decisions were deliberated, like 'Should the users be allowed or not to have total control over the rotation of the blocks?'; 'Could the users' manipulation of blocks could be aided from the view-point of the task at hand; and if so, how?' Moreover, decisions on feasibility/availability of characteristics of the virtual puzzle were discussed, for example the kind of devices to which we had access at the given time (for example, CAVE-like screen, wireless gloves, tracking system etc.), and what features were provided/available in the application chosen (for example properties associated with the projection and interaction of blocks like shadows, collision-detection, material properties, friction, effect of gravity, etc.)

Normally, the interactions techniques, the modalities and the devices used all go hand in hand, i.e. each can have an impact on others. However, it may not be possible to describe all of them concurrently, thus each one of them is discussed one by one.

7.1 Hardware/software

The virtual reality based experiment was developed using the 3DVIA Virtools software [Virtools] (Dassault Systèmes). It was conducted in one of the two CAVE-like screens at the LIMSI-CNRS at the former Virtual Reality room of the LIMSI-CNRS lab, at the University of Paris (South-XI), at Orsay, France and at the ECI (Ergonomie, Comportements, Interactions) Lab, of University Paris Descartes. The A.R. Tracking System [ARTracking] was used for the gestural modality, while “Windows Vista’s” built-in Voice Recognition System was used for voice training, recognition and input for certain parts of the experiment. Figure 7.1 shows the system functional diagram where the arrow (o_1) pointing towards the user signifies the output, while the arrows (i_1 , i_2 or i_3) emanating out represent the different devices/modalities through which the user gives input and interacts with the system.

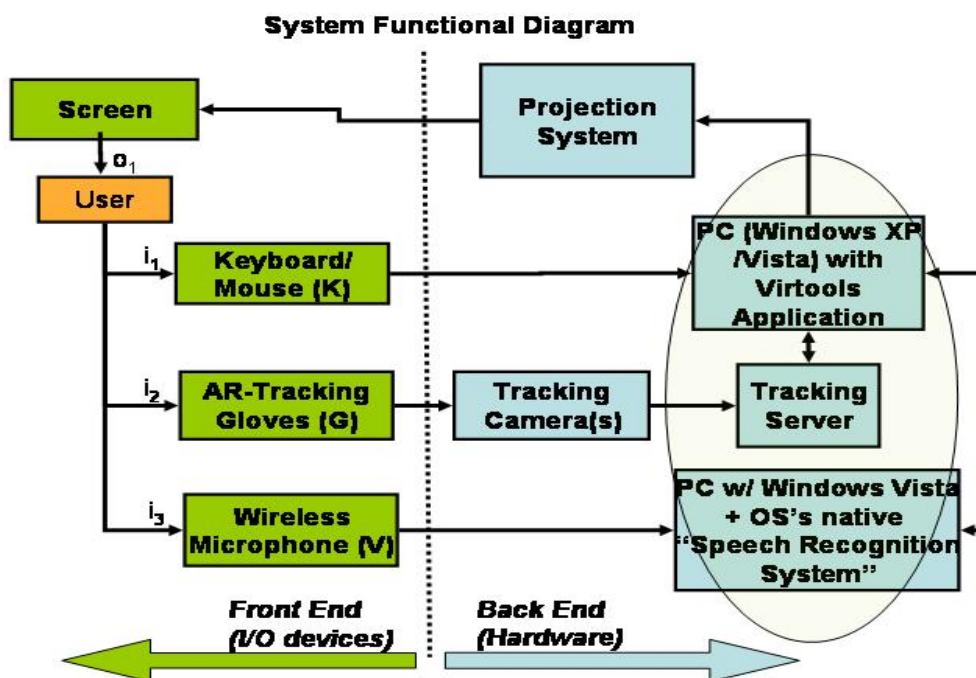


Figure 7.1 : System functional diagram

7.2 Big picture: The overall scene and the setup

The scene was projected on a big screen (2 metres width x 1.5 m height). The scene contained a big table with a lamp on it (both of them fixed).

As shown in Figure 7.2, the virtual puzzle contained the seven coloured blocks (not fixed) meant to be directly manipulated by the user to solve the 3D puzzle, and a sphere (view ball) which the user could use for changing the scene's point of view. Only one object was selectable at a time.



Figure 7.2 : The scene for the Virtual Puzzle containing 7 coloured blocks

The user was placed in front of the screen. His/her task was to construct a cube by assembling the seven given blocks. (See section 7.6 below for the details on exactly what the users were required to do).

7.3 Little picture: The components of the scene and the behavioural aspects common to all modalities

The following section (section 7.4) discusses the choices made in the design and implementation of the virtual settings experiment in some detail. But before we proceed

with the section on modalities, it may be in place to first introduce all the basic blocks and other elements used in the experiment in the virtual settings.

7.3.1 The convention for axes

Whenever the terms X, Y and Z axes are used, the X axis is taken to be the horizontal axis, the Y axis to be the vertical axis, and the Z axis to be the depth axis.

7.3.2 The blocks

Figure 7.3 (reproduced here from figure 5.3) shows the seven blocks.

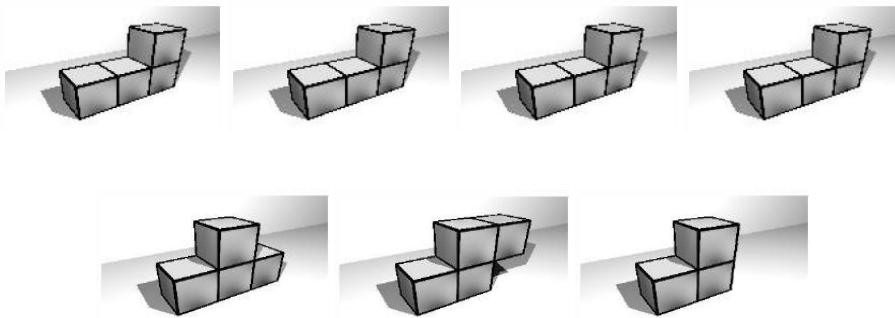


Figure 7.3 : The seven blocks - composed of 3 or 4 mini-blocks

7.3.3 Selected and deselected blocks

At any given time, only one block could be selected (and/or manipulated). The block when selected became white (see Figure 7.4). Any given manipulation could thus be applied over this white (i.e. selected) block. A block could be selected i) when no other blocks were previously selected, and ii) when another block was already selected. In the first case, the block upon selection simply changed its original colour (for example red, blue, yellow, etc.) to white. In the second case, when a new block was selected while another one was already selected, it led to two actions: a) the block that was previously selected got deselected (and

changed its colour from white to original); and b) the newer block got selected. How exactly a block was selected depended upon the modality. The input details of how precisely the selection of a block could be made, was different for each modality, and is thus is described under the details for each modality below.



Figure 7.4 : Colour of the block changes from original (in this case, yellow) to white

7.3.4 Arrow emanating from the moving objects denoting the axis of movement

Arrows emanated from the block in the direction in which it moved. Arrows of different colours were used for each axis – red for the X-axis (horizontal), green for the Y-axis (vertical), and blue for the Z-axis (depth).

7.3.5 Casting of the shadows

The blocks casted shadows on the table as well as over other objects. This characteristic and the two following this one were incorporated into the system from the perspective that they helped recreate the scene to be perceived closer to reality, as well as allowing the user to judge the respective positions of all the objects in the scene in the three dimensions more accurately.

7.3.6 Collisions

There was *collision detection* amongst the blocks, i.e. they did not occupy the same space at a given time, or did not merge into each other nor disappeared under the table. When the user moved a block, it came to a halt whenever there was any “obstruction” in the way of that block (for example the table or another block).

7.3.7 Occlusion

If two blocks were placed one behind the other, the one in the front was visible completely just as in real life, while the one at the back was occluded (blocked from view) by the one in front.

7.3.8 Six degrees of freedom (6-DOF)

Blocks could be moved (i.e. translated) on all three axes. While the movement of the blocks was perceptually continuous, however at the time of their deselection (i.e. when the block was released), they adjusted themselves as per an imaginary grid (making the resting space for blocks slightly discrete, or rather quantised). This grid was fairly fine, and thus the auto alignment could at most be perceived as a slight jerk. Moreover, since the user deselected the object, it seemed quite natural and was hardly perceived.

For the rotation, similarly, the block could be rotated around the three axes. Here too, the orientation that the blocks could acquire was *canonical* and they could be rotated in chunks (or multiples) of 90 degrees. This corresponded well to the behaviour of blocks in real life (which also could rest stable in orientation in multiples of 90° all of which were flat in correspondence to the surface of the table). When rotated, the movement (animation) of the block was smooth and not discrete, which made it easier to observe what action was being performed on the block. If the block was moved so that it didn't have enough space to

rotate all the way up to 90° (for example if it was obstructed by another block or the table), it returned back to its original position.

7.3.9 The view-ball

In addition to the blocks, there was a sphere – the “view-ball” as shown in Figure 7.5 which represented the user’s point-of-view – i.e. the point from which the user viewed the whole scene.

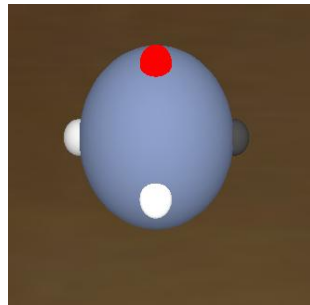


Figure 7.5 : The view-ball

The purpose of the view ball was for the user to be able to change the scene’s point of view. The user could either move around the blocks, or zoom-in/zoom-out of the place where the blocks were to be supposedly assembled – the centre of the table.

The user had four points of view available. He/she could rotate around the scene by first rotating the *view-ball* around Y axis. When the *view-ball* was rotated, the whole scene moved as a result.

Note that in comparison to the manipulation of blocks, there were some limitations to the manipulation of the view ball, in that it could only be rotated around the vertical (Y) axis

(allowing the user to change the view *around* the room, so for example it was not possible to view the blocks right from above (i.e. the top of the room), or from underneath the table).

As was the case with other blocks, the *view-ball* had to be first selected before it could be rotated. Similar to the other blocks, the *view-ball* too could only be rotated in chunks of 90°. (Imagine a camera rotating in a circle around the room, but which was always looking at and focused towards the centre of the table).

7.3.10 Virtual levitation (lack of gravity)

Note that there was *no gravity*, that is to say that if a block was left in midair and deselected, it remained there, i.e. it did not fall down at the table.

7.4 Existing research on modalities and interaction techniques and the choices made for the Virtual Puzzle

There exist a vast number of interaction related and system choices available for experiments in VEs. The devices and interaction techniques development has become quite stable and standardised in case of 2D devices and interaction techniques, with many devices and interaction techniques having become de facto standards [Brooks1999]. In the case of 3D, things are still very far from being considered as standardised. In fact, not only are new interaction techniques (and modifications to old techniques) being proposed all the time, but entirely new devices are also introduced from time to time (like The *Cubic Mouse* device and interface by [Fröhlich2000]).

Numerous VR/VE related studies have been done in the past that have tried different kinds of modalities like visual, gestural, vocal, auditory, haptic, gaze/eye-tracking, face expressions, body movements, etc., or some combination thereof, while some others are

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

more focused on taxonomic review of devices and modalities that have been tried. Typically, the term *multimodality* is used when more than one of these modalities are involved (see [Jaimes2005], [Zhai1998], [Brooks1999] for a vast list of devices of different modalities). It becomes quickly apparent that for 3D manipulation of objects in VEs, first of all, there are a number of different modalities such as the ones enumerated above (or a combination thereof for multimodality) that one can choose from; and secondly, each of those modalities has its own select choices of design, characteristics, and features, thus making the choice of modality and design quite a complex one.

In addition to that, even when the choice of modality has been made, there exist a plethora of choices available regarding the precise user-system interaction techniques [Bowman2006], [Bowman2002-b], [Poupyrev2000]. A taxonomy of devices and a near exhaustive treatment of different interaction techniques is available in [Bowman2005].

The selection of a specific modality (and within a given modality, the selection of a specific device and the associated interaction techniques) is quite vast. Classifications of the different input devices are available. Some of them are made from the point of view of the human senses of sight, touch, hearing, smell, and taste, since the input modalities of many computer input devices can be considered to correspond to human senses. [Jaimes2007].

- vision (sight): cameras, body, gaze, gesture
- audio (hearing): microphones
- touch: haptic, tactile
- olfactory (smell), and
- taste

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Some classifications contrast spatial vs. non-spatial devices (for example [Jaimes2007], [Stanney2003]). Within the spatial devices, further distinction is made between *isomorphically controlled* devices (i.e. devices measuring the position of the user's hand, such as mice or motion tracker) and *isometrically controlled* devices (devices measuring the force or tension applied by the user, such as joystick) [Zhai1993], [Bowman2005].

However, we found the taxonomy proposed by Bowman to be quite a comprehensive one. It is reorganised below in tabular form for easier review (table 7.1).

Table 7.1 : Taxonomy of input devices used in 3D VEs

Device category	Devices
Desktop input devices	<ul style="list-style-type: none">• Keyboards• 2D mice and trackballs• Pen-based tablets• Joysticks• 6-DOF-devices for the desktop
Tracking devices	<ul style="list-style-type: none">• Motion trackers (including magnetic tracking devices, mechanical tracking devices, acoustic tracking devices, etc.)• Eye trackers• Data gloves (including Bend sensing, Pinch gloves, combined)
3D mice	<ul style="list-style-type: none">• Hand held 3D mice• User-worn 3D mice
Special purpose input devices	<ul style="list-style-type: none">• Interactive slippers, transparent palettes
Direct human input	<ul style="list-style-type: none">• Speech input• Bio-electric input• Brain input

In fact, there are different points of view in even what defines a modality. For instance, for some, the input method or device is what really defines the modality (e.g., gestural, audio,

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

eye-tracking, etc.) with output being constant (for example [Ruiz2010]). However, one can find others who define a modality in terms of the kind of feedback the user receives for example as in [Nam2008].

Moreover, it is interesting to note that many of the studies in the VE domain (particularly the immersive VEs) are concerned usually (if not purely) with the visual aspects of the interface in some form. Whether they are concerned with the interaction techniques like "selection", "manipulation" and/or "navigation" of objects in the digital and VEs [Bowman1997], [Bowman2001], [Bowman2002-a], [Abbasi2006], [Poupyrev1997], [Poupyrev2000] or whether they are comparing the different input and output devices [Jaimes2007], [Zhai1998], there is quite usually some sort of visual aspect associated with those techniques. This is perhaps because regardless of which input device is used there is almost always a visual feedback involved. Moreover, whenever these works consider other feedback(s), it is *in addition* to the visual one, such as visual-only (V), visual-haptic (V + H), and visual-haptic-audio feedback (V + H + A) by [Nam2008]. Of course there are certain exceptions, and non-visual modalities and interfaces are studied as well, such as in the experiments that are specifically for the blind (for example [Peres2008]), but such studies are more of an exception than a rule.

In the classical digital world, where everything was in two dimensions, or at least where everything could at least be mapped onto two dimensions, it was possible to create fairly generalised interfaces which could work in virtually all situations. However, when things move on to 3 dimensions in the digital world, its complexities increase manifold. The arising situations and issues that one confronts are closer to real world, as compared to its sibling - digital 2D world. Therefore, in order to be able to deal with issues in the digital 3D world, the approach taken also takes its inspiration from the real world rather than the classical digital 2D world.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

If the above claim is indeed true, then as in real life, where we have different tools for performing different tasks (each having its unique interaction and operation techniques), and no one tool being ideal for all tasks, so too is the case with VE applications, there isn't any one specific universal device and interaction technique that could qualify to be the best for all situations and applications. Indeed, the decision as to which specific 'interaction technique' or 'modality' is best for any given situation is a difficult one because no one modality is best for all situations. What makes more sense for one kind of application may not be the best or so well suited in another situation. The appropriate interaction techniques ought to be evaluated as a function of the task and the devices available at hand [Bowman2005]. As a general rule, the devices themselves ought to be a function of the task at hand.

Whatever the input modalities for our virtual puzzle were going to be, it was thought that it would indeed be a good idea to compare them with the current standard (or the traditional) device as a sort of a benchmark for its digital counterparts, since without it, a direct comparison of real world with new devices does not make a lot of sense. Thus in our case a combination of keyboard and mouse were considered as the standard input devices.

From the general viewpoint of analysing the design of interaction techniques, a good interface design can be summed up as one possessing the following characteristics: *visibility* [Rehman2002], *feedback*, and *system transparency* [Norman1988].

Visibility: In a nutshell, the principle of visibility suggests that all things should be visible to the user. Characteristics/features of the system, the current state of the system, checking if everything is going fine, or if there is a problem, regardless of whatever the case is, the system should make that visible to the user. The user should be easily able to see what actions are available for him/her, and he/she should be able to perform those actions easily.

Feedback: The user should be able to get a feedback continually, even when the system is performing normally, even when a given action is successfully performed (and even when the system is taking the desired action as per the intentions of the user). The user needs to be reassured that everything is working correctly.

System transparency: As an overlap of the previous point, even if the system is in the desired state, the user should be able to expressly reaffirm that the system is indeed in that particular state by looking at the system; and if that is not the case, the user should be able to tell which state the system is in. Ideally, the user should also be able to easily gauge which actions are available that he/she could (or should) perform to remedy the problem [Norman1988].

