T.R.
GEBZE INSTITUTE of TECHNOLOGY
GRADUATE SCHOOL of ENGINEERING and SCIENCES

3D INTERACTION INTEGRATING 2D MOUSE and 3D HEAD TRACKING

ASLIHAN TECE
MASTER THESIS
COMPUTER ENGINEERING DEPARTMENT

GEBZE
2014

T.R.
GEBZE INSTITUTE of TECHNOLOGY
GRADUATE SCHOOL of ENGINEERING and SCIENCES

3D INTERACTION INTEGRATING 2D MOUSE and 3D HEAD TRACKING

ASLIHAN TECE

MASTER THESIS

COMPUTER ENGINEERING DEPARTMENT

SUPERVISOR
DOC. DR. MEHMET GOKTURK

GEBZE
2014

GEBZE YÜKSEK TEKNOLOJİ ENSTİTÜSÜ	**YÜKSEK LİSANS JÜRİ ONAY FORMU**

GYTE Mühendislik ve Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 20/01/2014 tarih ve 2014/4 sayılı kararıyla oluşturulan jüri tarafından 02/05/2014 tarihinde tez savunma sınavı yapılan Aslıhan Tece'nin tez çalışması Bilgisayar Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE
(TEZ DANIŞMANI) : Doc. Dr. Mehmet Göktürk

ÜYE : Doç.Dr. Hacı Ali MANTAR

ÜYE : Doç. Dr. İlyas KANDEMİR

ONAY

GYTE Mühendislik ve Fen Bilimleri Enstitüsü Yönetim Kurulu'nun .... tarih ve ..../........ sayılı kararı.

İMZA/MÜHÜR

SUMMARY

This work demonstrates a three dimensional (3D) multimodal user interface integrating 2D mouse input and 3D head tracking for 3D desktop virtual environments (DVE). A test-bed is proposed to evaluate the functionality and usability of this 3D interface. The traditional direct manipulation techniques are performed by the rotation of head to interact with the objects in 3D DVE composed of a relatively up-to-date desktop computer, a webcam and a mouse. User performance is compared during simple and complex translation and rotation tasks to understand both the effects of possible mappings of 2D interaction device in a 3D environment and the capability of head tracking as a hands-free 3D user interface for multimodal purposes. The results indicate how suitable head tracking is as a 3D interface integrated with 2D interaction device to be used in a 3D DVE. The next step after this test will be an experiment to identify benefits of this interface in an actual video game environment as a means to find out whether it is possible to transfer the findings to the game environment or not, and the usability level changes compared to the findings of the test-bed.

Key Words: Human Computer Interaction (HCI); Head Tracking; 3D Interaction; Multimodal Interfaces; Usability.

ÖZET

Bu çalışma ile üç boyutlu (3D) masaüstü sanal ortamlarda kullanılmak üzere iki boyutlu (2D) mouse ile 3D baş takibinin entegrasyonuna dayalı üç boyutlu çoklu biçimli bir kullanıcı arayüzü önerilmiştir. Bu arayüzün işlevselliğini ve kullanılabilirliğini test etmek üzere bir test ortamı da tasarlanmıştır. Güncel bir masaüstü bilgisayardan, web kamerasından ve bir mousedan oluşan test ortamında, geleneksel doğrudan elle müdahele teknikleri kafa hareketleri ile gerçekleştirilecektir. Basit ve karmaşık taşıma ve döndürme görevleri sırasında kullanıcı performansı ölçülerek 2D etkileşim cihazının üç boyutlu ortama olan izdüşümünün yarattığı olası etkiler ve kafa takibinin eller serbest bir üç boyutlu arayüz olmaya yönelik kabiliyetleri anlaşılmaya çalışılmıştır. Test sonuçlarının kafa takibi ve iki boyutlu etkileşim cihazının entegrasyonunun üç boyutlu masaüstü sanal ortamlarda kullanılmasının uygunluğuna yönelik ipuçları vereceği öngörülmüştür. Bu çalışmadan sonraki ilk adım önerilen arayüzün bir oyun ortamına entegre edilerek kullanılabilirlik testlerinin tekrarlanması ve elde edilen sonuçların test ortamı sonuçlarıyla karşılaştırılması olacaktır.

Anahtar Kelimeler: İnsan Bilgisayar Etkileşimi; Baş Takibi; Üç Boyutlu Etkileşim; Çoklu Biçimli Arayüzler; Kullanılabilirlik.

ACKNOWLEDGEMENTS

I want to thank to my supervisor Doç. Dr. Mehmet Göktürk for his guidance and support. Without his mentorship, this experience would not have been as valuable as it had been.

Also, thanks heaps to my beloved husband E. Emre Bayrak and my bigger family—mums and dads and awesome-friends for their help and support in many ways all through my life.

TABLE of CONTENTS

LIST of ABBREVIATIONS and ACRONYMS

Abbreviations and Acronyms		**Explanations**
2D	:	Two Dimensional
3D	:	Three Dimensional
3DUI	:	Three Dimensional User Interface
DOF	:	Degrees of Freedom
DVE	:	Desktop Virtual Environment
HCI	:	Human Computer Interaction
HMD	:	Head Mounted Display
HOMER	:	Hand-centered object manipulation extending ray-casting is a 3D interaction method.
MMHCI		MultiModal Human Computer Interaction
UI	:	User Interface
VE	:	Virtual Environment
VR	:	Virtual Reality

LIST of FIGURES

LIST of TABLES

1. INTRODUCTION

Aim of this work is to investigate the possibility of more efficient interaction for 3D Desktop Virtual Environments (DVEs) with the help of head tracking. A variety of usability experiments have been held for immersive environments for several years; however, there is not much research into non-immersive 3D environments supporting 3D interaction. Even though conventional metrics of Human Computer Interaction (HCI) are about the investigation on usability of the interactions and interfaces, in this study the goal is highlighting and investigating new ways to conceive the relationship between users and their electronic devices—machines aka. computers— that may lead to an innovation in the field.

While informally observing users playing video games or even during the performance of attention demanding tasks in everyday life, following actions with head is found to be quite common. Getting closer to screen while jumping over, rolling one's head right/left while driving, mimicking the rotations by head while flying an aircraft; hence, assisting the motion in a way with head is a part of almost all players' performance. This suggests that head tracking based interaction can be a quite powerful candidate as an intuitive 3D interaction interface.

This work started with the questions/ideas below (as the kick-start questions), leaded to some others, managed to create approaches to explore results of some of them, and left some new generated questions to be identified and experimented in future studies. I will name them as key questions, try watching around them within the test-bed and will match the findings of the experiments back to them.

Key Question #1: Might head tracking and 2D mouse integration be efficient for 3D interaction in DVEs?

First key question creates other questions around the efficiency of interaction, how we define it and how we evaluate it. Test-bed will try covering these series of questions leading to usability evaluation of interaction and interfaces.

Key Question #2: Might head tracking be able to solve the dimension problems in 3DOF rotation needs as of flight simulators and such game object interactions?

Key Question #3: Can head tracking be a good interaction interface for desktop virtual environment to increase immersion?

Bearing these questions in mind, this study tries to identify suitable design and experiments for 3D interaction in DVE to find out the capability of head tracking as a 3D interaction means mostly focusing on the first key question, not omitting the others; but, leaving them for future studies.

2. INTERACTION

Merriam-Webster defines interaction as the “mutual or reciprocal action or influence”. Quite similarly Human Computer Interaction can be defined as the reciprocal action between human and computer, defining the action and feedback cycle as the core of interaction. Interaction techniques are the means for user to accomplish a task providing the user with hardware and software elements towards their achievement. Interaction techniques require input technologies and interface design to gather the input, evaluate and respond. Interaction designer should understand the input technology and the needs of environment well enough to be able to provide efficient interface design. This includes everything starting with the data created by input device, the ways of interpreting this data to the feedback provided to the user as the tiny cycle of interaction with the environment.

Input device, in this sense, is the main character that starts the interaction in the first place, and they are used to complete elemental tasks on a computer. Ken Hinckley’s report [41] from the study of Foley, Wallace and Chan (1984), proposing six elemental tasks that all user interface transactions are composed of creates a meaningful base to start thinking about the tasks and interaction:

- *“Select:* Indicating object(s) from a set of alternatives.
- *Position:* Specifying a position within a range, such as a screen coordinate.
- *Orient:* Specifying a rotation, such as an angle, or the three dimensional orientation of and object in virtual environment.
- *Path:* Specifying a series of positions or orientations over time, such as drawing a freehand curve in a paint program.
- *Quantify:* Specifying a single numeric value.
- *Text:* Specifying a symbolic data such as a sequence of characters*”*.

Hinckley finishes his words by declaring if a computer system is allowing a user to accomplish all six of these elemental tasks, then in principle the user can use the system to accomplish a computer based task. Quite similarly, if the interface developed with this design for the particular purposes allows the user to accomplish the tasks designed and developed within this environment, then the system can be evaluated as successful since the user is able to use the system to accomplish the tasks within the virtual environment.

2.1. 3D Interaction

3D interaction is composed of behavioral primitives described as universal tasks [24]; navigation, selection, manipulation and system control.

- Navigation

Navigation encapsulates two components; *travel* to be the motor component as user displacement in space, and *wayfinding* to be cognitive component as a cognitive behavior to define a route in a VE [17].

- Selection

Selection can simply be thought as selection of an object among others. Manipulation of an object requires selection as well; therefore, selection task is mostly combined with manipulation.

- Manipulation

Manipulation refers to the modification of the state of an object after the object is selected. Manipulation is composed of two primitive actions that create the change of state: translation and rotation. Translation refers to the change in position of the object by means of location whereas rotation refers to the change in position by means of the orientation. Furthermore, manipulation can also be split into canonical tasks to break down the complexity as: (1) *position*, the task of positioning an object from an initial to a final, terminal, position; (2) *selection*, the task of identifying an object; and (3) *rotation*, the task of rotating an object from an initial to a final orientation [18].

- System Control

System control refers to a task in which a command is applied to change either the state of the system or the mode of interaction [17].

Several interaction techniques are experimented in conjunction with another to perform one of these tasks or a combination of these tasks in immersive and semi-immersive 3D environments. In this study, 2D mouse input and head tracking are integrated to perform selection and manipulation tasks; manipulation including travel and rotation of the objects. This integration is explained in Chapter 4. Research Direction.

2.2. 3D Interaction Techniques

3D interaction techniques are developed from the metaphors to carry out the universal tasks explained above. Only selection and navigation techniques will be mentioned in this study since manipulation is the core interaction in this particular virtual environment.

2.2.1. Selection and Manipulation techniques

Selection techniques are classified by task decomposition as each selection task is composed of three stages: indication of the object to be selected, confirmation of selection and feedback [18].

Any manipulation task can easily be considered to include selection as well since the object should be selected to perform any level of manipulation. Therefore, manipulation techniques do include selection and are capable of performing the canonical tasks mentioned before. The manipulation techniques are restricted by the capability of the input device; therefore, the design of interaction techniques should be suitable to the characteristics of the input device [18]. When a taxonomy for manipulation techniques are considered, Poupyrev et al [34] evaluates it in two groups depending on the metaphors described in subsection 2.1.2 as egocentric and exocentric metaphors. Here, ray-casting is mentioned as a technique for interaction by pointing; Virtual Hand and Go-Go are mentioned as direct manipulation techniques for interaction with virtual hand, or sometimes called *virtual hand* techniques[1].