7.4.1 Our chosen modalities

Gestural (G): Since ours was a manipulation-centric task that was also visual in nature, one of the obvious decisions taken was to test users' performance by enabling them a hands-on direct-manipulation interface, which would permit them to virtually directly manipulate the objects (or blocks). Thus, one of the chosen modalities was the gestural one.

Vocal (V): Imagine devising a working system that would enable the user to solve the puzzle by navigating the blocks by giving verbal commands, as if one were to give instructions to someone in the real life and in real time. Imagine a situation where you are giving directions to another person through a walkie-talkie, whom you could observe from a distance.

One would wish to pass instructions to the system as one would pass them to a person, using natural everyday language commands. This of course could also have applications especially in the field of wayfinding and navigation.

It is quite usual to combine the individual modalities to create the multimodality. In our case, a potential possibility was also to combine the gestural and vocal modalities and to be able to execute commands in the famous “*Put-that-there*” fashion. We wished to provide the user with a much broader set of manipulation choices available, where he/she may either use one method or the other (gestural or vocal) uniquely and in isolation, or the two in combination as per his/her personal preferences. Just as the choice of modality exists in real life also, for example if one so wished, then one could give a command in the ‘PUT-THAT-THERE’ fashion, or one can pass a command in entirely verbal terms such as ‘*put the ball inside the hole please*’ [Bolt1980].

7.4.2 Other related issues

For our experiments, the users were to perform the task on one of the screens in the CAVE-like environment, using different methods of input including the direct manipulation. But even with the choice of modalities made, there remained a long list of questions to be answered about the interaction techniques. Broadly speaking, for spatial manipulation, two broad categories of input devices exist. As mentioned above, *isotonic devices*, which are free-space devices (for example, a mouse), and *isomorphic devices*, which work on the principle of tension/pressure (for example, a joystick). Both kinds of devices have their advantages and disadvantages. The main advantage of *isotonic* devices is that they convey a more natural feeling, but on the other hand the devices have their space limitations (and thus also necessitating a clutching mechanism). [Abbasi2006] gives a simple introduction to some famous isotonic devices. *Isomorphic* devices may feel a bit less like real life, and in general they also provide less *sense-of-presence* to the users. But once the user develops a sense of mapping, the advantage of such devices is that the user can navigate large distances

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

or manipulate objects at large distances, even in restricted spaces (and without the need for a lot of space for the user to move in).

While some authors define “presence” as “the feeling of being there”, according to [Stanney2003] it is the subjective perception of experiencing oneself as being in a computer-generated environment rather than in one’s actual physical location. From the usability point of view, the users’ “sense of presence” and “immersion” are two similar yet different aspects of verifying the user's engagement. While the sense of presence of course does also depend upon the level of reaction of a VE system to the user's actions, which should be robust and have no perceivable lag or distortion of motion, it is not synonymous to the “realism” built into the system (for example, while some may measure it using methods such as post-experimental questionnaires, this experience is highly subjective. Moreover, another way of measuring the presence is by calculating the level of the user's motivation to continue to maintain involvement in VE interaction, which may be completely independent of realism, such as in video games. For example, this sense can be developed by creating situations that demand higher interaction on the part of the user, rather than creating more realistic systems). In short, while the realism of the VE system may play an important role in increasing the so-called sense of presence, it is believed that it is more closely associated with immersion than with presence.

As the field of VE progresses, these issues (including the issue of the sense of presence) are being increasingly considered as important ones from the usability’s perspective (see for example, [Zahoric1998], [Vinaya2004], [Nichols2000], [Brogni2003], [Slater1995]). Another related issue which used to be ignored but is being increasingly considered, is that of the “side-effects” of the VE/user interface. Side-effects include user’s comfort (or discomfort), sickness (e.g., motion sickness, muscle/eye strain due to prolonged usage of the VE), and psychological after-effects that may jeopardise the experience, the well-being, or learning capabilities of users both during and after VE interaction [Stanney2003].

7.5 Description of modalities for the experiment in the virtual settings

This experiment was conducted using three different modalities: Keyboard/Mouse (K), Gestural (G), and Vocal (V). The first one was the classical visuo-Keyboard and Mouse modality with visual feedback and keyboard and mouse as the input devices which we shall refer to as 'K' modality. The second one was the visuo-gestural modality, which we will call the *gestural* (or G) modality. In this modality, the users could control the blocks with their hands. The third one was visuo-vocal modality, which we will call the *vocal* (or V) modality. In this modality, the users could control the blocks with voice commands. The details for each modality are discussed below.

7.5.1 Keyboard/mouse: The classical modality

The keyboard and mouse combination was included in our study as the classical modality, firstly because it is the one with which most users are very well versed, and due to this very reason, there is no learning overhead specifically associated with interaction with the device; and secondly, because it could be used as the standard modality to compare the other modalities with.

7.5.1.1 Selection/deselection

For this modality, the item could be selected by simply clicking the object with the mouse. The selected object changed colour to white. If a block was already selected (white), it could be deselected by clicking over it again with the mouse. Since only one block could be selected at a time, if one block was already selected and the user clicked on the other, it resulted in two actions: a) deselection of the previously selected block (and that block turned back to its original colour), and b) selection of the new block, and the colour of the newly selected block turned into white (see figure 7.4).

7.5.1.2 Translation

The selected object could be translated on the three axes X, Y, and Z, and it could move in either direction (positive or negative) on those axes. So, this meant two buttons for each axis; and thus six for all three axes. Just as the traditional 2D systems displaying and manipulating 2D information, two buttons are normally used for each axis (for example in the classic game of Pac-Man, the user can manipulate the Pac-Man on the 2 (X and Y) axes using 4 cursor keys). Let us take the horizontal axis as the X axis; then traditionally, the user normally has two buttons to manipulate his Pac-Man on the X axis, the right arrow key to move to the right (the positive direction of the X axis), and the left arrow key to move to the left (the negative direction of the X axis). Thus just as manipulating objects in 2D generally requires four keys, thus for manipulation in three dimensions, we needed $(3 \times 2 = 6)$ six distinct keys.

We tried to capitalise on the established conventions by using the keys that are used most frequently for spatial tasks and which also have natural spatial association. The obvious choice was to reuse the four directional arrow keys traditionally used for spatial navigational purposes. In addition to those, we chose the numeric keypad keys '3' and '7' for the navigation in the third dimension.



Figure 7.6 : The direction keys

Figure 7.6 shows the four arrow keys - also known as the cursor keys (bottom left), and the keys 3 and 7 on the numeric keypad (right) used for translating the selected object in the third dimension. When pushed in combination with the left Control (Ctrl) key, these same keys changed the orientation of the block.

When the user pushed a key, the block moved in a particular direction with a constant speed. Releasing (un-pressing) the key made the block stop moving. That is, the keys had to be kept pressed continuously for the movement to continue.

7.5.1.3 Rotation

The selected objects could be also rotated on the three axes in either direction. The same six keys (used for translation) were reused. Thus, when used in combination with the 'Left Control' (Ctrl) key, it made the selected object rotate. This was a conscious decision, firstly since as only one kind of manipulation was possible at a time, and secondly because having chosen six different keys for rotation would have meant to oblige the user to remember 12 different keys for manipulation, which would have put a very heavy cognitive load on the user.

As explained above, the rotation takes place in multiples of 90°, thus even if the direction key was pressed briefly (in combination with the "Ctrl" key), once the selected object (i.e. the selected block or the view-ball) starts rotating, it continues to do so until 90° rotation is completed. In case something obstructed the block (for example another block or the table) from completing its 90° rotation, it retreated back to its starting position.

7.5.1.4 Zooming-in and zooming-out of the scene

It was possible to zoom-in and zoom-out of the scene by means of pressing the “+” and the “-” keys of the numeric keypad respectively. The keys ought to be kept pressed-down in order to zoom in or out continuously, and as soon as the keys were released, the zooming stopped. There was a limit to zooming-in to the scene, and zooming-out of it. The focus of the camera was always at a central position of the table.

It may be worth adding that the selection and movement of the view-ball was done in the same way as the other blocks, with two exceptions: The first was already discussed above that the view-ball may only be pivoted, i.e. rotated around the axis vertical (perpendicular) to the table. And second, the collision-detection does not hold over the view-ball. Thus it could be placed just about anywhere in the scene, and it could even be (sub)-merged inside other objects. Thus for example, it could be placed away from other blocks and partially merged beneath the surface of the table (such a thing was not possible for other blocks).

7.5.2 Gestural modality

7.5.2.1 Setup specific to this modality

The infrastructure setup for the gestural modality for virtual puzzle included Optical tracking system that used three tracking cameras and two hand-targets. Optical tracking systems allow three-dimensional input for virtual environment applications generally with high precision. They are wireless and allows to track the orientation of the hand and the position of the fingers. We used the one which is known by the name of A.R. Tracking system or A.R.T. [ARTracking]. It uses position and orientation of the back of the hand, as well as the hand gestures based on the movements of the thumb, and finger movements.

The hand-target was worn on each hand by the user. On the screen, the user saw two images of hands which corresponded to his/her own hands (figure 7.7). The setup required

an initialisation in which the user was to wear the hand-targets and keep his hands facing downwards, with thumb and fingers fully open and pointing to the central tracking camera (see figure 7.8). The movements required on the users' part were fairly natural except for the fact that the user was required to move his/her hands so that they were visible to the tracking cameras. Three tracking cameras were used, all of which were located above the user, which necessitated that the user's hands were always facing down.

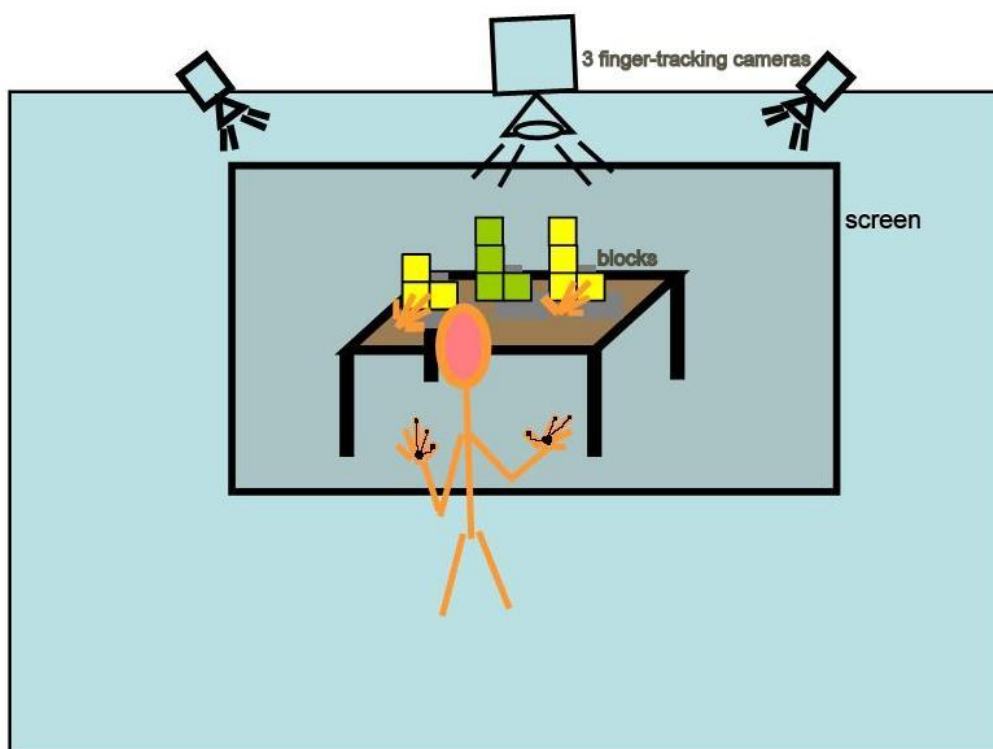


Figure 7.7 : The infrastructure setup for the gestural modality

Figure 7.7 shows the infrastructure setup for the gestural modality. The functional diagram includes three tracking cameras and two hand-targets. The hand-target included a tracking device to be worn over the back of the palm-top, as well as the thumb, the index and the middle finger.

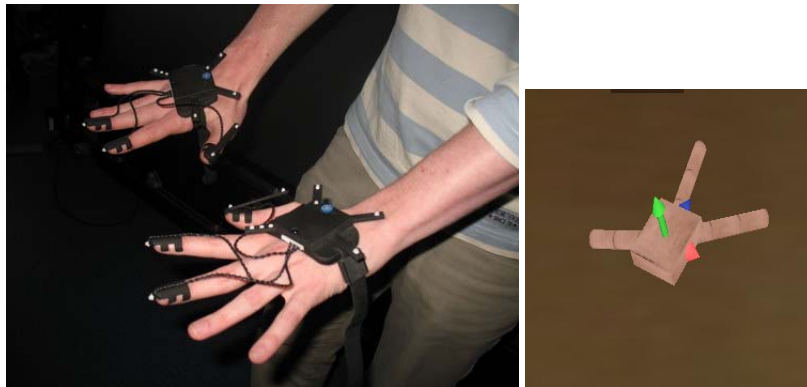


Figure 7.8 : Calibration for the gestural modality

Figure 7.8 shows the calibration (palm flat + fingers fully open), and a trident is projected on the screen representing the image of the user's hand and mimicking his/her hand movements.

7.5.2.2 Selection/Deselection

There were two ways to select an item using this modality: by the "pointing and selecting" method and the "grabbing" method, the near objects may be selected by either of the two methods while far objects may only be selected by the former (i.e. pointing and selecting) method.

Selection by the 'pointing' method

The pointing method involved a three-step process. The user was first required to point to the object with his/her index finger. When the user made the pointing action, a ray emanated from the user's index finger on the screen. At that point, the user needed to carefully point the emanating ray so that it passed across the desired object (figures 7.9 and 7.10). Then he/she was required to make a rapid (spasmodic) action of the pointing finger (and indeed the whole arm) towards the desired object so that the extending ray passed 'through' the object that one wished to select. We called it the *javelin-fishing method*.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

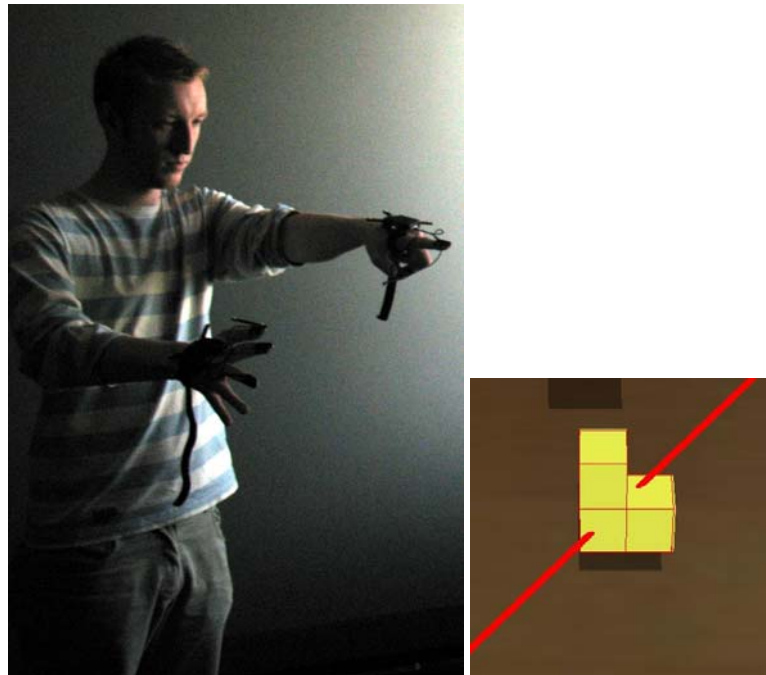


Figure 7.9 : Pointing at an object during the gestural modality

In figure 7.9, (left) a user is pointing his index finger towards the object; (right) a laser beam is emanating out of the user's hand and passing through a block.



Figure 7.10 : A user is pointing and selecting an object

Figure 7.10 shows that a user is selecting the object by 'pointing' using the 'javelin-fishing' method. The user is pointing towards the red block. On the screen the ray (laser beam) extends from the user's index finger, and passes through the red block. To select the object, he must now make the laser beam pass through this red block in a rapid (indeed jerky) action.

When the user makes this jerky action, the object comes flying into the user's hand. An advantage of this 'pointing' method is that one can easily get hold of objects that are placed far from the user.

Selection by the 'grabbing' method

The other method required the user to make a grabbing gesture to select a desired object. The grabbing gesture was similar to grabbing, however, it was in no way identical to the way

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

one grabs objects 'naturally' in the real world. Nevertheless, it was a constraint imposed by the system. The grabbing method required a rapid action of closing (clasping) the fingers in a way as if holding a pen, while keeping the back of the palm horizontally upwards and parallel to the floor (figure 7.11).



a)



b)

Figure 7.11 : Grabbing action

In Figure 7.11 a) for grabbing action in the left image the palm is flat with fingers open, and in the right image the fingers are closed rapidly towards thumb, while grabbing the object.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Care had to be taken that all those actions were visible to the cameras installed above the user as shown in the gestural infrastructure setup figure above. Note: this example is a recreation where the user is grabbing a real pen. b) Shows the hand reaching for the object, and once inside, the user can grab it by clenching his/her hand.

To deselect the selected object, one needed to swiftly release open the fingers away from the thumb (figure 7.12 or the two images in figure 7.11 but in reverse order).

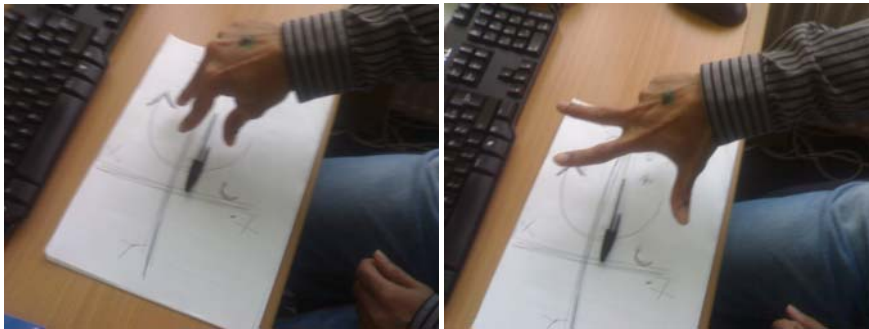


Figure 7.12 : Deselection of an object

As shown in figure 7.12 to deselect a block, the user lets it loose in a rapid action

7.5.2.3 Translation

Translation for this modality required the user to simply move his/her hand slowly and steadily in whichever direction he/she wanted. The only condition imposed by the setup was that the fingers should always be pointing downwards, for example as in figures 7.8, 7.9 and 7.10.

7.5.2.4 Rotation

The selected objects could be rotated on the three axes in either direction, while keeping the fingers closed, rotating the hand about 30-40 degrees around the desired axis and putting it back in its original position. This action maybe thought of like the action that one performs for turning the car using the steering wheel of the car, or the action of ignition of a motor car. In both cases, one rotates the hand and then rotates it back to the original position (figure 7.13).



Figure 7.13 : Rotation of a block

Figure 7.13 shows how to rotate objects in gestural modality. The hand must be turned around the desired axis and then quickly brought back to original position. In case the hand was not turned back to its original position, the selected block kept rotating on the given axis.

7.5.2.5 Zooming-in and zooming-out of the scene

Zooming-in and zooming-out was a little bit special for the gestural mode. First, holding the view-ball in one hand and then clasping the other empty hand (thus making a gesture using both hands as if one was holding an iron rod) makes a stick/bar appear. Now bringing both hands closer towards the centre of this rod (i.e. in the central grey area of the rod) makes the scene zoom-in, while if the hands go farther (white area of the rod), it makes the scene zoom-out.

It sometimes happens that the hand of the user escapes the centre of that small sphere that appears inside the highlighted block due to the collision feature. This is the case, for example, if the user continues to move a block while it comes in contact with another object like a table or another block. In such a case (i.e. when the block stays at the point of collision, while the hand of the user would have moved much farther in space), the user has to bring his/her hand back within the region of the sphere before it may again begin to follow the hand's movement. During all this time, the block continues to remain selected. If the user wants, he may deselect that block before reaching back for it, for example by unfolding the hand to deselect the object, and by then selecting another object in the scene.

7.5.3 Vocal modality

7.5.3.1 Setup specific to this modality

The Microsoft *Windows Vista's* voice recognition system was used at the base to communicate with the Virtual Puzzle system. The language used was French. The words in our vocabulary were fed to the system so that the system first tried to map what it heard onto those words at a higher priority.

An appropriate vocal command had to be used to move or to rotate the selected block, for example "*Bouger à droite*" (Move Right). The block kept moving in the said direction until the user said "STOP" or until the user gave another command to move it in some other direction, or until it collided with something.

Note: The users for this modality had to first go through a vocabulary familiarisation and voice testing phase before the actual experimentation was conducted. See Appendix A for the list of commands (in French) that the users had to read out first, in order to verify if the system recognised their voice well; and Appendix B containing the complete list of commands available to the users along with their descriptions.

7.5.3.2 Selection/deselection, translation, rotation, and zoom

Each of the seven blocks had a number associated with them, which was written over that block. The user could select that block by the French expression equivalent to “Select Block n” or “Choose block n”, where n is to be replaced by the number ranging from 1 to 7, for example “Choisir bloc un” (Choose block one (1)), or “Sélectionner Bloc cinq” (Select block five (5)).

All the actions were performed by means of passing vocal commands to the system.

7.5.4 Multimodality (G+V)

Since we had Gestural (G) and Vocal (V) as the base modalities, for multimodality, the idea was that the user would not only be able to perform the task either in one of those modalities or the other (or be able to switch between them) at will, but also be able to combine some of those interactions using both modalities in a natural way, hence the multimodality (V+G). [Oviatt2000] suggests that a well designed multimodal system "should be able to integrate complementary modalities such that the strengths of each modality are capitalised upon and used to overcome weaknesses in the other". Indeed, we decided to examine the situations where the user may be expected to perform an action more easily and/or swiftly in a multimodal context. For example, the user can place his/her hand at any desired position, and call out “attraper bloc n” (grab block n). Or, while pointing to a certain block, the user can call out “attraper ce bloc” (grab this block). In both cases, the chosen block swiftly flies into the user’s hand, saving a lot of manipulation time as well as effort.

7.5.4.1 Performing tasks in virtual puzzle with greater ease using multimodality

In real life, because of the gravitational constraint, movements of human beings are highly constrained. In various kinds of spatial tasks, for instance when delivering route directions, one may encounter situations involving the joint use of verbal and graphic information (for

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

example, while using the map information intended to help the user visualising route instructions). Or one can make use of body gestures, such as pointing the finger in a certain direction while stating "You go straight in that direction" (where the word "that" is accompanied by a pointing action, so that "that direction" corresponds to a direction shown by the finger).

Already, there are limits to our vocabulary for giving directions with the 2D-functional flat model of the world, but giving verbal instructions in a 3D space, things could get even more confusing and complicated. Moreover, for changing the direction, in the 2D-functional flat model, just the words "left" and "right" suffice, hence phrases like "You go straight in that direction, then turn the second right". Here once again when giving instructions in a 2D landscape, one can communicate quite easily and unambiguously where to turn. However, it could get rather complicated if one were to communicate with 3D maps, if it made sense to turn "up" and "down". Likewise, imagine if someone said "turn left", could that mean on the one hand, to pivot anticlockwise imagining oneself being looked at from the top; or could that mean, to tilt the whole body, by leaning the head in the left-downwards, until it touches the ground? As it may be seen, due to the way that the world behaves, and we move around the world, we do happen to have evolved a simple everyday vocabulary that is sufficient to give instructions as far as 2D maps are concerned. However, for 3D navigation, we lack simple everyday vocabulary to communicate the ideas, rapidly, easily and unambiguously, though we do happen to have only the technical vocabulary to give precise instructions for 3D navigation.

An interaction technique was thus implemented that allowed the user to move a certain object on a given axis and in a given direction by pointing towards it. For example the user could pass the following set of commands:

"Choisir bloc cinq"; (*Choose block 5*)

"Bouger vers cette direction" (*Move in this direction*)

had the following effect.

Choose block 5 (The block gets selected)

Move in this direction (while pointing in a certain direction, say, to the user's left)

(The block moves towards that direction (say, the user's left))

Similarly for rotation ("*Rotation suivant cet axe*" (*move around this axis*) would make the object move around the axis pointed to by the finger).

7.5.4.2 Performing tasks in virtual puzzle swiftly using multimodality

In comparison to our vocabulary for the V modality, choosing a certain block that is placed far from the solution, and bring it in close proximity of the solution is time consuming and may require a large set of basic manipulations from the user.

An interaction was added for the multimodal, that allowed the user to point towards an object by the laser technique as in the G modality and to give the system following command: "*choisir ce bloc*" (choose this block) or "*prendre ce bloc*" (take this block), as done so in the V modality. This would make that block rapidly approach the user's hand.

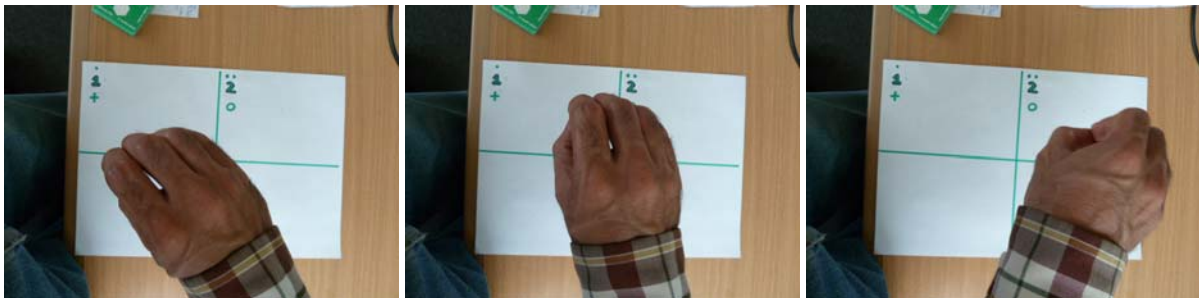
Alternately the user could say, "*Attraper bloc quatre*" (catch block four) and the block would still come flying right into the users dominant hand. Note that in this current case, there was no need to point towards the object, since the phrase ended with a number denoting the block (one, two...seven)

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

And finally, the user could select a block with the hand, and then rotate this block by passing the verbal command for example “tourner à gauche” (turn left), and the block would turn towards left in the depth perception. This could be useful since the user's hand may be rotated rather easily in certain directions, but not easily in certain others. With the help of gloves, the users could rotate the blocks in any of the three orientations by rotating their hands by 45° in that particular direction and then putting it rapidly back at its neutral position. Figures R-01 to R-03 below shows the hand in the two extreme and centre position. That is to turn left, the user would first turn the hand as in figure R-01, and then back in position R-02). Similarly in order to turn right, he/she must first turn the hand as in R-03, and then back again, as in R-02. This kind of rotation is also known as “yaw”, and for easier visualisation, this is the direction in which a boat or a ship can move.

The other three figures that immediately follow on the next line show the same positions of hand from a front view.



R-01-yaw-left-top-view

R-02-yaw-center-top-view

R-03-yaw-right-top-view

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

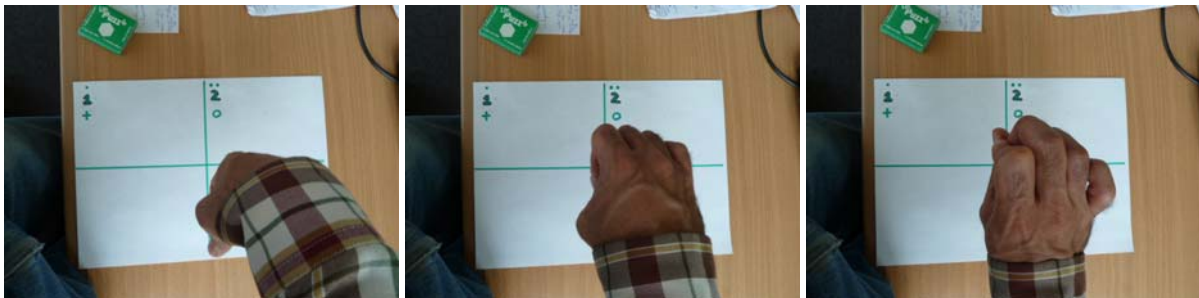


R-01-yaw-left-front-view

R-02-yaw-center-front-view

R-03-yaw-right-front-view

Similarly, there are two other ways in which the user could rotate the blocks. Let us first consider the direction in which one turns his/her hand when accelerating a motorbike (see figures R-04, R-05 and R-06 from the top and front view as well). This rotation is also known as “pitch”.



R-04-pitch-forward-top-view

R-05-pitch-center-top-view

R-06-pitch-backward-top-view



R-04-pitch-forward-front-view

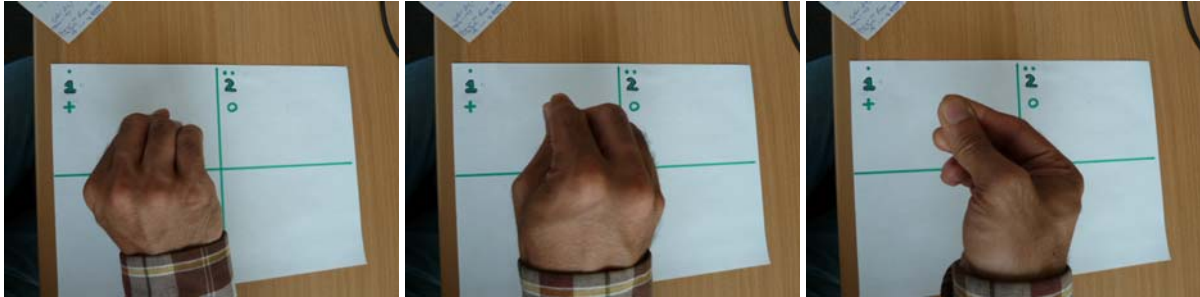
R-05-pitch-center-front-view

R-06-pitch-backward-front-view

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

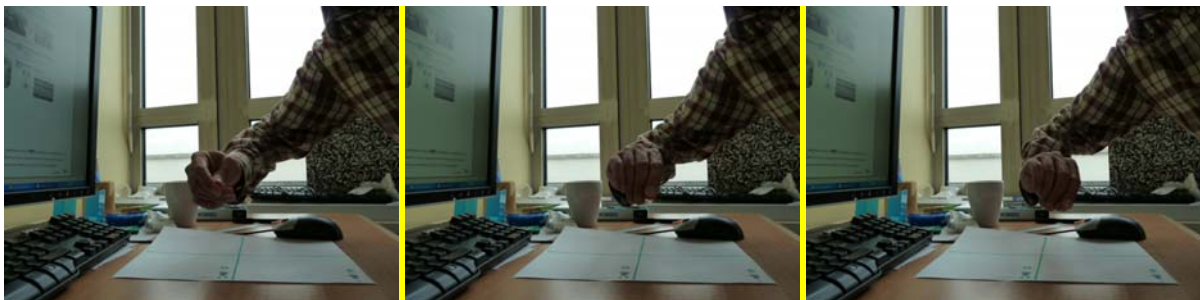
And finally, the user could rotate blocks in the direction as if one turns the key. This rotation is also known as “roll”. (see figures R-07, R-08 and R-09 from both the top and front views)



R-07-roll-left-top-view

R-08-roll-center-top-view

R-09-roll-right-top-view



R-07-roll-left-front-view

R-08-roll-center-front-view

R-09-roll-right-front-view

Note that in the last two examples of the vocal commands, while we say that the commands are executed multimodally, there is one modality in action at a given time, so for instance, in the last example while the block is selected by hand gesture, later on it is manipulated vocally. In other words, it may be argued that while the two different modalities (G+V) are used, the interaction is unimodal at any given time (or is unimodal sequentially). However, it may also be counter-argued that the interaction is in principle purely multimodal, since the user is not necessarily required to keep his/her hand static. Rather, he/she could also move the object spatially (translation), while concurrently passing the rotate command verbally. Thus it could either be sequential (G, followed by V followed by G) or concurrent (V+G).

7.5.5 Salient features of the interface

As a comparison of behaviour and interaction of objects in the virtual puzzle, here are some important differences with its real world counterpart:

- i)* One notable difference in the behaviour of the virtual puzzle blocks with respect to their behaviour in the real world is that the virtual blocks lacked gravity.
- ii)* In the virtual puzzle, only one object may be selected (or only one object will be active) at any given time. This also means that only one of them could move at a time.
- iii)* Rotation takes place in canonical form (in chunks of 90°) even when the block is placed flatly on the table. In the real world, it is possible to rotate on this axis at any angle.
- iv)* While rotation takes place in the virtual puzzle, the user may not be able to perform any other actions until it is finished, in real life, users are not only bi-manual, but also they are performing multiple actions over the same block simultaneously.
- v)* The reason for using grids was multi-fold: it allowed for the objects to be aligned perfectly and it helped with the docking of objects.

7.6 Method employed for the research

7.6.1 The overall protocol

The participants were given a brief introduction to the experiment (for example being told that the experiment involves 3D spatial-manipulation task). The participants had to first sign a paper covering the ethical aspects. The participants were informed that the block tasks during the experiment will be filmed and that they reserved the right not to sit for the experiment and the right to discontinue the experiment anytime during the experiment, if they wished to do so. Moreover, it said that the experiment will be used for the purpose of this research only, and thus will not be used for any commercial purposes.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

The experiment was conducted using three different modalities, the Keyboard/Mouse (K), Gestural (G), and Vocal (V). The overall description and protocol for the experiment remained the same, while only the modalities and the interaction techniques (which depended upon the respective modalities) changed according to the modality chosen. The overall protocol was as follows.

As a precursor to performing the “actual” task of solving the 3D virtual puzzle, the participants went through a training session of learning phase that introduced them to the different available controls, and allowed them to familiarise themselves with the interface. We shall call this the learning phase of the VE tasks (table 7.2, column c).

The learning phase could further be subdivided into four sub-tasks. For each sub-task during this learning phase, users were first shown examples by means of animation, and their task was to repeat the task that they had just seen in the animation. They had to perform four fairly straightforward sub-tasks within this learning phase, for which they had 12 minutes maximum. All those sub-tasks (like selecting an object, moving an object, rotating an object etc.) were necessary to be learned, before the user could be put to the real test of solving the real 3D Virtual Puzzle (table 7.2, column d).