---

[1] For further information in various interaction techniques as well as more information about the interaction techniques mentioned here, reader should refer to the 3D UI Theory and Practices by Bowman et al 0.

- Ray-casting

Ray-casting technique is a pointing based interaction technique based on a virtual ray pointed to the objects. Several versions of ray casting has been implemented within the years and other techniques, such as flashlight and aperture techniques originating from ray-casting, improved versions of ray-casting for precision and occlusion are developed to overcome the shortcomings of this technique [24], [29], [33]. Figure 2.1 below is sourced from [34].

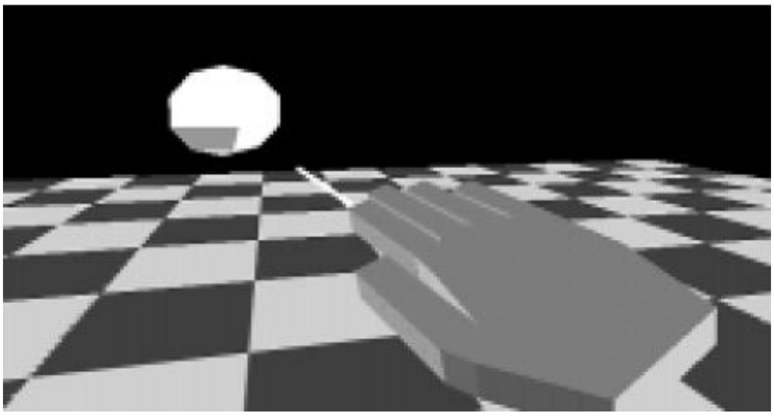

Figure 2. 1: Ray-casting technique is based on a ray sent from the user's hand to the object.

- Simple Virtual Hand

The virtual hand technique is a direct mapping of the user hand motion into the correspondent motions of the selected object in a virtual environment. This motion can be linearly or non-linearly scaled to establish the correspondence between the input and environment coordinate systems. The 3D cursor is generally shaped like a human hand. The selected object is attached to the hand and can be manipulated with the input device. Main shortcoming of this technique in immersive VEs is impossibility of reaching to an object that is far away since the objects should be in arm’s distance since interaction is supported with a wand or a haptic-glove. Simple virtual hand is also known as classical virtual hand. Figure 2.2 below is sourced from [30].

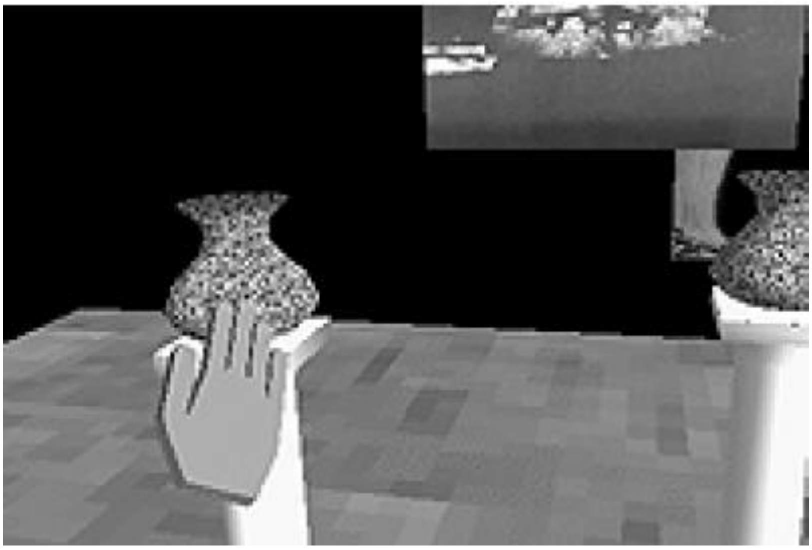

Figure 2. 2: Virtual Hand Technique: the user's reach is limited by arm length.

- Go-Go

Go-Go technique attempts to improvement virtual hand technique by allowing the user extend their arm as a virtual arm capable of selecting far away objects when the user's physical hand goes beyond a certain distance from the body to reach the object, via a ray casted from the hand and aligned with the direction of the arm. The movements of the virtual hand still correspond to the user’s hand movements with a nonlinear mapping to an infinitely growing virtual arm [30], possibly following a polynomial function. This technique was also developed for and used within immersive virtual environments mostly supported with a HMD haptic-glove or wand couple. Figure 2.3 below is sourced from [30].

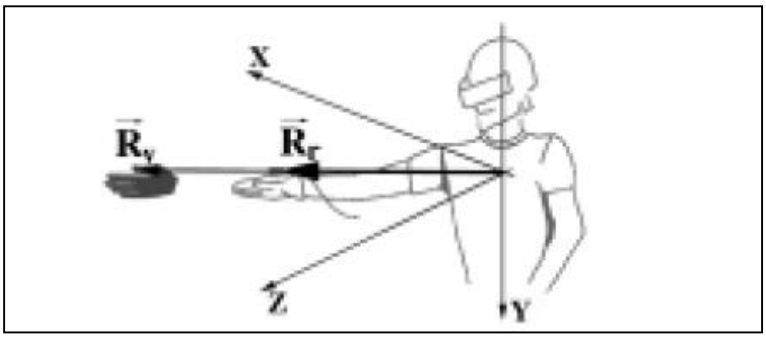

Figure 2. 3: Go-Go Technique: mapping between user's hand and virtual hand in coordinate system.

Also, there are several improved versions of Go-Go technique developed in several studies as Fast Go-Go Technique, Stretch Go-Go Technique, Indirect Go-Go etc.

Beside the originator techniques, there are also hybrid techniques that bring the capabilities of a couple of techniques together. HOMER is a hybrid technique that combines ray-casting and virtual hand manipulation [17], [29] to overcome the shortcomings of both techniques. In HOMER, the incapability of virtual hand in selection and the incapability of ray-casting in manipulation are eliminated by this combination. The object is selected by a ray-casting technique, and the user's virtual hand instantly moves to grab the object and attaches to it. From that moment, the user can manipulate the object exactly as direct hand manipulation with virtual hand.

2.2.2. Metaphors/ Mental models for 3D Interaction

3D Interaction techniques are based on metaphors that are originated from mental models. Mental models are quite important in user interface design by means of creating the connection between a familiar area of knowledge to a new introduced area or computer phenomena [36]. 3D interaction design shares the same approach with classical user interface design and values metaphors and mental models. The better the metaphor is understood and accepted to represent the interaction, the higher the success of the interaction is. Since the users are already familiar with real world artifacts and the ways to interact with them, understanding the purpose and method of using 3D interaction techniques that are based on the real world experience and interaction is said to be easier and commonly recalled as the real-world-metaphor.

Interaction techniques are classified by means of metaphors as well as task decomposition as described for selection techniques in previous section. Poupyrev et al proposed a metaphor based taxonomy for 3D manipulation techniques each of which forms the fundamental mental model of a technique, perceptually defining what users can do as *affordances* and what they cannot do as *constraints* [34]. Furthermore, exocentric and egocentric metaphors are defined based on basic interaction metaphors referring to the basic interaction techniques. Poupyrev et al explains in their study that the origin of the taxonomy "exocentric and egocentric techniques" depends on the studies of cockpit displays to distinguish between two fundamental frames of reference for user interaction with VEs. Exocentric interaction, which is also recalled as the God’s eye viewpoint, enables users interact with VEs from the outside of the environment as in the World-In-Miniature technique, which allows manipulation of objects by interacting with their representations in a miniature model of the environment held by the user [38]. On contrary, *egocentric* interaction depends on the user interaction from inside the environment. Therefore, user does not obtain external information about the environment as a whole. Virtual hand and virtual pointer metaphors are the two basic metaphors evaluated as egocentric metaphors.

Table 2. 1: Taxonomy of VE Manipulation Techniques by metaphor.

VE Manipulation Techniques

• Exocentric Metaphors
  • World-in Miniature
• Egocentric Metaphors
  • Virtual Hand Metaphors
    • Simple Virtual Hand Technique
    • Go-Go Technique
  • Virtual Pointer Metaphors
    • Ray-casting Technique
    • Apperture Technique
    • Flashlight Technique

In this study, virtual hand metaphor is the main model of direct interaction; however, interaction cannot be easily defined to be egocentric as defined in immersive VEs since the presence in 3D DVEs is limited compared to the immersive VEs. Even though means of interaction is centered within the environment rather than from outside of the environment as in world-in miniature, user has the chance to see the big picture and interact with the certain areas as an asset of the fish tank view. Therefore, the implemented technique will be neither exocentric nor egocentric even though it is based on the simple virtual hand technique, which is an egocentric virtual hand metaphor.

Since metaphors are the mental connections between user's knowledge from any known real-world experience to the interaction defined, they have impacts on the success and the performance of the interaction technique; moreover, interaction itself as an experience. See Section 2.2.3 for the performance of 3D interaction and effects of metaphors in 3D interaction.

2.2.3. Performance of 3D Interaction

Several studies have been conducted for the usage of haptics, HMDs, wands, joysticks etc. in VEs as well as traditional mouse and keyboards to evaluate task performance and usability of both 3D interaction methods and the metaphors leading to these methods. The capability of the input device, domain of interaction and ease of use are taken into account as other effects in interaction evaluations for VR and/or VEs [23], [24], [25] as well as the capability of the interaction technique. A good example referring to the capabilities of the interaction techniques is explained in Bowman and Hodges study [29]. In this study, Bowman and Hodges report that even though Ray-Casting techniques make grabbing virtual objects easy as a selection technique, they fail in manipulation as manipulation via the ray is difficult. They also report that arm-extension techniques provide natural and efficient manipulation; however, getting the hand in the correct position to grab the objects is hard. Therefore, the capabilities or shortcomings of the interaction techniques should also be considered so that the evaluation based on the environment variables are not affected by the side effects. Also, a technique can be vastly unsuccessful at certain tasks whereas successful at others. Hence, each technique should be used for particular environments and particular types of interaction that they are successful at.

In order to evaluate the success of interaction, different performance metrics are chosen thorough different studies. In the test-bed environment that Bowman used to compare two 3D interaction techniques to carry out simple tasks and generic 3D interaction the success rate of the users was the main measure for evaluation [26]. Also, in this study, the distance between the user and the targeted object, and the size of object to be selected were noted as external effects on performance for the selection task referring to the selection of an object among a group of objects within certain environment variables. In another evaluation that Poupyrev carried out [27], two interaction techniques, Go-Go and ray casting were evaluated across selection and manipulation tasks performed by the user revealing one technique to create better performance than other by comparing the environment variables; distance between the object and user for selection tasks, object size for manipulation tasks. These two experiments help us conclude that effects of environment variables should be taken into account cleverly to understand the success of the interface.