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Table 7.2 : The protocol for the virtual experiment. (The columns indicate the various phases, and break-up, of the protocol for the experiment, whereas the numbers underneath the columns indicate the time in minutes)

Preliminaries		Spatial Manipulation VE Task			Paper based Tests		Feedback
		<i>Main</i>	<i>Auxiliary</i>				
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Introduction to Experiment	Personal Information	Learning Phase	The Main Task	Mini-Task	MRT	MPFB	Feedback Questionnaire
10		12	15	5	8	8	10

All care was taken that the nature of tasks during the interface familiarisation phase (c) remained basic in nature, and that while it lets the users familiarise themselves with the interface (for example associating/mapping a command or an action) and learn how to make the interactions, but no further than that. The users ought not to get any insight into the resolution of the problem that would follow. For example, in the first sub-task, the users were led to interact with simple blocks (of size 1 x 1 x 1) for the selection and translation. Figure 7.14 describes in detail the first sub-task of the learning phase that focused on translation of blocks.

There were four subtasks in all. The first sub-task focused on selection and translation of blocks, the second on rotation of blocks, the third on change in the point of view on the scene (which included zooming, de-zooming, and observing and manipulating the blocks from different viewpoints), and the fourth on docking of blocks. Figure 7.14 shows multiple snapshots of the first sub-task, while figure 7.15 gives an overall view, and shows one snapshot each of the four sub-tasks of the learning phase.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

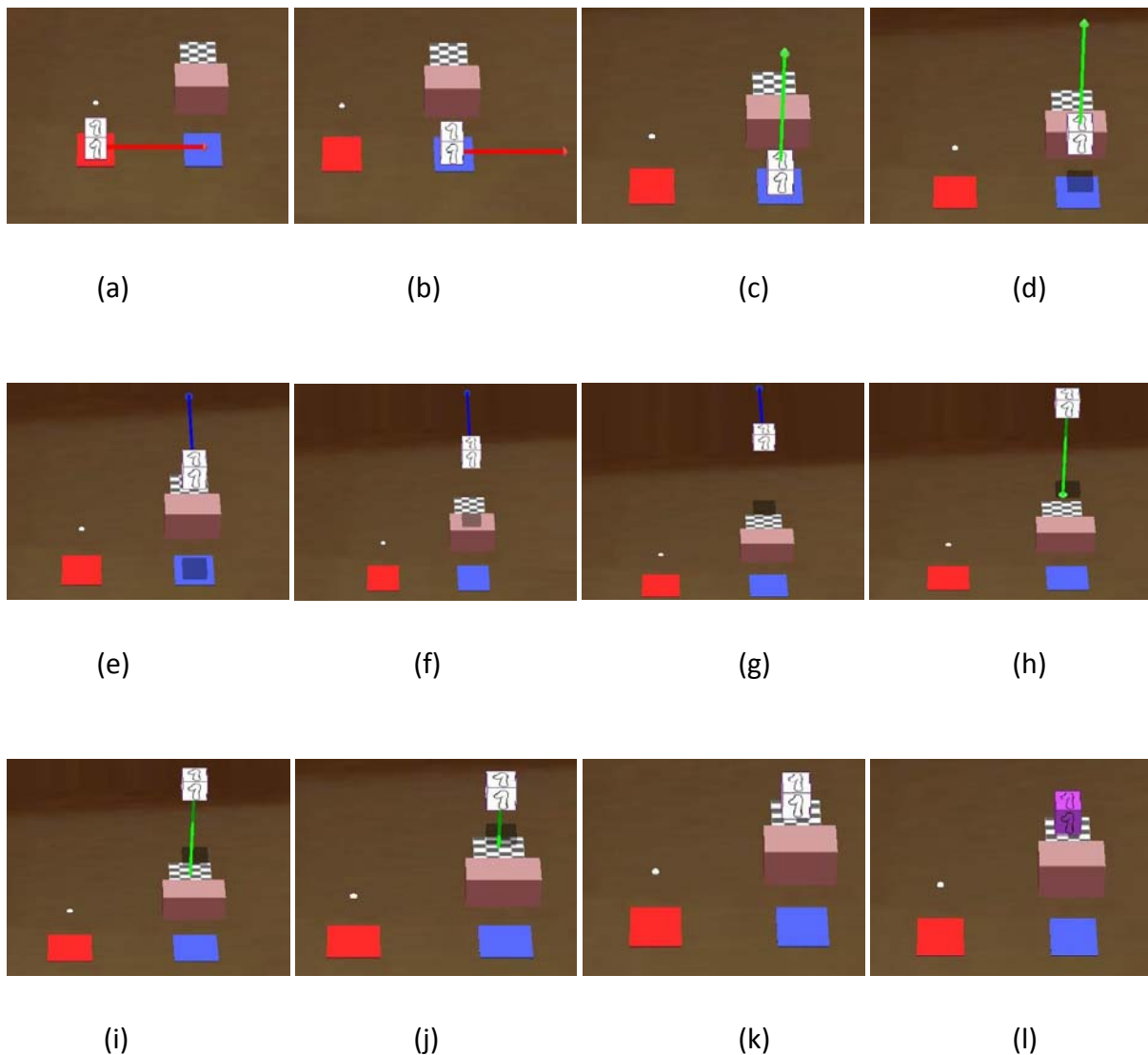


Figure 7.14 : Snapshots of the first sub-tasks during the learning phase

The objective of the first subtask of the learning phase was to teach the user how to translate objects before he/she attempts to solve the virtual puzzle experiment (figure 7.14). The user was first shown an animation and then he/she was to perform a similar action. The goal of this particular subpart included moving the small 1x1x1 block from the red square towards the blue square, and then to the chequered square. Since there was a solid brown block lying between these two squares (i.e. between the blue square and the chequered square), the user was required to avoid this big brown block as well. Here is a list of actions that the user saw.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Figure 7.14 (a and b): First, selection of block (upon selection, the block turns from purple to white). Then, movement of block towards the right. Notice the red arrow emanating from the little white block in the first two of these images, indicating the direction (axis) along which the block was moving. In these first two images, the block is moving towards the right (from position marked by the red square towards the position on the table marked by the blue square). In figure 7.14 c and d, the block moves upwards. This move is done in order to avoid the barrier (brown block) ahead. Notice the green arrow denoting the upward direction of the white block. Notice also the shadow of the little block as well, as it moves in the upward direction and away from the table.

It then moves depth-wise away from the user (figure 7.14 e, f, g). Notice the blue arrow and the shadow. And finally it moves in downward direction until it reaches the table at the chequered square (figure 7.14 h, i, j and k). Notice the downward green arrow and the shadow). Finally the block is deselected (figure 7.14 l). Notice that upon deselection, the block turns back to its original purple colour).

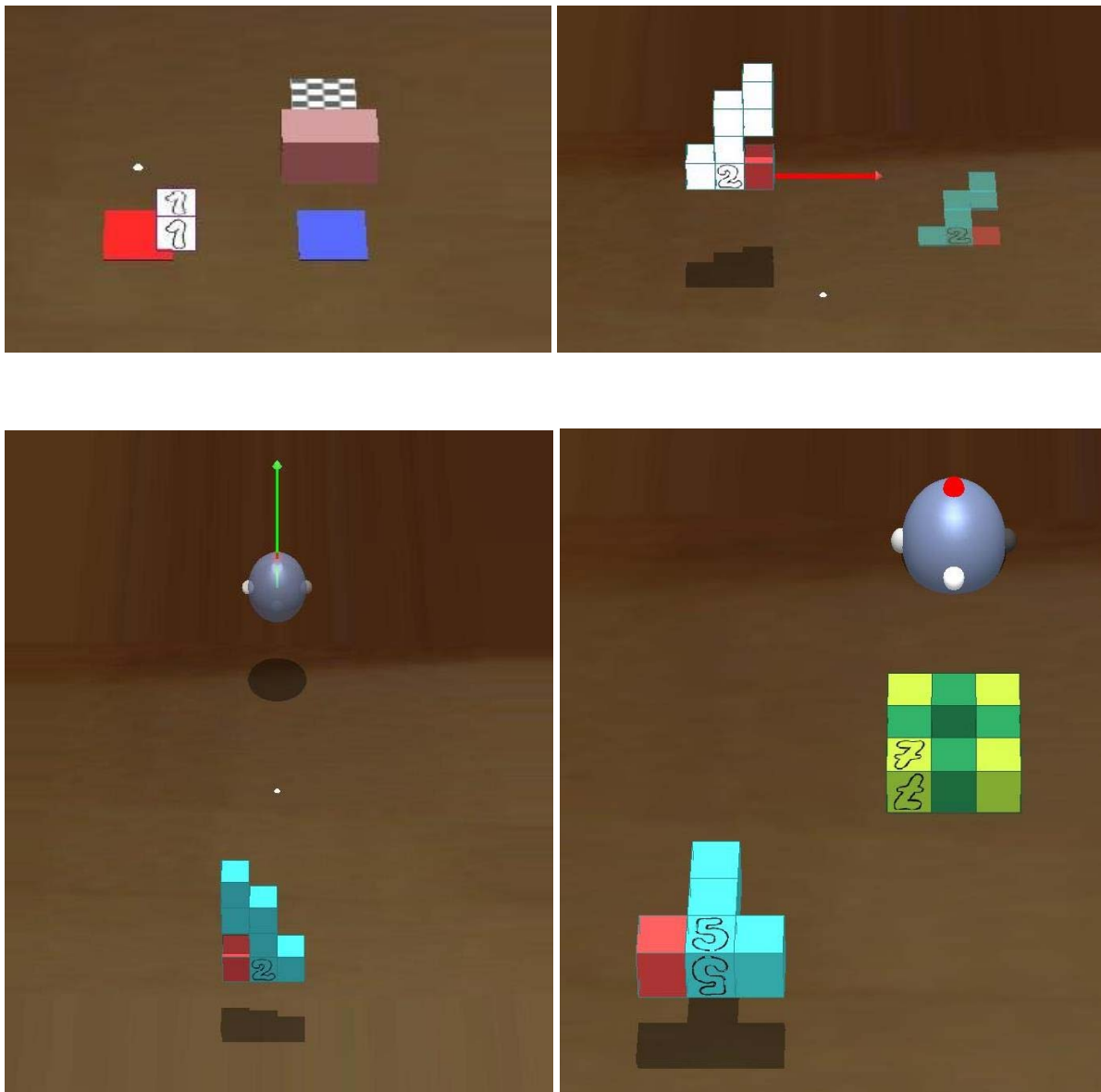


Figure 7.15 : The four different sub-tasks of the learning phase

Then, for rotation (second sub-part), the users handled just one block which was multi-coloured, so that they could visualise the rotations well on different axes. Since the users were often to encounter the task of docking in order to achieve their goal of assembling the cube during the main task (table 7.2-d), so it was thought that the task of docking should also be treated as a basic one, and ought to be learned by the user before the real task. Though yet again, during the learning phase (table 7.2-c), the docking scene contained just

two blocks, one of which was fixed (square shaped), while the other was movable (T-shaped), and the users' goal was to dock the head of the movable *T-shaped* block inside the empty square hole formed by the fixed block.

Following the learning phase, the participants sat for the real task (figure 7.10). They had a maximum of 15 minutes to solve it.

It may be apt to point out the few differences that were necessitated due to the peculiarity and nature of the given modality (for example testing if the system recognised the users' voice for V, or the preliminary hand gesture test for G). These modality-specific steps took place after the users had filled in their personal information, and just before the users embarked upon the learning phase (table 7.2, between b and c).

7.7 Results

7.7.1 Participants

A total of 40 participants took part in the various steps of the Virtual Puzzle experiment – the experiment in the virtual-settings. Twenty of them (10 females, 10 males) participated in the K group. The other participants (5 females, 15 males) were involved in the G and V groups. Originally, it was our intent to have the same participants, sit for both the gestural and the vocal conditions, since within-participants designs generally provide a better basis for the comparison of conditions. We thus started running the same people in both conditions (alternating those involved first in the G and secondly in the V conditions, and those involved in the reverse order). Four participants were run like that, but it turned out that the experimental sessions were very heavy for the participants, with sessions imposing substantial learning times. It was thus decided to move to a between-participants design, which would require every participant to learn only one modality. We then run 10 participants in the G condition and 6 other participants in the V condition. We inspected the

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

data profile of the first four participants who had been originally tested in the two conditions and checked that their performance was not discrepant at all with the performance of those participants who were tested in only one condition. The subjects did not differ significantly in terms of their MRT profile depending on whether they participated in one or two conditions (Two-sample Wilcoxon rank-sum (Mann-Whitney) test, $z=0.649$, $p=0.5166$) as well as on their number of moves (Two-sample Wilcoxon rank-sum (Mann-Whitney) test, $z=0.065$, $p=0.9484$). Consequently, rather than simply ignoring these originally tested participants, we included their data in the set of data of the respective G and V groups. As a result, the data reported below include those of 14 participants (all males) in the G group and those of 10 participants (5 females, 5 males) in the V group.

Tables 7.3, 7.4, and 7.5 show the distribution of the participants in each group in terms of gender, handedness, and age.

Table 7.3 : Gender based composition of participants

Gender ↓	Groups →	K	G	V	Total
Female		10	0	5	15
Male		10	14	5	29
Total		20	14	10	44

Table 7.4 : Dominant-hand based composition of participants

Dominant Hand ↓	Groups →	K	G	V	Total
Left		2	2	1	5
Right		18	12	9	39
Total		20	14	10	44

Table 7.5 : Age based composition of participants

Age ↓	Groups →	K	G	V	Total
Minimum		19	23	23	19
Maximum		27	31	45	45
Average		22.95	26.21	29.6	25.5

Looking at the spatial abilities of participants, we found that the three groups did not differ significantly in terms of spatial rotation skills as measured by the Mental Rotations Test (respect. medianG=14, Iqr=9; medianK=13, Iqr=9; medianV=8, Iqr=5; Kruskal-Wallis test, $\chi^2=4.402$, DoF=2, $p=0.1107$). Group V exhibited lower rotational skills than the other two groups, although non-significantly. They did not differ in terms of the MPFB as well (respect. medianG=21, Iqr=11; medianK=20, Iqr=9; medianV=15, Iqr=8; Kruskal-Wallis test, $\chi^2=1.732$, DoF=2, $p=0.4206$). Once again, we observed a (non-significant) lower performance in spatial skills test for the V group.

7.7.2 Overall performance

Table 7.6 shows the distribution of participants who succeeded and those who did not succeed solving the virtual spatial puzzle.

Table 7.6 Success rate of participants

Succeeded ↓	Groups →	K	G	V	Total
Yes		10	1	1	12
No		10	13	9	32
Total		20	14	10	44

In both G and V conditions, the success rate was extremely low (1/14 and 1/10, respectively) in comparison to the K condition (10/20). The test showed a significant effect of the conditions (Fisher test, $p=0.012$). Further tests showed that the G group performed significantly less than the K group (Fisher test, $p=0.040$, using the Bonferroni adjustment technique), while they showed no significant difference between K and V conditions (Fisher test, $p=0.111$) and between G and V conditions associated to the immersive environment (respect. 1/14 in G and 1/10 in V; Fisher test, $p = 1.000$). It is also noticeable that subjects succeeded worse in these three virtual conditions than subjects in the real setting condition reported in the previous chapter (17/24 i.e. 70%).

Across the K, G, and V conditions, subjects who succeeded were those who exhibited the significantly highest levels of visuo-spatial skills (medianMRT when Success=15, Iqr=6; medianMRT when failure=10, Iqr=8; Two-Sample Wilcoxon rank-sum test, $z=-2.334$, $p=0.0196$). Subsequent analysis replicated the same result within group K (Two-Sample Wilcoxon rank-sum test, $z = -1.969$, $p=0.0489$). No significant tests were found in G and V conditions due possibly to the very low number of subjects that succeeded in these two conditions. There was no significant difference in MPFB score across conditions between solvers and non-solvers (Two-Sample Wilcoxon rank-sum test, $z = -1.534$, $p= 0.1250$).