Poupyrev et al also conducted another study to find out the effects of *object distance*, *object size* and *visual feedback* on the user performance with basic interaction techniques on virtual object selection and repositioning tasks [34]. Their systematic evaluation includes comparison between the selection performance of virtual pointer and virtual hand metaphors for objects of different sizes located both close to the user and at-a-distance. All experiments and tests were conducted in a virtual reality test-bed with HMDs and tracking sensors. Therefore, even though this study is a significant study considering the environment variables, and gives good insight towards designing this system, the results may not correlate or give relevant head start.

On the other hand, McMahan et al. proposed a new technique to separate the effects of immersion and 3D interaction techniques for a clearer evaluation of the techniques claiming that direct comparison of immersive systems and non-immersive systems is insufficient [25]. They introduced a new Desktop Oriented Interaction Technique called DO-IT, which is a keyboard- and mouse-based technique to perform actions in Collaborative Analytical Visualization Environment (CAVE) environment. The results of their experiment show that the interaction technique had a significant effect on object manipulation time; however, the effect of input devices is not clearly stated to result in the performance differences observed.

2.2.4. Implementation Issues

- Selection

Object selection is one of the primitive tasks in any interactive VE. It is essential to incorporate adequate feedback while implementing object selection. The user must know which and when an object is chosen for selection and whether the task was successful or not. Generally the selection is acknowledged by highlighting the object or its bounding box. Other than the successful selection feedback, main issue arose from the appropriate selection technique for appropriate VE. Also, the success of this task can highly be dependent on the environmental factors as well as the shortcomings of the technique. Environmental factors can be listed as distance from user, object size, density of objects in the area, obstacles between the user and the target [26].

- Manipulation

Manipulation is one of the most important forms of interaction that can affect the presence in the VE quite highly. Interaction is realistic when the user grabs and moves a virtual object as he would grab and move objects in the real world. Taking into account that manipulation is composed of canonical tasks as mentioned before, the level of reality depends on these canonical tasks. Therefore, the issues, such as distance and size of the objects, that are valid for selection directly effects implementation of manipulation. Even though the virtual hand technique is intuitive, only those objects that are within the area of reach can be picked up, significantly limiting the technique's applicability in immersive CAVE-like virtual environments [30]. However, this will not be an issue for this design as mapping from the screen space of the mouse will be the plane of motion for the virtual hand attached to the cursor.

The precise manipulation can be difficult without the visible cues and constraints. In rotation with virtual sphere, even though the interaction is carried on by a 2D device, the sphere around the object is quite helpful to keep user in the rotation axes and act as a guide, since it is presumed that during direct manipulation, the mapping between the interaction device and the object might confuse the user.

2.3. 3D Interaction and Desktop Computing

Manipulation of the objects is a very important part of interaction in real world. Similarly, manipulation is very important in virtual environments (VE) as such to interact with the entities of VE. The quality of interaction depends on both the success of interaction techniques and the authenticity of the interaction form. Also, robustness, efficiency and fast reflective feedback are key requirements for interactive environments. When we consider the definition of 3D interaction to be "Human–computer interaction in which the user's tasks are performed directly in a 3D spatial context" as described by Froehlich, Kitamura and Bowman [19], 3D interaction with DVE is valid as long as the interaction happens in 3D spatial context. Therefore, even though mouse itself is a 2D input device, it can interact in 3D environment and function as a part of 3D interaction with an effective mapping to the universal tasks. Hence, interaction in 3D DVE should supply the quality level expected by means of robustness, efficiency and feedback for success.

2.3.1. 3D interaction via 2D input device: "the problem"

Simple 2D mouse is considered as a selection and navigation device as it allows the user to select and navigate (translate) the objects to a different location in virtual 3D space. In selection tasks in VE, the position of the cursor on the screen-space is transformed into a three dimensional ray (like a 2D to 3D reverse projection mapping) to be able to select any object along this ray as in the application of ray-casting technique in Schafer and Bowman's work to investigate spatial collaboration [20]. During translation tasks, mouse takes the role of direct-hand manipulation by means of translational tasks.

Since manipulating physical objects with the hand is intuitive for humans, direct-hand manipulation claimed to be the most natural technique. Implementation of the technique is based on attachment of a virtual hand to an object in immersive or semi-immersive world with an input device like gloves or wands. In this study, mouse pointer is represented as the virtual hand and performs direct hand translation for translation tasks (for object navigation purposes). Head-based tracking fills the rotation aspect of manipulation for direct hand manipulation method as well as the third dimension component of the translation task in collaboration with mouse input.

In desktop environment, the mouse has proven to be an excellent and easy to use input device as it is a standard two-dimensional pointing device (2D input device by nature). By nature, it is capable of translating the actions in 2D space, whereas there is a difficulty in aligning the 3rd dimension on mouse movement. Providing a natural mapping from 2D input to 3D position is a difficult problem, usually faced by DVE interface developers. The simplest solution suggested by Strauss in SIGGRAPH course notes is to provide handles or widgets for explicit 3-axis manipulation and has been adopted by many conventional systems [44]. Therefore, while using mouse as a direct input device for translation / rotation tasks in 3D DVE, we need to take a decision by means of mapping mouse to either x-y plane or x-z plane. Otherwise the conflict in dimensions will create difficulty in usage by means of differentiating the dimension to which the object is translated. Same problem was also addressed by Oh and Stuerzlinger in their study about moving objects with 2D input devices in desktop virtual environments [43]. Their main idea was finding a movement surface and mapping the mouse movement onto the movement on that particular surface by providing handles for all three dimensions and aligning the plane of motion with the user's intentions. Their approach was seeking for a visually smooth technique to prevent limitations in axis-aligned motions or predefined object behaviors as investigated in previous studies by means of smartly finding a snapping surface for the object in motion, and mostly snapping the object to the foremost surface behind the moving object.

Dimension mapping is also the same problem encountered while playing flight simulations on desktop computers without a joystick but with a mouse. Joystick easily solves the problem of dimension mapping not because it is capable of moving or creating movement in three dimensions but because the natural object of flying is a joystick as a wheel of a car, which naturally creates the capability from already known phenomena. Bowman et al. claims that the connection between the already known phenomena and created interface is where the success of interaction lies even though the device is not completely representing the mapping; therefore, they recommend special cut solutions to scenarios based on the environment and scope of interaction [17]. In this study, we choose the effective plane for 2D mouse input to be x-z plane mapping the virtual hand in the environment to the hand holding the mouse, since that is the plane of the surface that can intuitively map to the horizontal plane of the 3D virtual environment as the obvious plane of motion for the user.

2.3.2. Natural interaction for desktop computing

Natural interaction techniques are considered to perform interactions in a way similar to the real world by mimicking actions performed in the real world for similar tasks. This is still to be experimented; however, McMahan et al.'s empirical study shows that the success in interaction and feeling of presence are not always correlating [21]. On contrary, the comparison via their test-bed reveals that the performance of natural interaction was not as successful as expected even though it was more fun. Besides, natural interaction should also consider, the authenticity of the interaction also depends on the real world phenomena by means of providing the environmental behaviours of the real world such as occlusion, collision and physics behaviour. The correct physics behaviour of the virtual environment providing real collisions and Newtonian physics mimicking the real world would create more realistic environment also increasing the level of presence for the user, yet to be experimented.

Similarly, this test-bed identifies whether the proposed multimodal interface is capable of performing manipulation tasks in 3D DVEs, addressing whether head tracking is intuitive and efficient enough to be used as an interface for navigating a 3D object in 3D DVE with the integration of 2D mouse and head tracking as well as 3D object rotation. While there are experiments evaluating the efficiency and usability of 3D rotation by free-space 3D input devices [22], there is room to investigate the capability of head rotation as a free-space 3D input device.

2.3.3. Immersion

Immersion in video games arguably refers to how much the user is drawn into the game's world and how much they feel as though they are actually in the gaming environment surrounded by the stimuli of the game. In 3D VEs, immersion mostly depends on the feeling of presence as well as the human factors and intuitive interaction within the VE [17]. VEs are evaluated for immersion and accepted as immersive or semi-immersive environments depending on the level of immersion whereas 3D DVEs are mostly labelled as non-immersive. However, if the feeling of presence is perceived to be the origin of immersion, novel ways should be pursued to increase this feeling within the DVEs to support immersion in gaming and 3D interactive applications for desktop users. It should also be noted that realistic interaction with the environment might enhance the level of immersion in many ways when the behaviours of the objects in the environment are based on real world concepts of motion and behaviour by employing almost realistic Newtonian physics and effects supporting the look-and-feel of any environmental element.

This level of research into immersion is not the main pursuit of this study; however, as future work the further direction of the study can focus on immersion factors for DVEs utilizing the findings from this test-bed.

2.3.4. Game-viewpoint applications

Viewpoint simply refers to the camera angle in a video game. Depending on the genre of the game, this viewpoint could be attached to the first person point of view or third person point of view. In a first person point of view, the player sees the game world from the viewpoint of the avatar in the game. The visible part of the world is as large as the field of view of the viewpoint camera. Obstacles in the environment, corners of the walls are the boundaries in the virtual environment. Mostly these little problems are solved with a key mapping from the keyboard that brings in the capability to peek around corners or change the camera angle. Nevertheless, there have been numerous studies to utilize head orientation to enhance the immersion by providing natural solutions for any boundaries for viewpoint.

Teather and Stuerzlinger conducted a research by mapping head translation to the game-viewpoint of the player controlling the viewpoint via the translation of the head along the x-axis [5]. Similar to game-viewpoint control approach, a technique called orbital viewing technique was implemented by Koller et al in 1996 by using a Head Mounted Display (HMD) [37]. Rather than the name game-viewpoint, it was named as orbital viewing technique since it was providing the capability to look around the object. Head rotations were mapped to get an orbital view around the point of interest while user was wearing a HMD providing a fully surrounding VE as in Figure 2.4. Therefore, one should consider that such mappings in DVE have view point limitations to provide usable rotation versus viewpoint mappings and maintain particular views for long periods of time. Figure 2.3 below is sourced from [33].

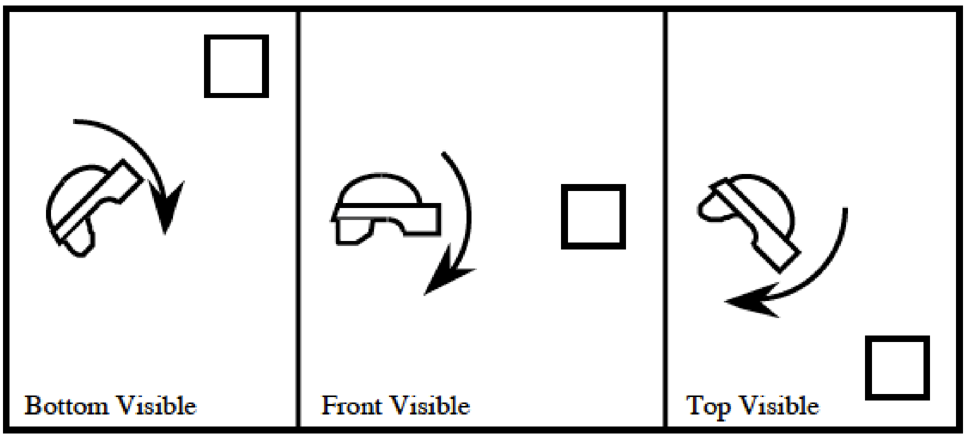

Figure 2. 4: Orbital-viewing approach via Head Mounted Display (HMD).