7.7.3 Partial success distribution for G and V

The number of solvers for the G and V modalities is very marginal. Thus, it may not be sufficient to classify the subjects only in terms of solvers (S) or non-solvers (NS). Therefore, we devised a method to count the number of blocks adjusted within the 3x3x3 *workspace* (see chapter 5 for description of workspace):

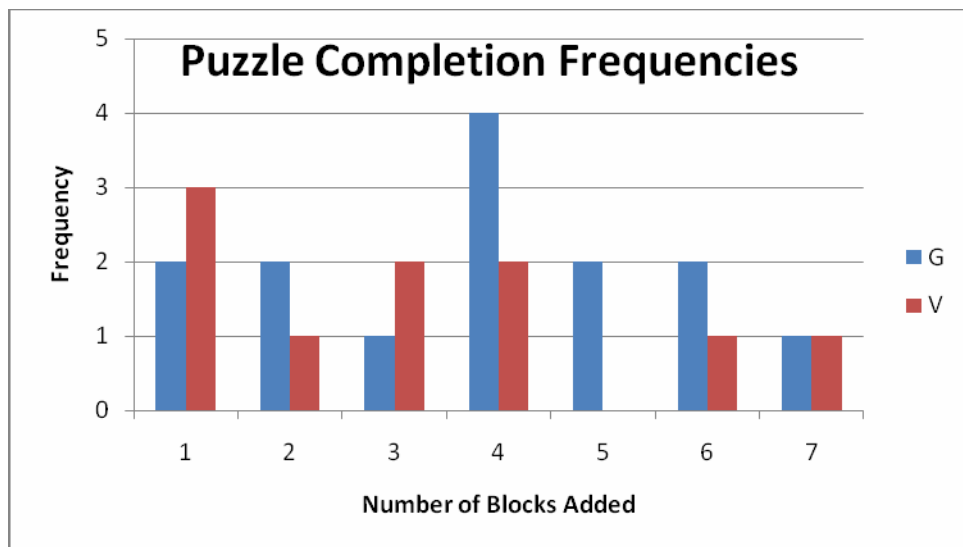


Figure 7.16 : Puzzle completion-level frequencies

In figure 7.16 above, the X axis shows the number of blocks completed towards the solution in the 3x3x3 workspace. The Y axis contains the number of participants who reached this level. Those who had not put together any blocks together at the end of their stipulated time appear against the value 1 (2 participants from G, and 3 from V). As to those who completed the cube (i.e. were able to place all the seven blocks within this workspace), their frequency is shown under the value 7 (1 for G and 1 for V).

If success is put at 5 blocks at least (higher end of the spectrum), then 5 participants in G and only 2 participants in V succeeded. If success is put at 4 blocks (mid to high), then 9 participants in G, and only 4 participants in V succeeded. On the basis of the above observations, we conclude that participants tended to perform better in G than in V, although the test failed to be significant (medianG=4, Iqr=3; MedianV=3, Iqr=3; Two-Sample Wilcoxon rank-sum test, $z=0.849$, $p=0.3961$).

7.7.4 Mean times for solution

Average time taken by S subjects in group K was the fastest with 622 seconds, while it was 875 for one unique solver in group G and 782 seconds for one unique solver in group V. If one compares the difference between the mean of K with the other two groups, the difference was more than 3 minutes with V, while there was more than 4.5 minutes of time difference when compared with G. Although the number of subjects who succeeded in G and V groups was low, the observed difference gave rise to a significant test (Kruskal-Wallis test, $\chi^2=11.00$, DoF=2, $p=0.0041$).

Table 7.7 : Average times for solving the puzzle

Ave time taken ↓	Groups →	K	G	V
Minimum		211	875	782
Maximum		900	875	782
Average		622	875	782

7.7.5 Comparison of moves in different modalities

7.7.5.1 Number of moves

Table 7.8 shows the break-up of the average number of “moves” made by participants for each modality.

Table 7.8 : Average number of moves made for each modality

(a) Moves ↓	Groups →	(b) K	(c) G	(d) V
Minimum		119	11	9
Maximum		690	31	26
Average		376.70	20.00	15.30

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

As shown in table 7.8, the average number of moves for K is 376, whereas for the G and V groups, it is 20.00 and 15.30, respectively. The result is similar if we consider separately subjects who succeeded and those who failed in completing the task. The former exhibited a median number of moves about 274 (Iqr=169) in group K against 22 in group G and 11 in group V (one subject only for both). The latter showed a significantly lower number of moves in the G and V conditions (respect. medianG=20, Iqr=9; medianV=16, Iqr=4) than in the K condition (medianK=439, Iqr=203; Kruskal-Wallis test, $\chi^2=20.676$, DoF=2, $p=0.0001$).

Since the number of attempts differed somehow depending on the conditions, we also computed the average number of moves per attempts. The data (table 7.9) show that subjects in the K condition exhibit a high number of moves per attempt whereas very few moves per attempt are observed in both G and V conditions. The V condition shows the lowest level with only 13 moves on average per attempt. The test is significant (Kruskal-Wallis test, $\chi^2=33.070$, DoF=2, $p=0.0001$).

Table 7.9 : Median, Inter Quartile Range (Iqr) for number of moves per attempt as a function of condition

Moves ↓	Groups →	K	G	V
Median		133	20.00	13
Iqr		71	8	5

The large differences observed between the K conditions and the other two immersive conditions have two possible explanations. Firstly, since the users are accustomed to and have traditionally had a lot of experience with the K modality before the experiment, they executed their commands rather rapidly. The second explanation is that the difference perhaps lies inherently in the selection /manipulation device and the respective interaction techniques. Furthermore, since users could select the blocks using a mouse, thus perhaps mouse has an inherent quality whereby blocks sometimes got selected by the users

inadvertently. If that is indeed so the case, then the user must once again select the correct block and may have three moves counted instead of one. But even considering such corrections, it remains that the manipulation in the K condition offers more opportunities of executing moves within the same time range.

7.7.5.2 Break-up of moves

Observations of the moves for the experiment in virtual settings revealed that they were quite similar, yet with some noticeable differences. Most of the moves for the virtual puzzle were similar to that of the ones observed during BMT (such as “Add”, “Remove”, “Rearrange”, and “Rotate”). However, certain other moves were observed during the BMT and were not really observed in the virtual puzzle (such as “Pick/Hold”, “Remove/Hold” and “Pick/Leave”) because these moves did not really make any sense under the virtual setting.

Let us consider, for example, the move “Remove/Hold” in the real setting. It is described as “The subject picks up a block up from the solution and holds onto it, without placing it either on the table or back on to the solution” (refer to table 6.1). But in the real setting, when the subject removed a block, it was a different thing if a) after removing the block, he/she held it in hand and contemplated what to do with it (in which case, it was counted as a “Remove/Hold” move, the word *hold* here signifying that the user chose to retain it in his/her hand); or b) after removing the block, he/she just placed it immediately on the table and moved his/her focus on another object (in which case, it was counted as simply a “Remove” move). This lack of difference between the two different moves was also due to the fact that in the real world, it was not possible to leave a block in mid-air. It either had to be held in the hand, or be put down on the table. Leaving the block in mid-air would mean that it would fall from the table, which was something that the users didn’t do in practice, whereas in the virtual setting, due to the absence of gravity, leaving a block in mid-air or placing it on the table didn’t make any difference. Thus, when the user wished to take off a

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

block, it was just taken away from the solution, and just deselected somewhere far (i.e. anywhere as long as it was a bit far from the solution).

In the same manner, there was one move that was observed which was exclusive to the virtual setting, “Displace”, and which was not observed during the BMT. As per this move, “the user moved a block from one place from the table to another place on the table”. This happened quite often during the course of the virtual puzzle as was revealed by the analysis. Sometimes, it was just due to the simple fact that while the user was moving a block from one place to another, it simply got deselected (e.g. in the case of the G modality) and the user simply had to once again select and move the object in the originally thought direction. At other times, it happened due to the longer execution times for moves in the virtual setting. That is, while the user was moving the chosen object from one place to another, he/she changed his/her mind during the course of that action, and simply changed his/her mind and moved (the focus) onto another object. This happened at times, if he/she didn’t succeed at reselecting the same block that had gotten deselected; while in other cases, the time to transport an object from one extreme corner of the manipulable area in the virtual puzzle to the other was so long that the subject sometimes decided to simply change his/her mind. Such “Displace” moves didn’t happen in the real setting because the execution time of a move was very rapid.

We did a further analysis of data regarding the type of moves in the immersive situation (table 7.10). Beyond the fact that “Displace” was logically the most observed action due to the virtuality of the environment, we can observe a weak but significant relationship between S's or NS's pattern of favoured moves (Cramer’s $V^2=0.033$; Fisher test, $p=0.007$). Analysing the Relative Deviations with an absolute value greater than 0.20, we observed that S subjects are characterised by comparatively less “Rearrange” ($RD=-1.00$) and less “Displace” ($RD=-0.40$) moves while they exhibit far more “Add” ($RD=+0.90$) and “Remove” ($RD=+0.65$) moves. This group of subjects favoured adding and removing pieces over

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

rearranging and displacing them. Inversely, no clear difference between the types of moves appeared among NS subjects. These subjects are characterised by a large number of “Displace” and “Rearrange” moves. “Rotate” was the only move that appeared to be independent from the fact of being a solver or not.

Table 7.10 : Frequency of the types of moves for subjects in immersive (G or V) conditions as a function of success or failure to complete the task

	Add	Remove	Rearrange	Displace	Rotate
Solvers	15	2	0	7	6
Non-Solvers	89	14	44	146	72
Total	104	16	44	153	78

We further decomposed our analysis by separating S and NS subjects in the G and V conditions in terms of an optional effect on the type of moves they favour. Among S (table 7.11), we observed a strong relationship between the condition in which S subjects carried out the task (G vs. V) and the moves they have favoured (Cramer’s $V^2=0.17$). The test failed, however, to reveal a significant effect (Fisher test, $p=0.196$) which could be due to the small number of S subjects in our sample. Relative Deviations suggest that both S groups do not differ in terms of the number of “Add” moves. Similarly, both of them did not perform any “Rearrange” moves. S subjects in group G were characterised by the fact that they favour the “Rotate” moves ($RD=-1.00$) avoided both “Remove” ($RD=-1.00$) and “Displace” ($RD=-.61$) moves. Contrastively, S subjects in the V group showed a trend to avoid the “Rotate” moves ($RD=-0.47$) whereas they favoured “Remove” (+0.58) as well as “Displace” (+0.35) moves.

Table 7.11 : Frequency of the types of moves for solvers in G and V conditions

	Add	Remove	Rearrange	Displace	Rotate
SolversG	9	2	0	6	2
SolversV	6	0	0	1	4
Total	15	2	0	7	6

Among NS (table 7.12), we found a weak relationships between the conditions (G vs. V) and the distribution of moves (Cramer's $V^2=0.024$). NS in the G condition favoured mostly "Rotate" (RD=+0.20), whereas NS in group V were characterised by doing more frequently "Rearrange" (RD=+0.22) and avoiding "Rotate" (RD=-0.33) and "Remove" (RD=-0.23). The test failed, however, to be significant (Fisher test, $p=0.065$).

Table 7.12 : Frequency of the types of moves for non-solvers in G and V conditions

	Add	Remove	Rearrange	Displace	Rotate
Non-Solvers G	58	10	24	83	54
Non-Solvers V	31	4	20	63	18
Total	89	14	44	146	72

7.7.6 Solution as a function of number of attempts

An attempt (*for definition, see chapter 6 as well*) is sort of a container for moves. An attempt may contain any number of moves and it carries on until the user takes apart all but one block from the cube.

Table 7.13 : Average number of attempts

Ave number of attempts ↓	Groups →	K	G	V
		Minimum	1	1
Maximum		5	1	2
Average		2.8	1	1.1

7.7.6.1 Number of attempts for K, G and V

Participants in the K modality performed an average of 2.8 attempts (ranging from a minimum of 1 to a maximum of 5).

With an exception of one subject who made two attempts towards the solution, all the other participants in the G and V groups made just one attempt during the entire experiment, i.e. all the subjects in the G group made just one attempt towards the solution, while the participants in the V group made an average of 1.1 attempts. Tests showed that subjects who completed the task did not exhibit a significant difference in terms of number of attempts whatever the condition (medianK=4, Iqr=2; medianG=1; medianV=1; Kruskal-Wallis test, $\chi^2=3.224$, DoF=2, $p=0.1994$). Inversely, subjects who failed in completing the task demonstrated a higher number of attempts in group K than in others (medianK=2, Iqr=1; median G=1; medianV=1; Kruskal-Wallis test, $\chi^2=19.240$, DoF=2, $p=0.0001$). Overall, these results reflect that when placed in the K condition, subjects did more attempts than in group G or in group V.

Furthermore, it is interesting to note that the average number of attempts is about 5 times less than the average number of attempts made in the experiment conducted in the real world (whose range too started from 1 but went up to as high as 15).

7.7.6.2 Number of moves per attempt

In each group, we calculated the average number of moves that the participants made per attempt. A higher number of moves per attempt may be interpreted as a reflection of user's "patience" and persistence in finding out the solution. We notice that for the groups K and G, this rate is quite similar (slightly higher) for S as compared to NS, while for the V group, this ratio is higher for NS than for S (see table 7.14)

Table 7.14 : Group-wise break-up of number of moves per attempt

Succeeded ↓	Groups →	Number of moves per attempt		
		K	G	V
Yes		138	22	11
No		132	20	14
Total		135	20	14

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

We compared the number of moves per attempt in S and NS subjects of the K condition (medianSolvers=127, Iqr=85; medianNon-Solvers=141, Iqr=57). The test for the difference failed to be significant (Two-Sample Wilcoxon rank-sum test, $z=0.378$, $p=0.7055$). The same comparison was not done for G and V because only one subject succeeded in each of these conditions. Subsequently, data show that the number of moves per attempt is mostly affected by the difference between a usual keyboard-based condition and the fact of being in an immersive condition, while no significant difference has been found between S and NS.

To account for these findings, we must consider that the manipulation of objects is more direct in the case of K and G as compared to V. The notion of "directness" of manipulation may be interpreted in two senses.

Firstly, the notion corresponds to the *mapping of the users' actions and reactions* on objects that are spatially close-by. So, as in the case of G, the user manipulated the blocks by actually making a grabbing action over the blocks and moving them in 3D space, which corresponded with the actual movement of the blocks in virtual space.

Secondly, the notion corresponds to the *level of reactivity of the system* and how it makes the user believe that he/she was in complete control over the blocks' manipulation. This happened in the case of K, where, even though the user had an indirect manipulation according to the first sense, the user could quickly adapt and create a link between his/her actions and the block's response. In the case of K, there were two reasons why the user developed this mapping so quickly: a) the user's previous experience with the device; and b) the reactivity and reliability of the system.

We nevertheless must be cautious with these explanations, since the number of S subjects for V was extremely small (i.e., one), so it could simply be the case that the one who did solve it just got lucky or had a natural acumen for utilising vocal commands.

7.7.6.3 Number of moves per minute

The success rate for a given modality seems to have a direct link to the ease with which the user is able to manipulate the objects in desired manner.

Overall, the average number of moves per minute for the K group was 30.52, to be compared with the average number of moves per minute for the G and the V groups (1.33 and 1.03, respectively). Tests show a significant effect of the condition on the average number of moves per minute for subjects who did not reach solution to the task (MedianK=32.42, Iqr=11.27; medianG=1.68, Iqr=0.77; medianV=1.14, Iqr=0.36; Kruskal-Wallis test, $\chi^2=21.598$, DoF=2, $p=0.0001$), whereas the test failed to be significant for subjects who succeeded (medianK=18.27, Iqr=11.27; medianG=3.19; medianV=0.74; Kruskal-Wallis test, $\chi^2=4.654$, DoF=2, $p=0.0976$). Furthermore, subjects in group K who succeeded in completing the task exhibited a significantly lower number of moves per minute than those of subjects in group K who failed in completing the task (Two-Sample Wilcoxon rank-sum test, $z=2.343$, $p=0.0191$).

The difference in manipulation speed could be explained by the fact that subjects had primarily a greater familiarity with the device and thus greater ease of use. It could be also due to the inherent characteristic of the device, which is prone to unintended selection and/or deselection.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

This was informally confirmed by the users, who found the experiment in the real settings rather entertaining (gave comments like, “fun”, “playful”, “amusing”, “funny”, etc.) as compared to the virtual puzzle which they found “quite tiring” and “exhausting”.

Even the users who were at relative ease at manipulating the objects comparatively well were not in any way even close to the amount of general ease with which one manipulates the objects in real life.

7.7.6.4 Order of blocks

Table 7.15 shows which blocks (simple or complex) were placed at each step of the construction, for the two solvers of G and V.

Table 7.15 : Number of times where complex blocks (A) and simple blocks (B) were used at each position in the (canonical) solution, for the participants who succeeded in G and V

Positioning	Complex blocks (A)	Simple blocks (B)
1st	1	1
2nd	2	0
3rd	0	2
4th	0	2
5th	1	1
6th	0	2
7th (last)	0	2
Total	4	10

The frequency with which the complex blocks (see chapter 5, section 5.2.4) were used during the first two positioning actions was 3 (that is, 1 + 2) out of 4 (75%), whereas for the last five positioning actions, the complex blocks were only once (0 + 0 + 1 + 0 + 0) out of 4 (25%). This reflects that for the virtual setup also, in line with our expectations, for S it is a more spontaneous procedure to start problem solving by using the larger and more complex blocks, to which smaller and simpler blocks are added later on. That is, as shown in table

7.15, for the construction of the cube, the S subjects placed at least one of the two complex blocks during the first three steps (canonical formation) during their successful attempt. A more formal analysis shows a strong relationships between the level of complexity of the blocks and the position they were used in (Cramer's $V^2=0.65$). Relative Deviations show that among the two solvers in G and V, complex blocks were attracted towards the first (RD=0.75), second (RD=2.50) and fifth (RD=0.75) positions whereas simple blocks were attracted to the third, fourth, sixth and seventh positions (all RTD=0.40).

Just as was the case in the BMT, in the current experiment too, not even a single participant solved the puzzle task by manipulating the complex blocks at the end (although, as we noted above in chapter 6, such solutions exist).

7.8 Comparisons and discussion

The comparison of the data collected in the V, G and K conditions as well as in the real setting condition led us to the following observations.