A following study tried replacing head rotation in this mapping with hand orientation in CAVE and reported that this change in orbital viewing technique improved the efficiency. Even though mapping the viewpoint to head rotation is more realistic and seems more natural, interaction was proven to be not as intuitive and efficient as expected due to the environmental factors in VE. Besides, DVEs already have view limitations. Therefore, rather than coupling head motion directly with the object or controller, a trigger to activate and deactivate the tracker could be more useful in DVE.

2.4. Interaction and Video Games

The approaches for 3D video game interaction can be classified under three groups [3]. The first group is the interaction via traditional devices: keyboard, mouse, joystick, and game controllers. The second group is based on the interaction via devices simulating the real world (natural techniques): steering wheels, musical instruments, weapons and tennis rackets etc. The third group is based on spatial tracking where users interact in and control elements of the 3D gaming world with their bodies, as in recent day's game consoles. Other than these three that LaViola addressed [3], tangible interaction should be considered as another group since these three groups are not enough to represent interaction in mobile gaming or gaming via tangible surfaces. The second and third groups are almost only appealing to game consoles. Desktop gamers do not really benefit from technological improvements within the interaction field enough except the interface improvements introduced by accessibility studies. Most of the interaction in desktop gaming is based on mouse and keyboard pair, and there is not much novelty despite the novel game mechanics and interactions that are introduced to the players via console games.

2.4.1. Interaction, Usability and Games

Games are interactive environments, and a game can arguably be defined as an interactive experience in which the player interacts with the game/game environment. This definition puts interaction in the center of the game experience since an input is required to play an interactive game. Ease of learning (*learnability*), *efficiency* and simplicity in controls with *effectiveness* are quite valuable aspects among usability goals of gaming. Therefore, this test-bed will nicely lead to a game-like test-bed as a next step towards usability analysis in games.

2.5. Multimodal Interaction

Multimodal systems are defined as systems that process two or more combined user input modes— such as speech, pen, touch, manual gestures, gaze, and head and body movements, mouse, keyboard etc.— in a coordinated manner with multimedia system output [2]. The development of novel multimodal systems has been enabled by various input and output technologies currently becoming available, including new devices and improvements in recognition-based technologies. Therefore, their design, implementation, advantages, shortcomings and usability status based on the generic usability goals are yet to be experimented, with the growing interest in multimodal interaction field of human computer interaction.

2.5.1. Input modes for multimodal interaction

Iterating from their definition, multimodal interfaces combine one or more user input mode in a coordinated manner. Therefore, a system with multiple keys will not be a multimodal system whereas one supporting both keyboard and mouse input will be. Similarly, a system that responds to a facial expression and hand gestures by only using camera input will not be a multimodal interface as the system is supporting only one type of input. Furthermore, in order to understand multimodality better, one should understand the user input modes.

User input modes can be classified in two modes based on the functioning style of the input device creating the input mode; one being active input mode, other passive input mode [42]. Active input mode and active input devices are based on the intentional action of the user as an explicit command, whereas passive input mode refers to the recognition of naturally occurring behaviour or actions with a passive monitoring mode. When these two modes are combined to create a multimodal interface, the creation is called as blended multimodal interface.

As an output of this study, integration of 2D mouse and head tracking creates a blended multimodal interface suggesting a novel way to interact with DVEs.

2.5.2. Application of multimodal interfaces

The applications of multimodal interfaces vary widely including but not limited to ambient applications, art/deco environments, mobile and wearable devices, public and private spaces, virtual environments etc. Users with disabilities are another focus of affective multimodal user interfaces. Especially in virtual reality environments, multimodal human computer interaction (MMHCI) is very attractive as it helps disambiguate the communication between the user and the virtual environment or the elements of virtual environment [2]. Speech and gesture recognition are highly used modalities for VE interaction purposes.

Suitable hardware towards multimodal interfaces for gaming purposes has already been manufactured and brought to players by game consoles. By combining controller input, gestural input and speech input, game consoles utilize multimodal interaction, and bring MMHCI to the hands of gamers. Kinect is a widely used device developed by Microsoft to be used along with Xbox 360 and its successors; moreover, it already has built in libraries for speech and gesture recognition to enable developers pursue new ways of interaction for their games. Any documentation could easily be reached via Kinect support website. Nevertheless, the playability and usability factors of these devices have not been fully experimented enough in scholarly studies yet leaving this field open to investigations.

3. COMPUTER VISION and HCI

In recent years, research in visual tracking gained acceleration leading to face recognition, head and gesture tracking etc. The progress in computer vision has been enabling human computer interaction studies to utilize this capacity for the benefit of HCI. Eye tracking has already been used as an evaluation tool by usability researchers for a long time to enable usable and effective design of interfaces as well as linking eye tracking data to cognitive processes to analyze what eye movement might reveal to help the research in psychology and physiology [1]. Body/gesture/gaze tracking and face recognition took their place as emerging technologies enhancing multimodal human computer interaction [2]. Even though it takes quite long time for many research results to reach the public domain, 3D spatial interaction or particularly gestural interfaces reached public quite fast with the help of video game technology, game consoles and video games. Moreover, 3D spatial interaction has addressed solutions to the usability problems originated from controller complexity in games that require maintaining high levels of expression and interaction as LaViola already addressed [3]. However, these solutions are yet to be identified by relative research and analysis.

3.1. Tracker as a 3D Input Device

Input devices are categorized depending on many different characteristics such as input type, physical interaction requirements, intention of usage etc. 2D mouse is widely used in desktop environments and can be thought as a purely active input device as it requires physical interaction to generate input information to the computer. Trackers are quite common in immersive VEs and called as monitoring input devices as they continually record and evaluate position/rotation information of the object tracked. They are also named as purely passive input devices as they don't require any physical action or interaction to function [18][2].

[2] Interested reader can find more information about input devices in 3D User Interfaces book by Bowman et al, pg. 89.

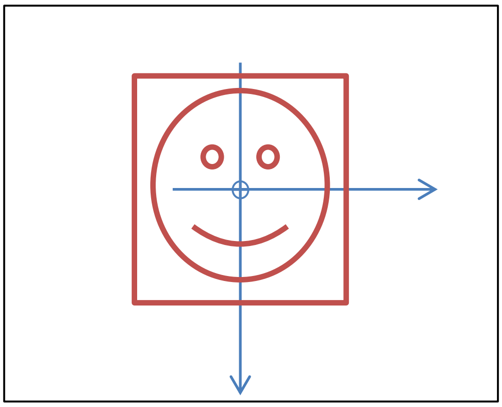

Figure 3. 1: Zero position for tracker can be assumed as in the figure above.

In real world scenarios, a reference head position is obtained as zero position to eliminate the issues possibly originating from postures of individuals. A home zone is located around this reference posture to prevent unnecessary rotations originating from small motions around the starting posture. Roll, yaw and pitch rotations can be evaluated relative to the reference position of the head as shown in the figure below.

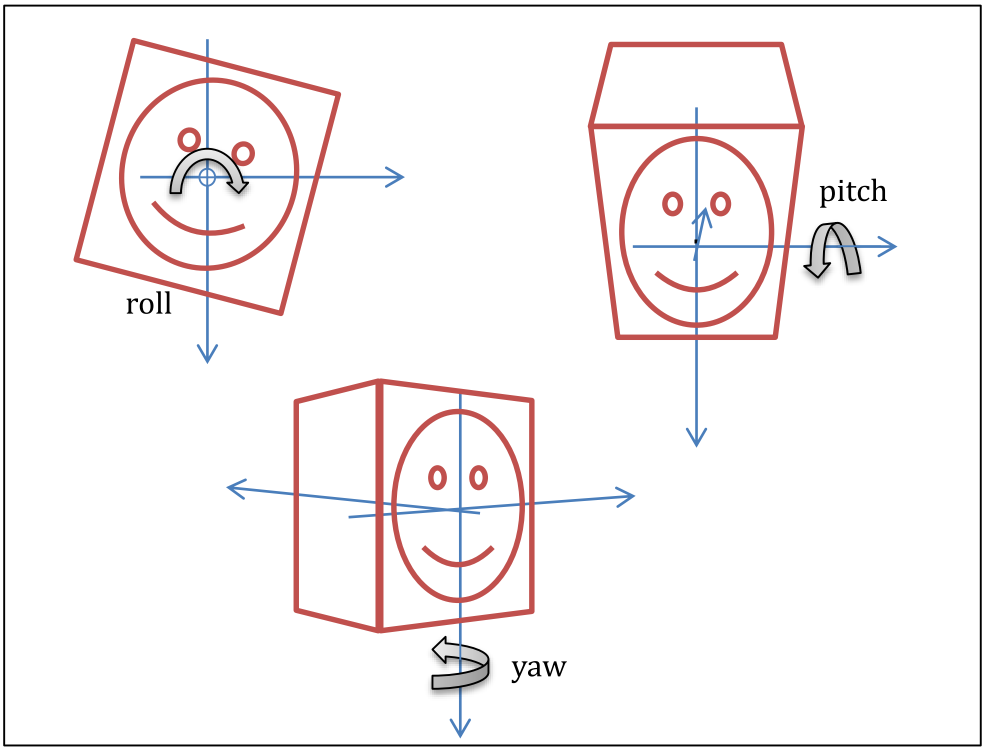

Figure 3. 2: Rotations of head: Roll-Pitch-Yaw.

3.2. Head Tracking in Previous Studies

Robustness, efficiency and fast reflective feedback are key requirements for interactive environments. Creating efficient versions of hands free interfaces has been an aim of human computer interaction researchers for both investigations towards novel ways and the benefit of disabled or elder people to enhance their interaction capability. With the improvement in the capabilities of desktop computers and better accessibility for public to the products/technology, human computer interaction, multimodal interfaces and intelligent systems can find more common ground to reach to individuals. Ubiquity of handheld gadgets with capabilities only bounded by imagination is quite normal for this new era. Moreover, tracking and recognition are important research areas empowering this technology peek. The capability in tracking also increases correlated with the progress in computing power and the capability of the systems.

There are several studies investigating head tracking approaches and head pose estimation as a challenge for computer vision. Every single head tracking system has their own detection approach, style of data processing and evaluation; therefore, performance results and suitability to environmental factors of these systems vary one another. Some of these works use sensors to track the head orientation as in [5], [6]; some use stereo camera based tracking as in [6], [7], [8] for positioning head; or, some use monocular head pose estimation as in [9], [10].

Tracking has been used in several researches to substitute the interaction device (mostly mouse). "*Camera Mouse*" project of Betke et al uses tracking to substitute mouse input by tracking body features to increase the computer interaction capability of disabled people. The output of this research is open to public use, especially to help people with severe disabilities [12]. Also, "Nouse", which stands for *Nose as a Mouse,* was another research that has considered the ubiquity of cameras and the decrease of camera cost as a good chance of enhancing the perceptual power of the computer and providing solution for an intelligent hands-free input device [7]. Not much later than that, a head tracking driven virtual computer mouse system, called "hMouse" was also developed by Yon Fu and Thomas S. Huang as a hands-free user interface [9] as a sequel of many other researches with same intention of navigating the cursor or triggering mouse click with the movement of eyes, nose and/or face. However, there are not many researches evaluating head tracking as a second or collaborative interaction device (referring to the collaboration between two input interfaces: head tracking and 2D mouse) for desktop computing rather than a substitution of mouse.

Some other researchers investigated head motion relative to hand gestures as aid for gestural interaction, either to validate the hand gesture or to lead the motion of the full body. One of these works uses head orientation as an additional feature to evaluate whether the pointing target is the intended target, and declares that significant gains in precision are obtained in pointing gestures recognition [6]. In [6], researchers tracked head rotation by using magnetic sensors attached to the user's head. In [13], Ashdown et al used head tracker via three cams to make the mouse pointer jump between monitors to facilitate mouse usage for window management and switching between applications while using multiple monitors and prevent losing the pointer when switching monitors.

The use of head tracking to improve experience in games and to enhance the view of the player has been investigated as well. By coupling the virtual camera to the player's head position for the game viewpoint control, Teather and Stuerzlinger used exaggerated head movements rather than rotations as an input to change the viewpoint of the user [5]. After their evaluation of different exaggeration levels applied to interaction to accomplish object movement tasks, they conclude that even though no significant difference is present for speed or accuracy by level of exaggeration, based on the questionnaires, participants are happy to have some level of exaggeration. Figure 3.3 below is sourced from [5].

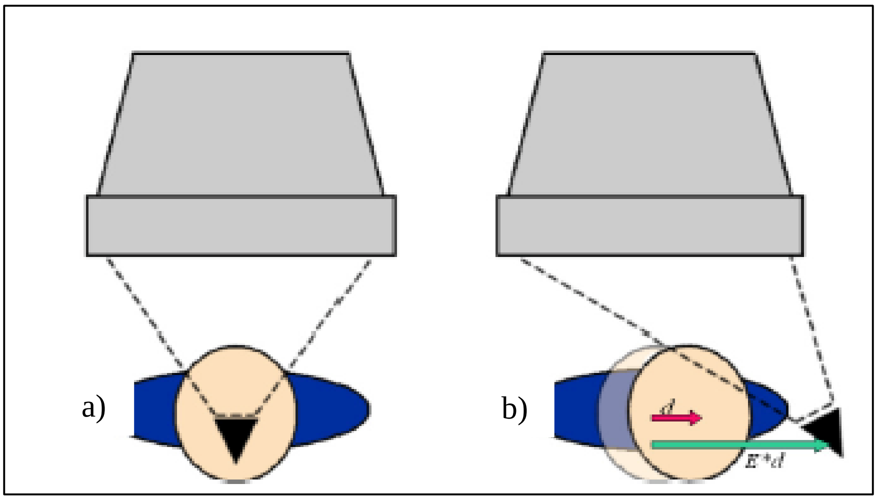

Figure 3. 3: Exaggerated head motions, a) Base condition, b) Relative shift.

LaViola et al used head tracking to find out the effects of enhancing vision for non-isomorphic rotational tasks [14]. In another work, head tracking is used to reduce the keyboard input complexity by transferring some tasks (like peeking around corners) from a button to head movement [3]. Besides, open source community developed an Api for computer games to enable free environment viewing (like looking around from within a vehicle) with the help of head tracking, and this Api is ported to several games with the efforts of open source community [15], [16]. However, no empirical or usability analysis are conducted to find out the usability/performance of this Api within the game context. It could be a very nice feature for some games whereas could be hindering the player performance in others; yet to be investigated.

When head gestures of the players are informally monitored during a task, it is observed that head leads the action within the game by following the action path. This natural movement suggests head directed movement to be a natural self-regulated interaction interface to organically channel player's habit of motion into 3D VE. Head already has this capability to rotate in 3D space without creating any conflict between the dimensions of interaction that might be present in the environment. It is big enough to track and movement is amplified enough to follow as long as the action does not require small precisions. Besides, the movement of head is a natural behaviour that assists most of the tasks in real world even without awareness of the individual.

3.3. Face Recognition and Immersion

Face recognition has been being investigated for several purposes in computer vision field. Recently, there are studies to implement face recognition to create realistic facial expression for avatars; however, they are not mature enough yet. Facial expressions are assumed to increase the immersion in games, also supporting the presence in the virtual environment by creating less synthetic looks reflecting the expression of the real player. Progress in this field of computer vision will certainly change the feeling of reality both in games and game-like environments. Especially for collaborative environments, facial expression may increase the feeling of presence and feedback cues as an important aspect of communication; yet to be experimented. Besides, detailed face recognition can help social studies' investigations as well by means of analysing individuals' responses/ reactions in certain conditions and deriving investigations via prescribed approaches capable of instantly analysing and grouping rather than simple observation.

4. RESEARCH DIRECTION

With this work, head tracking is proposed as a 3D interaction input device for 3D rotational tasks working in combination with 2D mouse input for translation tasks within a suitable test-bed to understand intuitiveness and comfort of such integration. Similar approach was experimented by Chen [4] via combining 2D mouse input with virtual rotational controllers to enable 3D rotation by using 2D control devices. Repeating alike test of Chen's to find out the usefulness of head tracking and mouse integration can be helpful by means of identifying another interaction opportunity for modelling software users as a domain specific interface. Therefore, an empirical study for rotational tasks comparing the Virtual Sphere techniques with 3D interaction with head-based input can be performed as a side study or in future work. This can fit in the task and domain specific interaction design approach as a generic interface to replace virtual sphere.

In this study, a test-bed for the integration of two input devices is defined to investigate the suitability of such integration in a basic interaction evaluation approach. Furthermore, this test-bed will lead to a more game-like test-bed environment to compare and contrast the findings between two similar test-beds. Hence, a repeatable and reusable test-bed framework would be constructed after a series of experiments.

4.1. Designing 3D Interaction

As a start to design 3D interaction, design guidelines explicitly advices to use existing manipulation techniques unless a good amount of benefit might be derived from designing a new one which is quite application specific [18]. The second step is finding the appropriate interaction technique for the input device. The capabilities of the device or the intuitiveness and the manipulation precision can affect the success of input device-interaction technique couple. Non-isomorphic techniques are said to be useful and intuitive as well as reducing the wasted motion (clutching) via an interaction metaphor deviated from the reality of real-world-metaphors. Virtual hand-based techniques are advised for manipulation tasks even though ray-casting and other pointing techniques are said to be providing notably better selection performance.

Based on the recommendations of Bowman et al. as stated above, direct manipulation— virtual hand interpretation— has been chosen to be the existing manipulation technique in use; moreover, virtual hand for both 2D mouse and tracker are chosen as the appropriate interaction technique for the input devices. Hence, 3DOF head tracking for rotation, 2DOF with 2D mouse for translation in 2D (x-z plane) and 1DOF with head tracking for translation in one dimension (y-axis/vertical axis) achieves to create interaction in 6DOF noting that precision in this interface depends on the capability of head tracking Api.

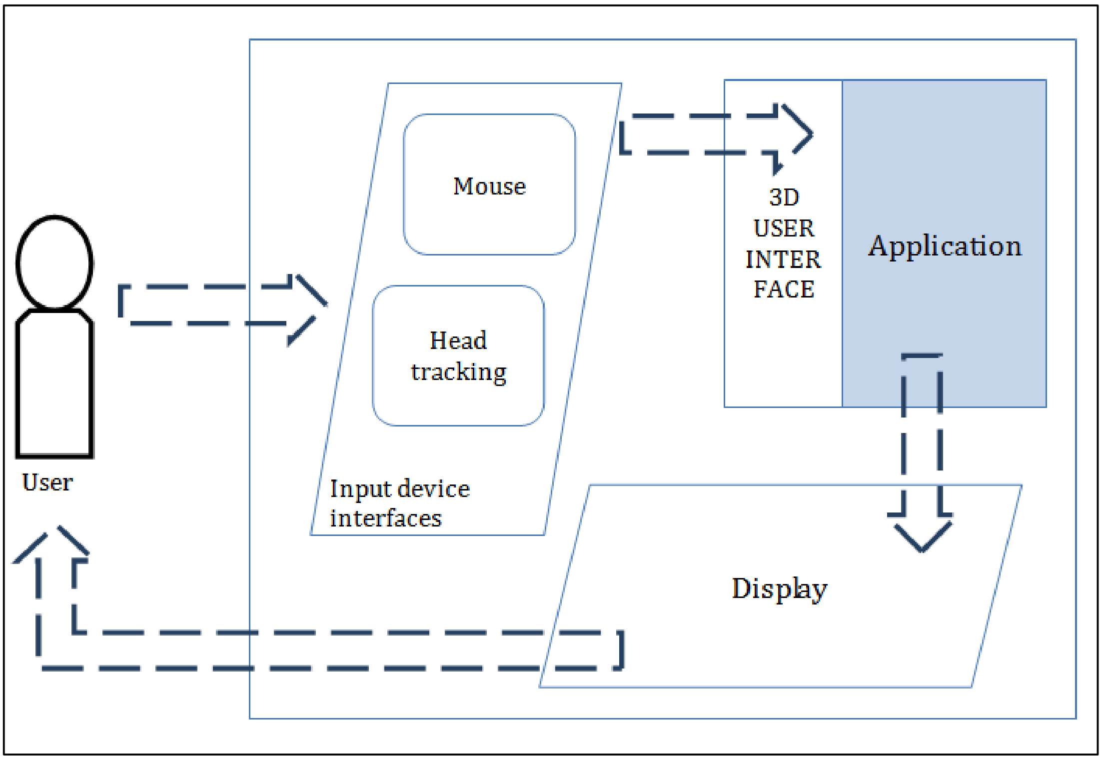

Figure 4. 1: Block diagram of the system.

Figure 4. 1 displays the big picture of the system as a whole open to interaction by the user. Multimodal input obtained by input device interfaces is evaluated by the 3D user interface, and via this interface user is able to interact with the application which is a 3D VE. The feedbacks/outputs of this interaction is displayed to the user.

Tasks and the composition of interaction that creates the test-bed are explained in detail in Section 5.2. Interaction.

5. TESTBED and EXPERIMENTS

In order to carry out the tasks and evaluate the efficiency of interaction, simple sets of objectives are planned to perform in a simple DVE. Simplicity is important by means of eliminating negative environmental effects on performance and find out the success of integration in the creation of this multimodal interface. Before and after the tasks are set, users to be asked to fill in questionnaires to be evaluated as a measure of usability by means of the feel of interaction. The questionnaires are explained in relevant sub-section. Use-case diagram for the test-bed environment is visualized with Figure 5.1 below.

Following parts in this section will explain test environment, interaction tasks and how canonical tasks are performed, experiment sets, task creation and complexity of tasks.