As shown in figure 7.17, there was a highest success rate (percentage of solvers) found in the real setting as compared to the three virtual conditions. In the virtual settings, the success rate for K (50%) was better than either G or V (which were both under or equal to 10%).

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

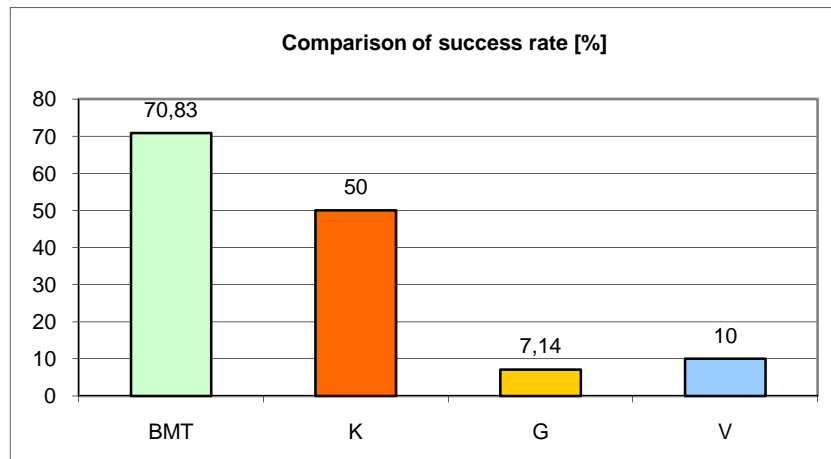


Figure 7.17 : Comparison of success rates in percentage for different modalities

Because the problem space was similar in the four different conditions, observed differences in performance related to solving could be explained by individual differences as well as by the specific properties of the conditions and systems used to interact with, or by interactions between individual and environmental features. Regarding the first line of explanation, groups of subjects in each conditions were not significantly different in terms of visuo-spatial skills: the lowest median score on MRT was 8 for subjects in the V condition, followed by 10 for subjects in the real setting condition, 13 for subjects the K condition and 14 for the G condition. The average score of subjects in the real setting condition is thus even lower than those of subjects in the K and G conditions. For the MPFB, we found an average score about 19 for the group in real setting condition against 20 for the group in the K condition and 21 for subjects in the G condition. Subjects in the V conditions were again characterised by a lower (but non-significant) score of 15 at MPFB.

An effect of spatial skills was however observed on the overall performance of subjects in all conditions. In the real setting, solvers exhibited a significantly higher score on MRT than non-solvers, but an only marginally significant higher score on MPFB. The same pattern of results was replicated in the virtual conditions where we found a significant difference between solvers and non-solvers for the score on MRT, but no significant difference for the

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

score on MPFB. We infer from these results that the four conditions shared the same component of visuo-spatial reasoning, although performance was clearly also affected by the specific properties of interface and interaction.

Having made clear that individual characteristics in groups are not responsible for the difference between the real setting and the three virtual conditions, let us examine more closely the effects of specific properties related to the four conditions and their possible interaction with individual characteristics. On average, it seems that subjects who solved the puzzle took more time in virtual conditions than in real with a median time between 622 to 875 s for virtual conditions against 200 s in real environment. This is consistent with several studies that have reported that the time to complete an assembly task in virtual environment is much longer (multiplied by two or even more) compared to the same task in real environment [Boud2000] [Unger2001]. Observed time to solve can be related to at least two different sources. First, subjects may need more time because they require more moves to achieve the task of completing the puzzle, which can be due to the fact that a similar goal requires more or less basic commands/moves to be achieved, depending on the condition. Interestingly, we can show that this is probably valid at least to explain within-group differences. For example, we observed in the real setting condition that the fastest solvers exhibited a median of 20 moves to solve the puzzle whereas solvers who took more than 200 s showed a median of 88 moves to achieve the task. Second, this may be because the same move – in terms of goal – requires more time in one condition than in another.

There was a difference in terms of speed and grain-level of moves related to the manipulation of the pieces of the puzzle in the four conditions, as shown by an indicator like the rate of moves per minute. We found a rate of 13 and 17 moves per minute (respect. Solvers and Non-solvers) in the real settings. The rates were considerably slower in the G condition: 0.74 and 1.14 moves per minute (respect. Solvers and Non-solvers), and in the V condition: 3.19 and 1.68 moves per minute (respect. Solvers and Non-solvers). Contrastively,

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

these rates are clearly lower than the rates of 18.27 and 32.42 moves per minutes (respect. Solvers and Non-solvers) in the K condition. This latter result could be interpreted as an action parcelling effect associated to the K interface: moves are not directly translated into an expected goal position by subjects and they imply fine grain translations using the more “atomic” command language implemented into the keyboard interface.

Looking at the number of moves per attempt shows the same pattern as observed in terms of moves per minute rate. We found on average that subjects did 13 moves per attempt in the real setting condition as well as in the V condition. They did slightly more moves per attempt in the G condition (median= 20). Oppositely, we observed a high number of moves per attempt in the K condition (median=133), which could be explained at least partially by the parcelling constraint. Interestingly, such a parcelling constraint was not associated to poorest performance in K as compared to G and V situations. This could be explained by the familiarity of subjects with keyboard-based interfaces as opposed to the novelty of Vocal and Gestural interfaces. Although the manipulation was not direct and required some mapping of the keys with the spatial directions, and even though there were many of them, nine (6+1+2), to learn, still they have had a lot of practice, which played a role in quick adaptation.

Putting together results in terms of rate and number of attempts (the latter being decreased dramatically from real to virtual settings), we are led to suggest that the low score in G and V conditions in terms of task completion may be also due to the fact that not enough time was provided to subjects regarding the specific temporal cost of using the G and V interfaces. An informal analysis of the videos shows that some users were struggling with a certain manipulation taking sometimes tens of seconds, and occasionally even minutes. For example, one female participant took more than three minutes just to perform an Add operation (adding a second block onto the solution).

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

1. Finally, a property of the real settings is that the manipulation was direct, and automatic: Direct manipulation as well as learning curve and mapping were minimal in real settings for the participants, and thus their focus was entirely on the strategy.
2. In the G and V modalities, most of the users, especially the ones not associated with the VR/VE domain did not have experience of using gestural or vocal systems. Even though we checked that learning of the G and V procedures was completed in the two groups, it is likely that part of the cognitive activity of the participants was occupied by retrieving the correct commands, which has added a cost. Nevertheless, it seems that both G and V groups showed a similar behaviour in terms of number of moves per attempt, but at a slower rate. If we look closely to subjects who solved the puzzle, it is also interesting to look at the respective patterns of moves found in the different conditions.

We have shown that the nature of moves differed between S and NS in both real and virtual settings, at least in the G and V groups (table 7.16). As already stated, some of the moves were similar across conditions (like “Add” or “Remove” while other were specific to one conditions (like displace specific to virtual conditions, or “Pick/Leave” specific to real setting conditions).

Table 7.16 : Frequency of moves observed in the real and virtual (G and V) conditions

Moves	Real setting	Virtual conditions
Add	841	104
Remove	432	16
Rearrange	139	44
Displace	0	153
Rotate	30	78
Try	124	0
Pick/Leave	172	0
Try/Hold	40	0
Pick/Hold	44	0
Remove/Hold	60	0
Leave	32	0
Disassemble (all)	32	0

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

What the three conditions have in common is that “Add” is one of the most frequent moves whatever the condition, which makes sense since the task was to add pieces one after each to obtain the solution to the puzzle. In real settings, “Add” is the more frequent whereas it is the second most frequent in G and V virtual conditions. Real and virtual conditions differ however globally, which can be related to the specific properties of the different conditions. Whereas “Add” is the most frequent in real setting, “Displace” is the most frequent in virtual conditions. This move corresponds to interface-based commands associated to changing place and positions of a block within the workspace. It could have also an origin in usability problems like subjects unexpectedly selecting and moving a block while attempting to move another one into the solution. This dominance of “Displace” over “Add” may generate a cognitive cost for subjects in the virtual conditions, providing a good candidate to explain the longer times required to find a correct solution. “Remove” was the second most frequent move in real setting, whereas under virtual conditions, it was only in the fifth position in terms of frequency, after both “Rotate” and “Rearrange” moves. The global differences are reflected by the relative deviations calculated on the basis of the reported table of categories of moves. In virtual conditions, we observed less “Add” (RD=-0.35) and “Remove” (RD=-0.79) moves and more “Rearrange” (RD=+0.42), “Rotate” (RD=+3.28) and “Displace” (RD=+4.93) moves proportionally as in real environment.

Finally, we interestingly observed some common patterns shared between the real and some of the virtual conditions. Solvers both in the real and in the K conditions appeared to make a smaller number of moves per minute than non-solvers. Since the number of solvers for the G and V conditions was small, it did not allow us to draw any sound conclusion, although they show consistent tendencies. We have interpreted this pattern in terms of a reasoning-oriented strategy as opposed to a more trail-and-error oriented approach.

We also observed in real, G and V conditions the same trend related to the effect of blocks complexity on solution path. Indeed, analysing the order in which solvers are assembling the

pieces to achieve the puzzle, we show that the more complex blocks are favoured as the starting basis of the solving path, whereas simpler blocks are mostly used to follow up.

7.9 The multimodal condition (with both G and V)

In addition to the main experiments reported above, we conducted a separate exploratory investigation of a condition of a group of participants, who were tested in a multimodal condition where they were invited to use simultaneously the gestural and the vocal modalities. Due to some technical reasons (system latency and stability issues), this investigation could not be completed with all the participants. Since the data size for this multimodal group (G+V) was restricted (4 sets of data), one cannot draw any definitive conclusions on this basis. However, the observations may provide us with some valuable indications.

We observed a total of 3 participants (2M, 1F). One of the male participants sat for the same experiment twice. We thus have a total of four sets of data. The participant who sat for the same experiment twice succeeded both times, while the other two participants didn't succeed in solving the puzzle.

7.9.1 Success rate

Table 7.17 : Success rate of multimodal (G+V) participants

	Total
S	2
NS	2
All	4

The success rate for the G+V group was 50%. Incidentally, it was the same subject who sat for the same task twice. There are two things which could potentially have impacted the performance of this user during his second trial: one was the greater familiarity with the interface and the second was the impact on performance due to the knowledge of the problem at hand. As it turned out, his performance was similar in both cases. In fact, it took him more time to solve the puzzle the second time as it did to solve it the first time. One explanation is more psychological than performance-based. The subject presumably had an advantage, both in terms of greater familiarity with the interface with the problem at hand. Although he could have just repeated the same moves to reconstruct the cube in the same manner, however, he tried to find a different solution to the problem without being asked to do so. Even though he succeeded, it took him marginally (37 seconds) longer to solve the puzzle, during the second sitting.

7.9.2 Average time for solving the puzzle

The average time for solving the puzzle was 370 seconds. This is remarkably less than for any of the other groups (K, G or V).

7.9.3 Average number of moves

The average number of moves for G+V is 13.25. This was expected, as it is much closer to the groups G or V than it is to K.

Table 7.18 : Number of moves by S and NS multimodal (G+V) participants

Succeeded ↓	Moves →	Minimum	Maximum	Average
Yes		7	8	7.5
No		16	22	19
Total		7	22	13.25

7.9.4 Number of moves per attempt

The moves per attempt were found to be much less for S (7.5) than for NS (19).

Table 7.19 : Moves per attempt by S and NS multimodal (G+V) participants

	Moves per attempt
S	7.5
NS	19
All	13.25

7.9.5 Number of attempts

The number of attempts for the multimodal group (1) is very similar to that of groups G and V, and dissimilar to that of group K, which was quite expected.

7.10 Conclusions

Although the best performers solved the puzzle in smaller number of moves, however, it was observed that the success rate was highest in the modality where the average number of moves and attempts was high, which reflected the ease with which the user could interact with the system.

The reason behind the greater success in the multimodal group (50%) could be explained by the fact that different users have different tendencies and preferences and thus when given both choices, they could opt for either purely one interaction technique or the other as per their individual inclinations.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Moreover, if one treats every move in isolation, then some users may have a preference for a certain interaction using the G modality, and another interaction using the V modality, which is possible in the case of multimodality or if, for some reason, a user is unable to perform a certain move. This happens when a user cannot place a block very precisely using the G modality, then he/she has the option of trying that interaction using the V modality as well.

Finally, the multimodal group allowed the subjects for certain interactions which were unique to this group (i.e. in the multimodal group, it was allowed to make all the interactions possible in the G group, as well as all the interactions allowed in the V group, as well as certain interactions which could neither be performed in either the G or the V group, for example the ones which required a combination of the two modalities. These interactions were found to be more usable than the counterpart moves of the single modalities.

Our exploratory results about the multimodal solution suggest that when operational, it could provide a better environment to support the puzzle solving (with 1/3 succeeded) than G only and V only conditions. The various indicators suggest a proximity with both G and V conditions. However, the limited size of the sample does not enable to draw any general conclusions and further investigations are still needed.

8 Chapter 8: Discussion, Conclusions, and Perspectives

This is the concluding chapter. It gives the quick overview of our work, and tries to draw general conclusions followed by our suggestions and recommendations. We provide distinct sets of conclusions regarding the contribution of our study to ergonomics and to informatics and human-computer interactions. In the end, we discuss applications for our work and future possibilities.

8.1 Overview

Part I of this thesis was devoted to the theoretical aspects. After some introductory discussions and the motivations behind our work, this thesis described what problems are, and their different categories. It then discussed the puzzles, and proposed a categorisation of a specific kind of puzzles – the spatial puzzles.

Part II of this thesis discussed the experimental aspects of our selected experiment. This second part defined in some detail our puzzle and the two experiments that were conducted in real and virtual environments.

One of the objectives of this empirical contribution was to study and identify some of the cognitive processes and factors that influence them in terms of performance and strategy associated to the task of solving spatial puzzles by humans. We used a 3D spatial puzzle as an object manipulation task. This task has been chosen because it provides a simpler but somehow representative case of solving an assembly task. Such a task has been shown as being still difficult to operate and assist in virtual reality up to now.

A second and complementary objective was to investigate the issue of implementing multimodal assistance based on analysing the task as it is approached “naturally” in the real environment. In other terms, the initial idea was to look deeper in the activity developed by people to solve the problem in a real environment, in order to extract constraints on the commands to be implemented in the virtual interface. The question is thus how the knowledge acquired in the experiment in the real settings could be used in developing the Virtual Puzzle, and how the interface for the Virtual Puzzle could be designed making use of not only the insights gained but also mixing it with the currently known virtual environment interaction techniques.

8.2 Contribution to ergonomics knowledge

The results reported in empirical part of this thesis provide us the following information on the contributions of our work to ergonomics knowledge about assembly tasks, as well as about the way virtual environments might support them or not.

People who performed the experiment in the real settings exhibited faster performance and higher success rate than in virtual environments settings including the K, G and V modalities. In these conditions, the rate of success was significantly lower. The rate of moves, defined as the number of moves per minute performed by the user for a given condition, was considerably higher in the K modality than in the G and V modalities.

Several explanations to the pattern of results can be suggested from our work. First, usability problems explain part of such results. Indeed, for the G and V modalities, the users struggled to move each and every block from one point to the other. Sometime, they spent a lot of time to perform even the simplest tasks of selecting and then moving a block from one place to another. Even when the users knew exactly what they had intended to do, their execution took a lot of time and effort. An additional factor that distinguishes between the real and the virtual environments was that the interaction in all the virtual conditions did not use two-handed manipulations. This provided subjects with the constraint of operating their actions one after the other while they attempted to solve the puzzle. Furthermore, the different conditions had not systematically the same level of granularity, which may have influenced some aspects of the control on their activity. In particular, we have evoked a possible phenomenon of atomisation of actions that can require more time and effort from the subjects, in order to achieve the same objective.

Another factor, possibly underestimated until now, is the familiarity of subjects with the activity and the technology involved in the different conditions. However, we should stress

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

that all subjects had training before starting the experimental trials. Furthermore, only G and V technologies were actually innovative for subjects, whereas the K condition made use of a widely known technology and interaction paradigm, i.e. command language with a keyboard.

We observed that the effect of some individual features like visuo-spatial skills had a similar effect whatever the real or virtual condition. This factor mostly affected the probability of solving the task in the time window of the experiment. On the one hand, this result is interesting since it means that the virtual conditions were designed to enable the subjects to use their own same skills as they used in real situations, even in the context of the K condition. On the other hand, the poor performance of the subjects in the other virtual conditions (G, V, and G+V) means that the design of interaction and devices did not provide them with any significant help. At least, it should be improved by capitalising upon the various available and evolving technologies used as opportunities to enhance human capability and to help them in solving problems and achieving tasks. We have also shown that successful solving was mostly associated to strategies focused on reasoning on the problem space rather than on manipulating blocks. Indeed, subjects who succeeded with the puzzle displayed slower rates in terms of moves per minute, less attempts, as well as specific patterns of favoured blocks during their solving path. In particular, more complex blocks were favoured by solvers to start the solution in both real and virtual conditions.

Finally, we have also shown that time is a possible factor to explain the weak results in virtual conditions. Indeed, the number of moves per minute strongly depended on the conditions, whereas the number of moves per attempt was relatively similar whatever the conditions. This can be interpreted in the sense that systems with a slow rate require more time to enable subjects to achieve the same correct solution. We replicate here (and provide an explanation for) the results often reported in the literature, that assembly tasks

performed in virtual environments usually require more time than the same tasks performed in a real environment.