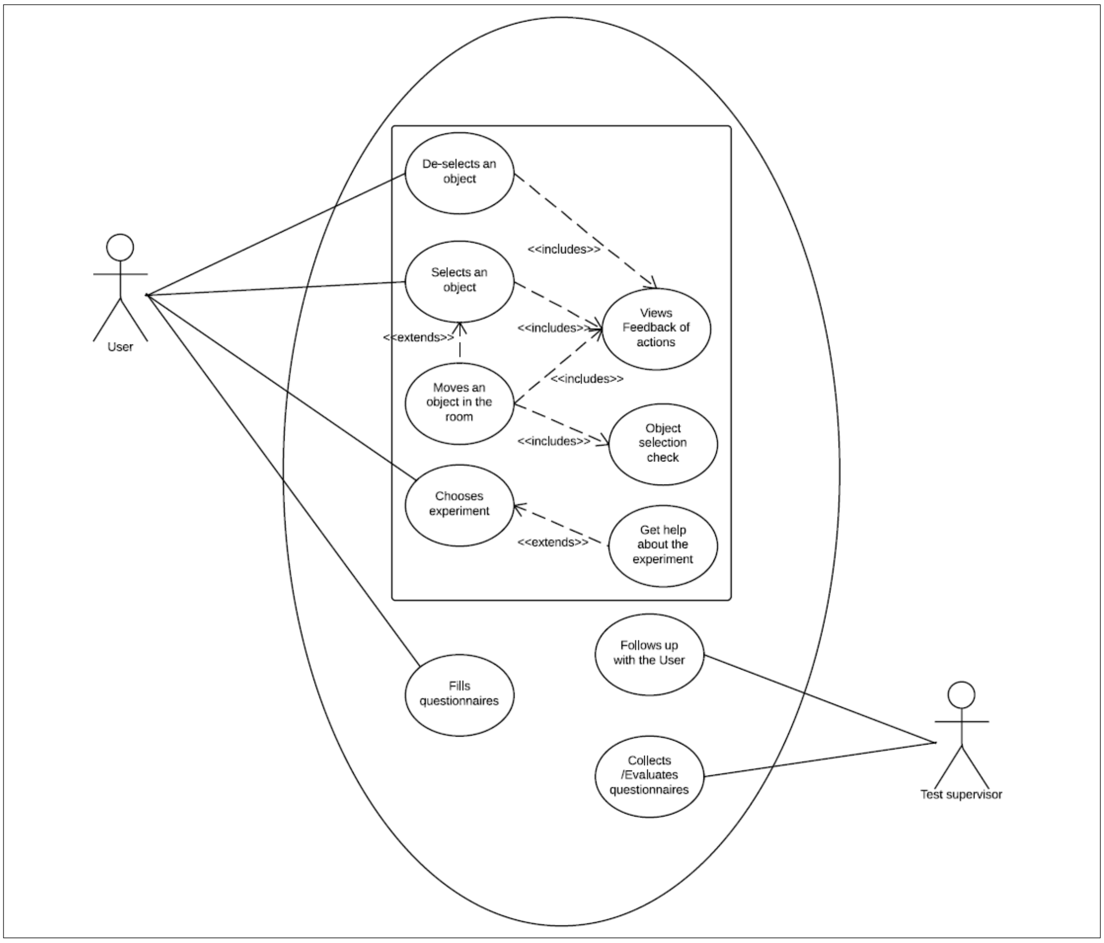

Figure 5. 1: Use-case diagram is useful to visualise the user behaviour within the test-bed. See Appendix for a bigger version of this Use-case Diagram.

5.1. Test Environment

Physical test environment is composed of a standard desktop computer with moderate graphics capability, a standard mouse and a standard plug and play camera mounted on the display.

Physical test equipment is composed of a standard PC with minimum 2GB Ram and Nvidia Graphics Card supporting DirectX10-11. A custom software that is developed in C++ using DirectX Api for graphics and a non-commercial application interface (API) for head tracking via a standard camera mounted to the screen contains the task sets for the test-bed to utilize and evaluate the designed interaction. Single camera is used for tracking purposes, as single camera is more common amongst end users. Performance of interaction via head tracking obviously is affected by the performance of the selected API. Since developing own API that has high performance capability would require extensive amount of work, and that is not the particular aim of this study, there are not many options other than using FaceApi [15], [16] for monocular tracking and Watson library for stereo tracking [28]. Therefore, any result that is obtained from this study is dependent on the precision of the head tracking API. This fact might create a performance metric for future APIs by means of comparing their performance on introduced tasks.

5.2. Interaction

5.2.1. Tasks of 2D Mouse and Head-Tracking Based Input

In order to eliminate the possible difficulties in performing canonical tasks of manipulation, left button enables selection/de-selection, and any selected object is attached to the virtual hand and navigated with the motion of mouse. 2D mouse maps onto two-dimensional space in x-z plane; therefore, navigation task in third dimension is carried out by head motion. Vertical shifting of the head (This might look similar to pitch, but includes shifting the head vertically.) supplies the input for vertical translation when the object is in navigation. Right button of the mouse enables rotation of the object in 3D space, and head tracking controls rotation via a non-isomorphic mapping. Non-isomorphic mappings let users interact with virtual world objects at an amplified scale, in contrast to isomorphic mappings (i.e., one-to-one mappings) that maintain a direct correspondence with the physical and virtual worlds. Roll, pitch and yaw rotation of the object is performed with the roll, pitch and yaw of user's head. When two buttons are pressed at the same time, selection overpowers manipulation to eliminate possible difficulties. Also, the ease of interaction depends on the capability of tracking as the precision in rotation is subject to factors like precision in tracking, and any kind of latency.

5.2.2. How Selection Works?

Selection is to be performed via mouse ray-casting with a projection of 3D onto 2D as HOMER [17]. When the virtual hand is attached to the object via mouse button, user will be capable of moving the object in VE.

When the object to be selected is pointed by the mouse cursor (which is a virtual hand), mouse button is used to confirm the selection. Selected object is indicated visually so that the appropriate feedback is provided to reflect selection is successful. After the object is selected with left mouse button, it is attached to the virtual hand and can be translated (moved) around with either mouse or head motion until the button is pressed again and the object is de-selected. De-selection of the object clears the visual indication of selection and object is no longer attached to the virtual hand. A separate comparison test can also be constructed to identify the best button usage and selection/de-selection feedback cues.

5.2.3. How Navigation Works?

Moving an object is either possible in 2D via the motion of the mouse or in 3D via the integration of mouse and head tracking based input. Navigation mostly refers to the translational tasks in this study as describing the motion of the object from a position to another.

Translation in x-z plane is performed as a non-linear mapping of mouse motion on x-z plane. Vertical motion of mouse behaves as the extension of arm in Go-Go enabling the translation of the object along the depth dimension.

Translation in vertical axis is based on the relative vertical shift of head. Head tracking based input maps to the vertical motion of the object and is in action as long as the object is selected and attached to the virtual hand. Figure 5.2, below displays the look-and-feel of the system based on user motion updating the screen state.

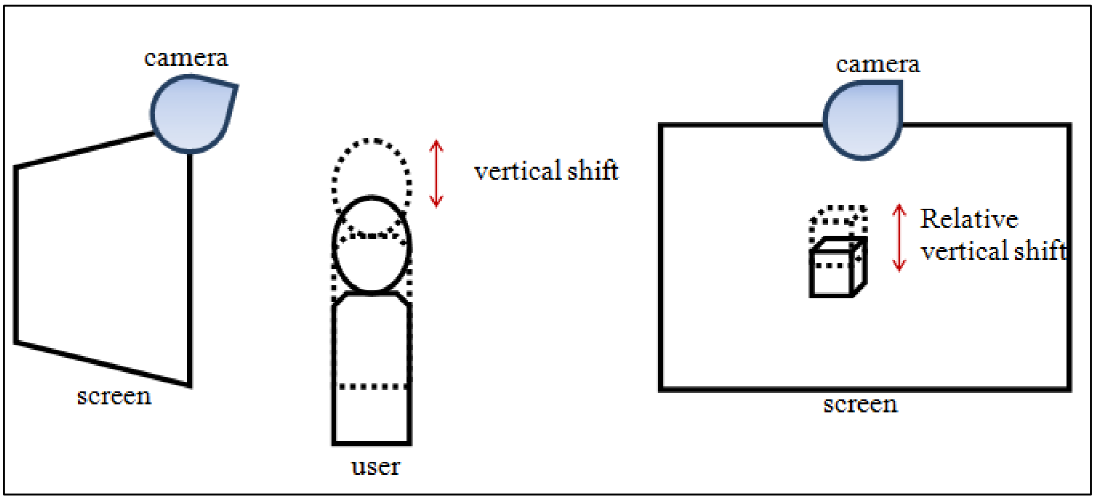

Figure 5. 2: Visualisation of head shifting for navigation of the object in vertical dimension.

The vertical shift of user's head is captured by head tracking Api, translated into the system and evaluated for a relative vertical shift for the position change of the selected object.

5.2.4. How Rotation Works?

Turning one's head is quite natural; however, turning too far in a DVE or in a game will narrow the field of vision. Therefore, linear mapping for rotations is not useful to create the interaction via the head orientation. This is why exaggeration is required to map head rotations to the object rotation (See Figure 3.2. Rotations of head: roll, pitch and yaw for reference to head rotations.). Since the idea of exaggeration is explored in a few 3D interaction techniques and found useful as reported by studies [29], [30], some level of exaggeration will be used in this study. Also, the study conducted by Teather and Stuerzlinger states that a modest amount of exaggeration is preferable by the subjects even though no significant differences were found for speed and accuracy when the exaggeration levels are compared [5]. As far as the head tracking API supports, a "home" zone will be located around the head so that unintentional motion or rotation of the head will not be evaluated as a manipulation by the tracker.

Rotation capability is activated via the right mouse button, only when an object is in selection. When rotation is activated, head shift is no longer evaluated for translation of the object as the rotation of the head is mapped to the rotation of the selected object.

5.3. Experiments

Two experiment sets are defined. First experiment is to identify the capability of mouse to perform positioning tasks—selection and translation tasks separately and in conjunction with head tracking based input. Second set includes rotational tasks as well as selection and navigation. Task order intends to reflect the difficulty order; starting with the easiest task which only requires the selection with mouse, ending with combination of mouse and head tracking for selection, navigation and rotation consecutively or simultaneously.

5.3.1. First Experiment Set

First experiment set does not contain any rotational task, as it only aims to evaluate the navigation tasks carried out by 2D mouse input and a multimodal input of 2D mouse integrated with 3D head tracking. For an easier approach to evaluation, this experiment set is also split into four experiment sets, two of which is for navigation via 2D mouse and the other two to be navigated by multimodal input. In order to eliminate relative difficulties brought into interaction by 2D mouse, first group of tests (first eight tests) only include interaction via mouse as a base. Second group of test (last eight tests) are performed for multimodal input to perform one navigation task.

- Navigation (Translation) test for 2D mouse input

Subjects are given an object to be translated to another specific location and left there. Navigation mapping of 2D mouse is horizontal and depth (x- z) dimensions. When the mouse is moved vertically up, the selected object moves away from the user and vice versa.

Below is a generic task list for translation with mouse; listed in increasing difficulty. First part of the tests helps creating familiarization for the test environment and evaluating the capability in performing the tasks by using just mouse; besides, addresses the difficulties originated from the primitive interaction method when multimodality is in test.

i. Object selection

ii. Object selection and navigation in two dimension; x and z (there is only one object in the scene); leaving the object at the designated place (there should not be any other object around the designated place)

iii. Object selection among more than one objects and navigation in two dimension; x and z; leaving the object at the designated place (there should not be any other object around the designated place)

iv. Object selection among more than one objects and navigation in two dimension; x and z; leaving the object at the designated place among other objects

- Navigation (Translation) test for multimodal input (2D mouse input integrated with 3D head tracking)

Head tracking is integrated to map the third dimension that mouse is not capable of covering for navigation tasks. Vertical shifting of the head (this might look similar to pitch, but includes shifting the head vertically) supplies the motion on y-axis (vertical dimension), as object will move up if head is shifted up (nose is almost pointing above the natural look horizon) and vice versa (See Figure 5.2). Vertical shift of the head is tracked by the head tracking Api and applied to the object relatively for the motion of the object in vertical axis.

Mouse is responsible for selection as in the other test and navigation in x-z plane. As future work, this test can be extended to compare the integration of mouse and head tracking to the general integration of mouse and keyboard under same task set resulting in the comparison of interaction device integration complexity based on accuracy, usability and intuitiveness.

For the integration of mouse and head tracking to perform navigation, below is the generic task list:

i. Object selection

ii. Object selection and navigation in three dimension; x, y and z (there is only one object in the scene); leaving the object at the designated place (there should not be any other object around the designated place)

iii. Object selection among more than one objects and navigation in three dimension; x, y and z; leaving the object at the designated place (there should not be any other object around the designated place)

iv. Object selection among more than one objects and navigation in three dimension; x, y and z; leaving the object at the designated place among other objects

5.3.2. Second Experiment Set

Second experiment set is separated from the rest just because the nature of interaction is more complicated as rotational tasks are included as well. 2D mouse is used for selection and translation tasks, and head tracking based input is mapped to rotational tasks and navigation in third dimension. After an object is selected, any of the tasks, translation and rotation, can be performed by user disregarding the order.

To enable a smooth transition combining navigational tasks with rotational tasks, splitting this experiment set into two groups would be helpful. Therefore, rotation via head orientation is accomplished before moving to the combined tasks of manipulation.

- Rotation via head orientation

  i. Object selection (separate object) via mouse, changing the orientation of the object to match a proposed model as in Chen's experiment with virtual sphere method [5].

- Navigation and rotation via mouse and head tracking input

Below is a task list listed in increasing difficulty for rotation via head tracking based input combined with the translation via 2D mouse input:

i. Object selection (separate object), navigation of the selected object among other objects, rotating the object, leaving the object at a designated place after changing the orientation

ii. Object selection (separate object), navigation of the selected object among other objects, rotating the object, leaving the object at a designated place among other objects after changing the orientation

iii. Object selection among more than one objects, navigation of the selected object among other objects, rotating the object, leaving the object at a designated place among other objects after changing the orientation

5.4. On Task Creation and Difficulty of Tasks

Tasks should be easy to understand and simple enough to perform. A clear instruction for every single task should be displayed on the screen.

Different levels in difficulty of tasks should be declared to examine capability level of the interface. It is essential to identify the possible difficulties in performing the tasks and what makes a task difficult. The difficulty can be grouped depending on the number of objects in the environment, the number of objects that is possible to interact, the number of actions required to accomplish a task, the expected accuracy etc. This is quite similar to car parking example; as parking in between two other cars seems to be more difficult than parking at a certain place without any other cars around. Anxiety of the user is another parameter among the difficulty reasons that can be addressed within the qualitative part of the research, the questionnaire. Therefore, the first task of each task list is simply interacting with just one object without others in the scene that acts as a baseline similar to "home zone" for each user to eliminate positive/negative effects originated from personal familiarity.

Also, the environmental effects on performance might affect the difficulty perception of tasks such as smaller objects versus bigger objects for selection, narrow target space etc. Most of these environmental effects examined as factors influencing performance both in Bowman's [31] and Poupyrev's [27] experiments. Bowman especially defined them as outside factors influencing performance in immersive VEs in [26]. This study assumes it is quite similar for non-immersive VEs leaving it open for a future study to find out the effects of outside factors in non-immersive 3D VEs.

5.4.1. Tasks Generated from First Experiment Set

Simple docking tasks are planned iterating from the generic task list explained above in Section 5.3.1. Flow of a simple task is explained with the diagram below. Feedback is especially listed within the task flow since main information cue for the user is supplied by the feedback provided by the interface. Therefore, feedback is an indispensable part of the task flow. Figure 5.3 below, displays the coupling of user's action with the relevant feedback.

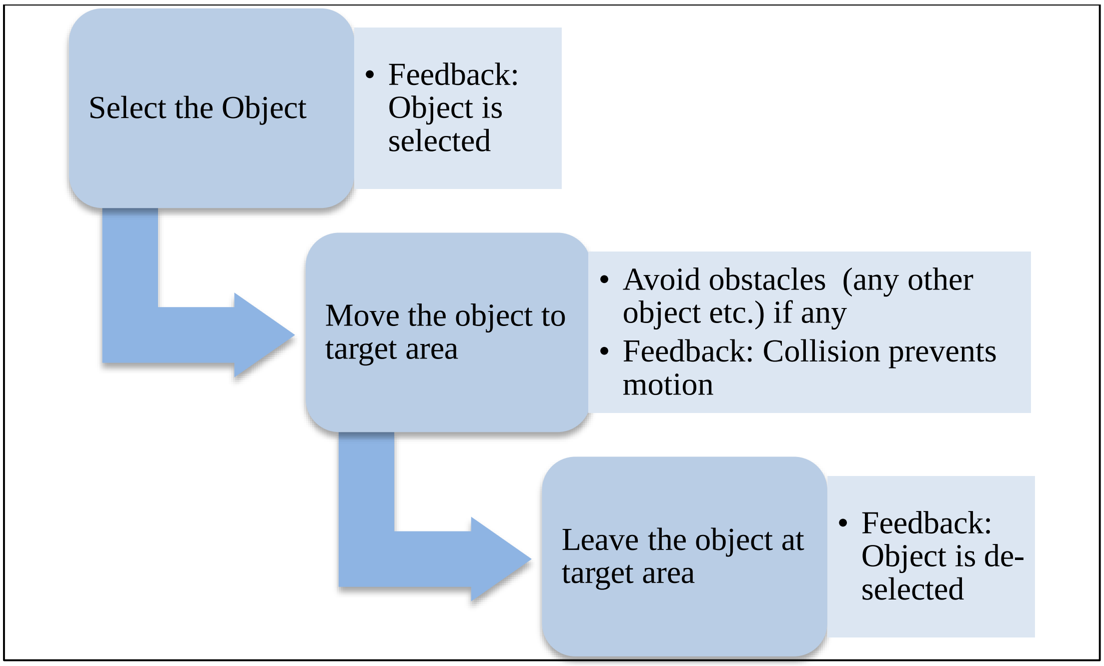

Figure 5. 3: Flow of a simple task.

For the clarity of the evaluation and efficiency in testing, generic first Experiment Set is split into four experiments; two experiments carried out just with 2D mouse, and two experiments carried out with multimodal input. By this approach, comparison between the experiments and tasks is simplified to be relatively easier. Each experiment is planned to have a constant, as either target complexity or environmental variable complexity, with the expectation of easier comparison and evaluation of achievements. Below is the table for experiments, their definitions and aims.

Table 5. 1: Experiments, scopes and aims.

Experiment	**Experiment Definition**	**Experiment Aim**
Experiment 1	This experiment is composed of two tasks, Task 1 and Task 2. Once Task 1 is completed with its stages, test-bed proceeds to Task 2. Across both of the tasks, the target complexity is constant, environmental variables do change.	Tracking the completion time of each task. Effect of environmental complexity on task completion. Tracking the accuracy errors.
Experiment 2	This experiment is composed of two tasks, Task 3 and Task 4. Once Task 3 is completed with its stages, test-bed proceeds to Task 4. Across both of the tasks, the target complexity is constant, environmental variables do change.	Tracking the completion time of each task. Effect of environmental complexity on task completion. Tracking the accuracy errors.
Experiment 3	This experiment is composed of two tasks, Task 1 and Task 2 including the vertical dimension. Once Task 1 is completed with its stages, test-bed proceeds to Task 2. Across both of the tasks, the target complexity is constant, environmental variables do change.	Tracking the completion time of each task. Effect of environmental complexity on task completion. Tracking the accuracy errors.

Table 5.1. Continued.

Experiment	Experiment Definition	Experiment Aim
Experiment 4	This experiment is composed of two tasks, Task 3 and Task 4 including the vertical dimension. Once Task 3 is completed with its stages, test-bed proceeds to Task 4. Across both of the tasks, the target complexity is constant, environmental variables do change.	Tracking the completion time of each task. Effect of environmental complexity on task completion. Tracking the accuracy errors.

5.4.2. Tasks for 2D Mouse input

All positioning tasks that are performed in Experiment-1 and Eperiment-2 (eight tasks in total) are created to evaluate interaction in 3D DVE via 2D mouse input. As discussed before, starting with a familiar phenomenon in interaction—2D mouse and gradually incrementing task complexity are presumed to help evaluation.

Iterating from the experiment set information provided in the section above and Table 5. 1, Experiment-1 and Experiment-2 are connected by means of the progress of complexity within the experiments since they both start with constant target complexity then increasing environment complexity. Also, they are connected by means of Experiment-2 being the successor of Experiment-1 based on the target complexity. The change in environmental variable complexity is the same across the tasks of Experiment-2 as with Experiment-1 whereas the target complexity of Experiment-2 is higher than Experiment-1. Yet, results of the experiments show the validity of the complexity assumption about the target area.

Adhering to the task complexity level and suggestions about user aid in simplifying actions required within the task flow, tasks are defined along with their helper text aiding to achieve the goal as listed in Table 5. 4 in Appendix-A. This table is detailing the definition of tasks based on First Experiment Set as defined in Section 5.3.1, utilising 2D Mouse Input. Tasks defined in Table 5. 4 in Appendix-A do not require head tracking but 2D mouse input since interaction is in two dimension—x-z plane, as explained before.

For evaluation purposes and easy comparison, Table 5. 2 would be helpful, since it is grouping the tasks by means of target clarity, environmental variables and target location. Figures of each task can be found in Appendix-B: Screenshots.

Table 5. 2: Short table for positioning tasks by interaction via 2D mouse.