8.3 Contributions to informatics and human-computer interaction

8.3.1 Difficulties to map the knowledge from real setting onto the interactions commands in a virtual environment

We have adapted the moves and strategies found in the real setting to specify the design of interaction in the K, G and V conditions. However, it has been difficult to directly apply them, due to the specific constraints of devices and modalities. A consequence is that commands (or moves as reported in the empirical chapters) are only partly common.

It would be interesting to have a finer analysis of the differences due to the specific constraints of devices and technology exploited in the interface. In particular, we lack of a systematic documented method to match the design of commands in virtual environments with goals and sub-goals spontaneously formed by the subjects as they do the task in a real setting.

8.3.2 Performance with multimodal inputs and suggestions for improvement

As was remarked at the end of chapter 7, the highest success rates (percentage of solvers) were achieved by the users in the real settings while the users in the G and V groups had the lowest performance. Here is our analysis of the reasons behind the lower performance of the users for these groups.

The primary source of the inconvenience for the users is the time lag between their actions and the reactivity of the system, which includes the input taken by the system from the user and performing the requested action and showing the output. In the surgical robotic

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

community, the maximum standard time lag acceptable is 50 ms for haptics/gestural interaction and 300 ms for the vocal interaction.

Secondly, in the current system, the users did not have a fully realistic stereoscopic 3D vision. That is to say that in our case, it was a 2D projection of a 3D space on a large screen. In particular, due to this factor, the users lacked a precise and accurate sensation for the depth perception. In fact, there are different factors related to human visual perception that the technology makes (or can make) use of, for creating better and more usable systems. These factors include the various depth cues (such as shadows of objects, occlusion, and the size of objects, etc.) that humans use to gather information about the depth perception of a scene. Besides the depth cues, the other factors include the refreshing rate of the screen, the screen geometry, and ergonomics of the display (e.g. strain, comfort and unobtrusiveness), which can all play an important role in the human performance.

A third lacking aspect specific to the gestural modality was the lack of haptics and/or tactile feedback. This does not prevent us from looking for more reliable tracking and vocal recognition systems. Once one is through with this prerequisite step, only would it be possible to delve into the other issues (deeper issues, such as interfaces). At the moment, response times are too slow. In short, one requires fast, reliable and responsive systems which users are well-versed with.

Improvement in all these three aspects would enhance the usability as well as the users' sense of presence.

8.4 Perspectives

8.4.1 Research perspectives

Informal observations revealed that there were different kinds of interaction related errors that were made by the users. An analysis of these errors could help understand them better and help create a better and improved system in the future.

In the future, as an immediate extension of our work, one could evaluate errors of the different modalities used under virtual conditions. For example, for the V modality, three types of errors could occur. First, an invalid command is given by the user and unidentified by the system. In this type of error, the user has given the system a command that does not exist in the predefined command list. Second, an invalid command is given by the user and the system identifies it as some other command existing in the command list. Third, a valid command is given by the user and the system is not able to identify it correctly. A complete statistical analysis of all these errors will help us improve the vocal system.

For the G modality, the informal observations revealed that there were misinterpretations of the gestures by the system as well. Such errors were again of three types, as mentioned for the V modality. Additional complexity of the G modality was the time lags that were due to the network congestion etc. For future experimentation, we propose that the entire system support infrastructure should be locally available.

And last but not the least, in the G+V environment, the additional phenomenon that was informally observed was that users assumed that many more combinations of interactions were possible than was really the case. It would be interesting to observe and take note of such commands that the users spontaneously try to give to the system during pre-tests, (or to get users feedback by other methods such as questionnaires or interviews) and incorporate such commands as valid ones in the future version.

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

Perceived affordances of objects take factors such as *social-conditioning, past experiences* and *users' perception* of the object's behaviour and interaction into account. We propose that such experiences should be integrated in the system. Thus, for example in our case, since the blocks were supposedly of wooden material, they should make a noise that reinforces the user's perception as he/she were in fact working with real wooden blocks.

More concretely, in the current system, when a collision occurs, the feedback given to the user is purely visual. In the future perspective, we intend to introduce audio feedback for collisions. In addition to improving the usability and the user performance, it would also help increase the user's sense of immersion in the system. For the accomplishment of the goal of the user, in the existing system there were no visual aids how the task should be achieved (in our case, to build a cube). In future systems, visual aids such as step indicator arrows or the translucent structure of the workspace might help the user visualise the solution.

One could also try to make different improvements in the currently used modalities, for example one such idea is that in order to track the users' gestures more globally (and freeing him/her from the burden of always having the back of his/her hand always pointing upwards), a greater number of sensors should be placed all around him/her, permitting the user to move his/her hands in any orientation, which effectively would make it easier for the user to interact with the system.

Moreover, one could also create new ways and techniques for the user to manipulate the objects in VE. For example, for the G modality, one could experiment selecting an object by making an action as if shooting the block with a pistol.

Besides, one could possibly make use of other devices such as haptic joystick (with force feedback) or gyroscopic sensor based devices in the extension of work (already, iPhone/Wii remote are using the accelerometer/gyroscope). It would be particularly interesting to try out an isometric device like the joystick (and to compare its performance with the K, G, and V conditions) and to verify if joystick-based manipulation is better suited for our kind of free space assembling task. Normally, it is said that isometric devices perform better when greater precision is required. And since our results revealed that “Add” move was overall the most frequently performed one, and since in its final stages it requires docking, thus it would be interesting to see how usable the joystick is for the same task (and particularly towards the end, i.e. at the time of docking of the block).

Finally, one could consider making improvements in the set of instructions given to the user textually and verbally during the training session for learning the interactions with the system for the different modalities.

8.4.2 Applications (of manipulation of objects in 3D space)

Our work may find its applications in many industries, particularly the car manufacturing industry, or any other manufacturing industries involving assembly/disassembly or manipulation of objects in 3D space. Moreover, since many industries are increasingly moving their focus towards the use of VEs, any industry which makes use of VEs involving the manipulation of objects could find applications of our work.

Surgical robotics industry – which in fact came into being as a result of collaboration between surgical systems industry and the robotics industry – is using systems which help make manipulation of human organs by the surgeons. *Da Vinci* is one such example of a commercially available system. The issues involved and under focus in the current context are related to having greater control and precision of the surgical robots and making minimal

Object Manipulation in Virtual Environments

Thesis Report by Sarwan ABBASI

invasion of the human body. This field has many manipulation related open issues to solve and our work could possibly have applications in this domain as well.

And finally, the gaming industry may find use of our work, since already, our puzzle could in fact qualify to be a (strategy) game even in its present form and there are many other similar ones which are in fact already available that involve strategy and problem solving (e.g. Rubik's Games).

Appendix A

List of Commands

Following are the instructions and the list of test commands that all the users who had to solve the puzzle using the vocal modality had to first read out. The basic idea was to make sure that all the critical words were correctly recognised by the system, which was a pre-requisite, if the user was to solve the virtual puzzle.

Some of the commands produced the exact same effect (e.g. "Bouger à gauche"; "Déplacer à gauche"; "Déplacement à gauche" and "Bouger sur moins X" produced the same effect). The objective behind this was to allow more options to the user and let the user decide whichever option he/she preferred and found more natural. Moreover, in the case where the system did not understand one of the keywords, the user had the possibility to use an alternative word. There were however certain critical words for which there were no alternatives (e.g., the numerals "1" through "7"); thus it was essential that the system correctly recognised and interpreted those numerals, and in case it did not, then it was meaningless to carry on the experiment.

The precise instruction and the list of commands, that users were given, are as follows:

Instructions

Please read each phrase in the list below loudly. The sound recognition system can be activated and deactivated using the phrases "Commencer l'écoute" and "Arrêter l'écoute", respectively. ';' (semi-colon) denotes a long pause and ',' (comma) denotes a short pause.

In case of a problem with the sound-recognition system, e.g. in the output window, if the system shows a stream of gibberish etc., please read out the following "Annuler; Arrêter l'écoute;". Repeat if necessary until the gibberish ends and the system goes to the sleep state. Then, you must again start by the phrase "Commencer l'écoute", and then continue from the point you had left.

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

List of test commands

"Commencer l'écoute" ;

===

Choisir bloc numéro 1;

Prendre bloc numéro 1;

Prendre bloc 3 ;

Activer bloc 6 ;

Bloc [1, 2, 3, 4, 5, 6, 7] ;

Choisir la caméra ;

===

Bouger à gauche;

Translater à gauche;

Déplacer à gauche;

Déplacement à gauche;

Bouger sur moins X ;

[à gauche; à droite; en arrière; en avant; vers le haut; vers le bas] ;

===

Basculer à gauche;

Basculer à droite;

[vers l'arrière, vers l'avant] ;

Pivoter dans sens positif ;

Pivoter dans sens négatif ;

===

Désactiver ;

Annuler sélection ;

Zoomer / dé-zoomer ;

===

Accélérer / ralentir ;

Stop ;

===

"Arrêter l'écoute" ;

Appendix B

List of Commands and their Descriptions

1. List of commands / Vocabulary:

The table below contains the complete list of commands that were available to the user during the V modality.

How the following table is to be interpreted?

The commands can be placed in three broad categories (+“others” for commands that don’t fall in any of those categories), namely:

- | | |
|------|------------------------------------|
| i. | Selection |
| ii. | Movement (Translation/Rotation) |
| iii. | Stopping of translational movement |
| iv. | Other |

In the table below, the second column (column ‘b’) contains the initial part of the vocal command which may need further arguments (columns ‘c’ and ‘d’) to complete it. Some of the last arguments had their technical equivalents, (and thus the argument in column ‘d’ could be replaced by the argument in column ‘e’). The right-most column (column ‘g’) identifies the action category (*i.e. selection, translation, rotation, stop*).

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

Everyday natural language commands used for the V modality

(a)	(b)	(c)	(d)	(e)	(f)	(g)
#	Command in French	Argument	Argument completion	Equivalent technical alternative for argument completion	Command category/ Meaning	Action Category
1	Sélection Sélectionner Prends Prendre	bloc élément composant	Numéro [1..7]		Pick, Choose, Select, Take/ Selection of block	Selection
2a	Translater, Translation, Bouger, Déplacer, Déplacement	sur, suivant, selon	droite	[X Y Z]	Move, Translate, Displace/ Translation on +X axis	Translation
	"	vers la	gauche	"	" -X axis	
	"	en	haut	"	" +Y axis	
	"	en	bas	"	" -Y axis	
	"	en	arrière	"	" +Z axis	
	"	en	avant	"	" -Z axis	
2b	Rotation, Tourne, Tourner	sur, suivant, selon	l'avant	[X Y Z]	Rotation on +X axis	Rotation
	Rotation, Tourne, Tourner	vers	l'arrière	"	" -X axis	
	Pivoter		« Sens des aiguilles d'une montre »	"	" +Y axis	
	Pivoter	dans le	« Sens inverse/contraire des aiguilles d'une montre »	"	" -Y axis	
	Rotation, Tourne, Tourner	à	droite	"	" +Z axis	
	Rotation, Tourne, Tourner	à	gauche	"	" -Z axis	
3	STOP, Pause, Arrêter, Annuler				STOP, enough, Pause, Quit, Exit / Stops movement of the selected object	STOP

2. General comments about the vocal interaction technique

- Selection of a block implies both highlighting (designation) and selection of the block.
 - So the command “Sélectionner Bloc un” (which is French for “Select Block one”) implies highlighting and selection of the block numbered one.
- At any given time, only one object is selected, thus selecting one object automatically deselects the previous one.
- The translation (displacement) command shall make the block start moving in the said direction (to make the block stop would require an explicit STOP command. If the STOP command is not forthcoming, then the block will not stop until it collides with some other object or until when it comes in contact with the boundary of the manipulation area.
- The blocks rotate by exactly 90°.

3. Formal descriptions of the commands available to the user

Description of how those words would be interpreted by the system, (followed by examples), to perform ‘Selection’, ‘Translation’, ‘Rotation’, ‘Stop’, ‘Other’ commands:

a) Selection (to be interpreted in three steps)

{Sélection | Sélectionner | Prends | Prendre}
{bloc | élément | composant}
{numéro} {1 | 2 | 3 | 4 | 5 | 6 | 7} |

Note that in these formal descriptions (in green), the vertical bar sign ‘|’ signifies “OR” (i.e. only one of those words would need to be performed). The square brackets ‘[]’ signify “optional” (i.e. it may or may not be performed)

Examples:

Sélectionner bloc numéro 3
Sélectionner élément numéro 5
Sélectionner bloc 5

b) Translation (three steps)

{Translater | Translation | Déplacer | Déplacement}
{sur | suivant | selon}
{X | Y | Z}

Examples:

Translater suivant X
Déplacer sur Y

c) Rotation

{Rotation, Tourne, Tourner}
{sur, suivant, selon}
{X | Y | Z}

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

Examples:

Rotation selon X ;
Tourner sur (moins) X

d) Stop

STOP | Pause | Arrêter | Annuler

Bibliography

- [Abbasi2006] Abbasi, S., Ajaj, R. (2006), "*Co-manipulation and Co-modelling in Virtual Environments*", Report for Master's course on Virtual Reality, 2006.
- [Abbasi2009-a] Abbasi, S., Burkhardt, J.-M., Denis, M. (2009), "*Spatial problem solving: Assembling three-dimensional puzzles in real and virtual environments*", Cognitive Processing, Springer vol 10, pages 179-181.
- [Abbasi2009-b] Abbasi, S., Burkhardt, J.-M., Denis, M. (2009), "*Assister le montage d'objets en réalité virtuelle: étude préliminaire de la résolution de problèmes tridimensionnels en environnement réel*". In actes du 5eme Colloque de Psychologie Ergonomique EPIQUE, Nice, France 28-30 September 2009.
- [Axelsson2001] Axelsson, A.-S., Abelin, Å., Heldal, I., Schroeder, R., Wideström, J. (2001), "*Cubes in the cube: A Comparison of a Puzzle-Solving Task in a Virtual and a Real Environment*" CyberPsychology and Behavior. Vol 4, Issue 2: 279-286.
- [Bolt1980] Bolt, R. A. (1980), "*Put-that-there: Voice and gesture at the graphics interface*", ACM Siggraph '80: 7th annual conference on Computer graphics and interactive techniques, 262-270.
- [Boud2000] Boud, C., Baber, C., Steiner, S. J. (2000), "*Virtual Reality: a tool for assembly*", Presence: Teleoperators and Virtual Environments, vol. 9, number 5, 486-496.
- [Bowman1997] Bowman, D., Hodges L. F. (1997), "*An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments*", ACM 1997 Symposium on Interactive 3D Graphics
- [Bowman2001] Bowman, D., Kruijff, E., LaViola, J., Poupyrev, I. (2001), "*An Introduction to 3D User Interface Design*", Presence: Teleoperators and Virtual Environments, Vol. 10, No. 1, pages 96-108, 2001, MIT Press
- [Bowman2002-a] Bowman, D. A., Gabbard, J. L., Hix D. (2002). "*A survey of usability evaluation in virtual environments: Classification and comparison of methods*", Presence: Teleoperators and Virtual Environments, vol. 11, no. 4, pages 404-424, MIT Press
- [Bowman2002-b] Bowman, D. (2002), "*VE Input Devices*", Course Lecture Notes for Computer Science 5984, Virginia Tech.
- [Bowman2005] Bowman, D., Kruijff, E., LaViola, J., Poupyrev, I. (2005), "*3D User Interfaces, Theory and Practice*", Addison-Wesley 2005

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

- [Bowman2006] Bowman, D. A., Chen, J., Wingrave, C. A., Lucas, J., Ray, A., Polys, N. F., Li, Q., Haciahetoglu, Y., Kim, J.-S., Kim, S., Boehringer, R., Tao, N. (2006), “*New directions in 3D user interfaces*”. The International Journal of Virtual Reality, 2006, 5(2):3-14
- [Brogni2003] Brogni, A., Slater, M., Steed, A. (2003), “*More Breaks Less Presence*”, Presence 2003, 6th Annual International Workshop on Presence, 2003
- [Brooks1999] Brooks, F. (1999). “*What's Real About Virtual Reality?*”, IEEE Computer Graphics and Applications, 19(6), 16-27.
- [Burkhardt2003] Burkhardt, J.-M., Bardy, B., Lourdeaux, D. (2003). “*Immersion, réalisme et présence dans la conception et l'évaluation des environnements virtuels*”, Psychologie Française, 48, 35- 42.
- [Butler1994] Butler, B. E. (1994), “*Spatial puzzles: A guide for researchers*”, Canadian Psychology Psychologie Canadienne, Vol 35(1), Jan 1994. pp. 47-65. [Journal Article]
- [Clément1997] Clément, E., Richard, J.-F. (1997), “*Knowledge of Domain Effects in Problem Representation: The Case of Tower of Hanoi Isomorphs*”. Thinking and Reasoning (Psychology Press), Volume 3, Issue 2 April 1997, pages 133 - 157 (1997)
- [Demaine2007] Demaine, E. D., Demaine M. L. (2007), “*Jigsaw Puzzles, Edge Matching, and Polyomino Packing: Connections and Complexity*”, Springer-Graphs and Combinatorics (2007) 23[Suppl]:195–208
- [Denis2007] Denis, M., Loomis, J. M. (2007), “*Perspectives on human spatial cognition: memory, navigation, and environmental learning*”, Psychological Research (2007) 71:235–239, DOI 10.1007/s00426-006-0079-x
- [Eigler1930] Eigler, P. (1930), “*The Kohs Block Design Test*” (in Brief Reports), Child Development, Vol. 1, No. 4. (Dec., 1930), pp. 341-342
- [Férey2009] Férey, N., Nelson, J., Martin, C., Picinali, L., Bouyer, G., Tek, A., et al. (2009). “*Multisensory VR interaction for Protein-Docking: The CoRSAIRE project*”, Virtual Reality, 13(4), 273-293
- [Fröhlich2000] Fröhlich, B., Plate, J. (2000), “*The Cubic Mouse : A New Device for Three-Dimensional Input*”, Conference on Human Factors in Computing Systems, Proceedings of the ACM - SIGCHI conference on Human factors in computing systems (2000).
- [Fuchs2003] Fuchs, P., Moreau, G., Papin, J.-P. (Eds.). (2003). “*Traité de la Réalité Virtuelle*”, Paris: Presse de l'Ecole des Mines de Paris.
- [Fuchs2006] Fuchs, P., Moreau, G., Arnaldi, B., Guitton, P. (Eds.). (2006). “*Le traité de la Réalité Virtuelle : Les applications de la réalité virtuelle*”, vol. 4, Paris: Presses de l'école des mines de Paris.

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

- [Ganier2004] Ganier, F. (2004), “*Factors affecting the processing of procedural instructions: implications for document design*”, IEEE Transactions on Professional Communication 47 (1) (2004), pp. 15–26.
- [Gardner2008] Gardner, M. (2008), “*Origami, Eleusis, and the Soma Cube*”, Cambridge University Press, New York, NY, 2008.
- [Greeno1978] Greeno, J. G. (1978), “*A study of problem solving*”, in R. Glaser (Ed), Advances in Instructional Psychology (vol. 1, pp 13-75). Hillsdale, NJ: Lawrence Erlbaum Associates
- [Guiost2008] Guiost, B., Debernard, S., Poulain, T., Millot, P. (2008), “*Towards a Multi-sector cooperation in Air-Traffic Control supported by a Meta-Common Workspace*”, Proceedings of the 17th World Congress, The International Federation of Automatic Control
- [Hegarty2004] Hegarty, M. (2004), “*Mechanical reasoning by mental simulation*”, Trends in Cognitive Sciences, vol. 8, issue 6, June 2004, Pages 280-285
- [Hegarty2006] Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., Lovelace, K. (2006), “*Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning*”, Intelligence, vol. 34, issue 2 (2006), pages 151– 176
- [Heldal2005] Heldal, I., Anthony, S., Schroeder, R. (2005), “*Evaluating Collaboration in distributed Virtual Environments for a Puzzle Solving Task*”, Proceeding of the 11th International Conference on Human Computer Interaction, July 2005, Las Vegas
- [Hinckley1994] Hinckley, K., Pausch, R., Goble, J. C., Kassell, N. F. (1994), “*A Survey of Design Issues in Spatial Input*”, Proceedings of the 7th annual ACM symposium on User interface software and technology.
- [Ho2006] Ho, C., Eastman, C., Catrambone, R. (2006), “*An investigation of 2D and 3D spatial and mathematical abilities*”. Design Studies, Vol. 27, issue 4, July 2006, Pages 505-524
- [Jaimes2005] Jaimes, A., Sebe, N. (2005), “*Multimodal Human Computer Interaction: A Survey*”, Proc. 11th IEEE Int’l Workshop Human Computer Interaction (HCI), 2005
- [Jaimes2007] Jaimes, A., Sebe, N. (2007), “*Multimodal human-computer interaction: A survey*”, Computer Vision and Image Understanding, vol.108, no.1-2, pages 116-134, October, 2007
- [Kabbash1993] Kabbash, P., MacKenzie, I. S., Buxton, W. (1993), “*Human Performance Using Computer Input Devices in the Preferred and Non-Preferred Hand*”, Proceedings of the INTERACT’93 and CHI’93

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

conference on Human factors in computing systems, pages 474-481,
ACM

- [Kadri2007] Kadri, A., Lécuyer, A., Burkhardt, J.-M., Richir, S. (2007), “*Visual Appearance of User’s Avatar Can Influence the 3D Manipulation of Both Real Devices and Virtual Objects*”. In proceedings of 3D User Interfaces Symposium jointly to IEEE Virtual Reality 2007, March 10-14 Charlotte, North Carolina USA.
- [Kato2003] Kato, Y., Takeuchib, Y. (2003), “*Individual differences in wayfinding strategies*”. Journal of Environmental Psychology, vol. 23, no. 2, pages 171–188 (2003).
- [Kaufmann1979] Kaufmann, G. (1979), “*Visual imagery and its relation to problem solving*”, Bergen: Universitetsforlaget.
- [Keehner2006] Keehner, M. (2006), “*Learning a Spatial Skill for Surgery: How the Contributions of Abilities Change With Practice*”, Applied Cognitive Psychology, vol. 20, no. 4, pages 487–503, John Wiley and Sons (2006).
- [Kotovsky1985] Kotovsky, K., Hayes, J. R., Simon, H. A. (1985), “*Why are some problems hard? Evidence from Tower of Hanoi*”, Cognitive Psychology, vol. 17, no. 2, 248-294, Elsevier
- [Kotovsky1990] Kotovsky, K., Simon, H. A. (1990), “*What Makes Some Problems Really Hard: Explorations in the Problem Space of Difficulty*”, Cognitive Psychology, vol. 22, no. 2, pages 143-183 (1990), Elsevier.
- [Larousse1991] Larousse, Grande dictionnaire de la psychologie (1991 edition), Paris, Larousse.
- [Lautenbach1999] Lautenbach, M. A., Schegget, I. S., Schoute, A. M., Witteman, C. L. (1999), “*Evaluating the usability of web pages: A case study*”, available at: <<http://www.phil.uu.nl/preprints/ckipreprints/PREPRINTS/preprint011.pdf>>, on 19-10-2010
- [Lécuyer2002] Lécuyer, A., Megard, C., Burkhardt, J.-M., Lim, T., Coquillart, S., Coiffet, P., Grau, L. (2002). “*The effect of haptic, visual and auditory additional information on an insertion task*”, In Immersive Projection Symposia IPT-VR 2002, 24-28 march 2002, Orlando.

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

- [Li1999] Li, G. Y., Haslegrave, C. M. (1999). “*Seated work postures for manual, visual and combined tasks*”. *Ergonomics* 42 (8), 1060–1086.
- [Lok2003-a] Lok, B., Naik, S., Whitton, M., Brooks, F. P. Jr. (2003), “*Effects of handling real objects and avatar fidelity on cognitive task performance in virtual environments*”, *Virtual Reality*, pages 125-132, Proceedings of IEEE March 2003
- [Lok2003-b] Lok, B., Naik, S., Whitton, M. C., Brooks, F. P. (2003). “*Effects of Handling Real Objects and Self-Avatar Fidelity on Cognitive Task Performance and Sense of Presence in Virtual Environments*”. *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 6, pages 615-628, MIT Press
- [MacKenzie1994] MacKenzie, C. L., Iberall, T. (1994), “*The grasping hand*”, *Advances in psychology*, vol. 104, Amsterdam, Netherlands: North-Holland, Elsevier Science (1994)
- [Nam2008] Nam, C. S., Shu, J., Chung, D. (2008), “*The roles of sensory modalities in collaborative virtual environments (CVEs)*”, *Computers in Human Behavior*, Elsevier, vol.24, no.4, pages 1404-1417, July 2008.
- [Newell1972] Newell, A., Simon, H. A. (1972), “*Human problem solving*”, Prentice-Hall.
- [Nichols2000] Nichols, S., Haldane, C., Wilson, J. R., (2000), “*Measurement of presence and its consequences in virtual environments*”. *International Journal of Human-Computer Studies*, vol. 52, no. 3, pages 471–491
- [Nielsen1994] Nielsen, J., “*Heuristic evaluation*”, In Nielsen, J. And Molich, R.L (eds.) *Usability Inspection Methods*, New York: John Wiley & Sons, 1994. Available at: <<http://teddyso.com/Materials/Staffordshire/SDM/Lecture%208%20Usability%20Heuristics.doc>>, on 19-10-2010
- [Norman1988] Norman, D. (1988), *The Design of Everyday Things*. New York: Doubleday/Currency.
- [Oviatt2000] Oviatt, S.L., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., Ferro, D. (2000), “*Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research*

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

directions", Human-Computer Interaction, vol. 15, no. 4, pages 263–322, L. Erlbaum Associates Inc. 2000.

- [Peres2008] Peres, S. C., Best, V., Brock, D., Shinn-Cunningham, B., Frauenberger, C., Hermann, T., Neuhoff, J. G., Nickerson, L. V., Stockman, T. (2008), "*Auditory Interfaces*", HCI Beyond the GUI Design for Haptic, Speech, Olfactory, Other Nontraditional Interfaces 2008, Pages 147-195, Elsevier Inc.
- [Pettinaro1999] Pettinaro, G. C. (1999), "*Behavior-based peg-in-hole*". In *Robotica*, 17, (1999), 189–201.
- [Petzold2004] Petzold, B., Zaeh, M. F., Deml, B., Egermeier, H., Schilp, J., Clarke, S. (2004). "*A Study on Visual, Auditory, and Haptic Feedback for Assembly Tasks*", *Presence*, 13(1), 16-21.
- [Poupyrev1997] Poupyrev, I., Weghorst, S., Billingham, M., Ichikawa, T. (1997), "*A framework and Testbed for Studying Manipulation Techniques for Immersive VR*" *Virtual Reality Software and Technology Proceedings of the ACM symposium on Virtual reality software and technology* (1997)
- [Poupyrev2000] Poupyrev, I. (2000), "*3D Manipulation Techniques*", Lecture Slides, presentation, as part of the course: 3D User Interface Design: Fundamental Techniques, Theory, and Practice, delivered at SigGraph-2000
- [Rasmussen1980] Rasmussen, J. (1980), "*Notes on human error analysis and prediction*", in *G. APOSTOLAKIS, S. GARRIBA ans G. VOLTA (Eds)*, Proceedings of the NATO advanced study institute on synthesis and analysis methods for safety and reliability studies, pages 357 - 389, Plenum Press, New York, 1980
- [Rasmussen1983] Rasmussen, J. (1983), "*Skills, rules, and knowledge; signals, signs and symbols, and other distinctions in human performance models*", *IEEE Transactions on systems, man cybernetics*, vol. 13, no. 3, pages 257-266, 1983
- [Rehman2002] Rehman, K., Stajano, F., Coulouris, G. (2002), "*Interfacing with the Invisible Computer*", *NordiCHI*, 2002

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

- [Richard1993] Richard, J.-F., Poitrenaud, S., Tijus, C. (1993). “*Problem-solving restructuration: Elimination of implicit constraints*”. *Cognitive Science*, 17, 497-529.
- [Rieber1995] Rieber, L. P. (1995), “*A historical review of visualization in human cognition*”, *Educational Technology Research and Development*, vol. 43, no. 1, pages 45-56. 1995
- [Roussou2006] Roussou, M., Oliver, M., Slater, M. (2006). “*The virtual playground: an educational virtual reality environment for evaluating interactivity and conceptual learning*”, *Virtual Reality*, vol. 10, no. 3, pages 227-240, Springer
- [Ruiz2010] Ruiz, N., Chen, F., Oviatt, S. (2010), “*Multimodal Input*”, *Multimodal Signal Processing*, 2010, Pages 231-255
- [Shepard1971] Shepard, R. N.; Metzler, J. (1971) “*Mental Rotation of Three-Dimensional Objects*” *Science*, vol. 171, no. 3972, pages 701-703, 1971
- [Shepard1982] Shepard, R. N., Cooper, L. A. (1982), “*Mental images and their transformations*”, Cambridge: MIT Press/Bradford
- [Simon1976] Simon, H. A., Hayes, J. R. (1976), “*The Understanding Process: Problem Isomorphs*”, *Cognitive Psychology* vol.8, no. 2, pages 165-190, (1976)
- [Slater1995] Slater, M., Usoh, M., Chrysanthou, Y. (1995), “*The Influence of Dynamic Shadows on Presence in Immersive Virtual Environments*”, *Second Eurographics Workshop on Virtual Reality*, M. Goebel, ed., Monte Carlo, January 1995.
- [Stanney2003] Stanney, K. M., Mollaghasemia, M., Reevesa, L., Breaux, R., Graeber, D. A. (2003), “*Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design*”, *International Journal of Human Computer Studies* 58 (4), 447–481.
- [Svendsen1991] Svendsen, G. B. (1991), “*The influence of interface style on problem solving*”, *International Journal of Man-Machine Studies*, vol. 35, no. 3, pages 379-397, (1991)
- [Torkington2001] Torkington, J., Smith, S. G. T., Rees, B. I., Darzi, A. (2001), “*Skill transfer from virtual reality to a real laparoscopic task*”, *Surgical Endoscopy*, vol. 15, no. 10, pages 1076-1079, Springer 2001

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

- [Unger2001] Unger, B. J., Nicolaidis, A., Berkelman, P. J., Thompson, A., Klatzky, R. L., Hollis, R. L. (2001), “*Comparison of 3D Haptic Peg in Hole Tasks in Real and Virtual Environments*”, In International Conference on Intelligent Robots and Systems IROS, Maui Hawaii, 29 Oct-3 Nov 2001
- [Vandenberg1978] Vandenberg, S.G., Kuse, A.R. (1978), “*Mental rotations, a group test of three-dimensional spatial visualization*”, *Perceptual and Motor Skills* 47 (1978), pp. 599–604
- [Vinaya2004] Vinayagamoorthy, V., Brogni, A., Gillies, M., Slater, M., Steed A. (2004), “*An investigation of presence response across variations in visual realism*”. The 7th Annual International Presence Workshop, pages 148-155, October 2004
- [Wang1997] Wang, Y., MacKenzie, C. L., Summers, V. A. (1997), “*Object manipulation in virtual environments: human bias, consistency and individual differences*”. Extended In Abstracts of the Conference on Human Factors in Computing Systems CHI '97 /ACM, 349-350.
- [Wang1998] Wang, Y., MacKenzie, C. L., Summers, V. A., Booth, K. S. (1998), “*The Structure of Object Transportation and Orientation in Human-Computer Interaction*”. Proceedings of the SIGCHI conference on Human factors in computing systems, pages 312-319, ACM press, 1998
- [Wartenberg2004] Wartenberg, C., Dukic, T., Falck, A. C., Hallbeck, S. (2004), “*The effect of assembly tolerance on performance of a tape application task: A pilot study*”, *International Journal of Industrial Ergonomics*, 33(4), 369–379
- [Werkhoven1998] Werkhoven, P. J., Groen, J. (1998), “*Manipulation Performance in Interactive Virtual Environments*”, *Human Factors*, 40, 3, pp 432-442
- [Wideström2000] Wideström, J., Axelsson, A.-S., Schroeder, R., Nilsson, A., Heldal I., Abelin, Å. (2000), “*The collaborative cube puzzle: a comparison of virtual and real environments*”, *Collaborative Virtual Environments*, Proceedings of the 3rd International Conference on Collaborative Virtual Environments (2000)
- [Zacks2005] Zacks, J. M., Michelon, P. (2005). “*Transformations of visuospatial images*”, *Behavioral and Cognitive Neuroscience Reviews*, vol. 4, no. 2, pages 96-118, 2005

Object Manipulation in Virtual Environments
Thesis Report by Sarwan ABBASI

- [Zahoric1998] Zahoric, P., Jenison, R. L. (1998). “*Presence as being-in-the-world*”, Presence: Teleoperators and Virtual Environments, vol. 7, no. 1, pages 78–89, MIT Press 1998
- [Zhai1993] Zhai, S., Milgram, P. (1993). “*Human Performance Evaluation of Manipulation Schemes in Virtual Environments*”. Proceedings of the IEEE Virtual Reality Annual International Symposium, 155-161
- [Zhai1998] Zhai, S. (1998). “*User Performance in Relation to 3D Input Device Design*”. Computer Graphics, vol. 32, no. 4, pages 50-54, ACM 1998
- [Zhang2005] Zhang, Y., Sotudeh, R., Fernando, T. (2005), “*The use of visual and auditory feedback for assembly task performance in a virtual environment*”, Proceedings of the 21st spring conference on Computer graphics, pages 59-66, ACM Press 2005

Internet sources (URLs)

- [ARTracking] Advanced Real-time Tracking GmbH, <<http://www.ar-tracking.de/>>, retrieved on 19-10-2010
- [PuzzlesCat] Malcolmson, Kathleen, “*What are puzzles and why are they puzzling?*”, <<http://www.cleverwood.com/>> , retrieved on 19-10-2010
- [Virtools] Virtools, <<http://www.virttools.com/>>, retrieved on 19-10-2010
- [WolframMathW] Weisstein, Eric W. “*Soma Cube*”. From MathWorld--A Wolfram Web Resource. <<http://mathworld.wolfram.com/SomaCube.html/>>, retrieved on 19-10-2010