Task Definition in Test-bed	**Target Clarity**	**Env. Variable**	**Target Location**	**Task id**
Grab the red cube object and carry it to the target area marked with yellow square. No other objects adjacent to the target area are present.	Clear target[3]	Empty space	Rear	Task 1.1
			Right	Task 1.2
		Obstacles	Rear	Task 2.1
			Right	Task 2.2
Grab the red cube and carry it to the target area marked with yellow square. There are other objects present adjacent to the target area.	Complex target[4]	Empty space	Rear	Task 3.1
			Right	Task 3.2
		Obstacles	Rear	Task 4.1
			Right	Task 4.2

5.4.3. Tasks for multimodal input

Same tasks are used to create the second part of the first Experiment defined in Section 5.3.1, that utilises multimodal input to navigate towards the target are. Experiment-3 and Experiment-4 (eight tasks in total) are derived from the composition of generic tasks defined in Section 5.3.1. The tasks, this time, include navigation in 3D space with target areas and navigation requiring translation over another object. See Table 5. 5 in Appendix-A, for detailed explanation of tasks that are performed via multimodal interaction as combination of head tracking and 2D mouse input.

Table 5. 3 below is helpful for evaluation purposes and easing the comparison, since it is grouping the tasks by means of target clarity, environmental variables and target location. Figures of each task can be found in Appendix B: Screenshots.

[3] *Clear target*: No other objects adjacent to the target area.
[4] *Complex target*: Another object is adjacent to the target area.

Table 5. 3: Short table for positioning tasks by multimodal input.

Task Definition in Test-bed	Target Clarity	Env. Variable	Target Location	Task id
Grab the cube object and carry it to the target area marked with yellow square. No other objects adjacent to the target area (other than the one underneath the target) are present.	Clear target[3]	Empty space	Rear	Task 5.1
			Right	Task 5.2
		Obstacles	Rear	Task 6.1
			Right	Task 6.2
Grab the red cube and carry it to the target area marked with yellow square. There are other objects present adjacent to the target area.	Complex target[4]	Empty space	Rear	Task 7.1
			Right	Task 7.2
		Obstacles	Rear	Task 8.1
			Right	Task 8.2

5.5. On Evaluation within Test-bed—Expectations and Foresights

Since same tasks are reused to test the usability of 2D mouse and multimodal interaction, it is also possible to directly compare tasks of Experiment-1 with the tasks of Experiment-3 by the interaction method. For further descriptions about the tasks, see Table 5. 2 and Table 5. 3 for a short rationale describing the whole picture of similarities, constants and differences across the tasks defined within test-bed and Table 5. 5 in Appendix-A for a detailed explanation. Components utilised in test-bed scenes except the interaction method is constant across both experiment groups. Therefore, such comparisons give insights into the complexity of the multimodal interaction by using 2D mouse results as a baseline. Moreover, the gradual nature of the experiments enables deductions via eliminating constants defined in the test-bed.

Below are some examples about evaluation approaches in order to sample the comparisons all across the experiments within the test-bed. In all scenes, yellow sphere represents the “virtual hand” (See Section 5.2.2 for explanation about “virtual hand”).

5.5.1. Example One

Task 1 and Task 2 can easily be compared based on the task completion times since any time difference between the completion of these two tasks depends on the complexity of the navigation among the other objects—defined as environmental variables within the scope of this study. Therefore, the comparison of completion times of Task 1.1 and Task 2.1 concludes to the effect of other objects'—obstacles presence in the environment (See Figure 5.4 and Figure 5.5).

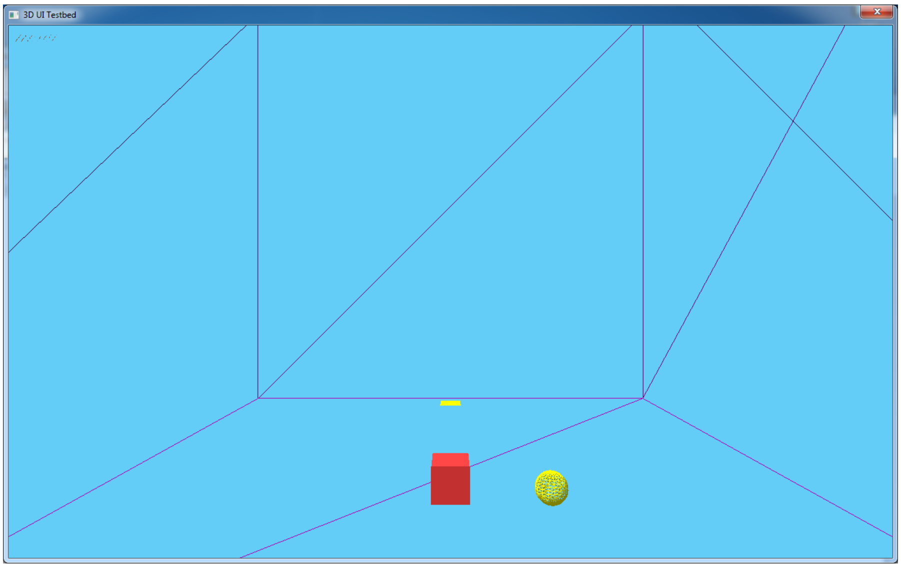

Figure 5. 4: Scene of Task 1.1 with rear target location and no obstacles.

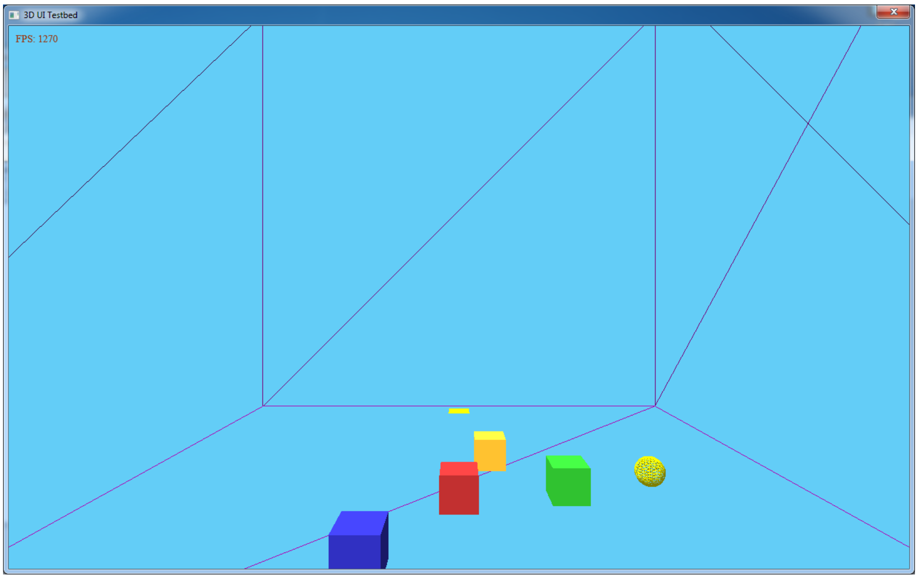

Figure 5. 5: Scene of Task 2.1 with rear target location and obstacles in the environment. Yellow sphere is representing the "virtual hand".

5.5.2. Example Two

Quite similar to the first example, when second iterations of these tasks are compared, the comparison concludes to the same result of evaluating the effects of obstacles. However, if the progression in Task 1.1 to Task 1.2 and Task 2.1 to Task 2.2 is considered and compared, the comparison of progression in completion time helps deducting about the effects of target location—far target, near target as well as hints about the depth perception in 3D DVE (See Figure 5.6 and Figure 5.7).

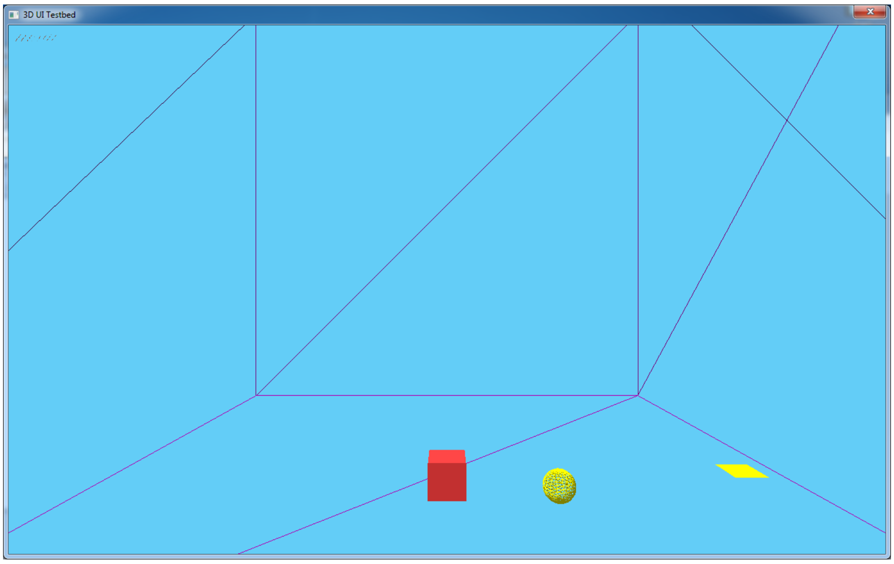

Figure 5. 6: Scene of Task 1.2 with right target location and no obstacles.

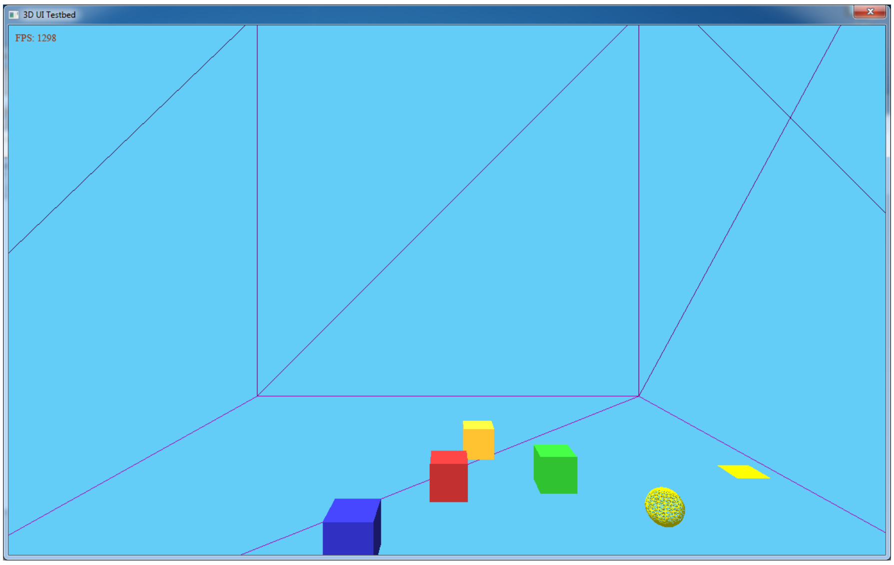

Figure 5. 7: Scene of Task 2.2 with right target location and obstacles in the environment. Yellow sphere is representing the “virtual hand”.

5.5.3. Example Three

Comparison of completion time and the accuracy errors between Task 1 and Task 3 display the effects of change in target complexity—from simple to complex (See complex target[4]—another object adjacent to the target area). This change can be compared with the change in the same set in Experiment-3 and Experiment-4 to evaluate the efficiency of multimodal input based on the target complexity. See Figure 5.8 and Figure 5.9 for visual representation of the scenario.

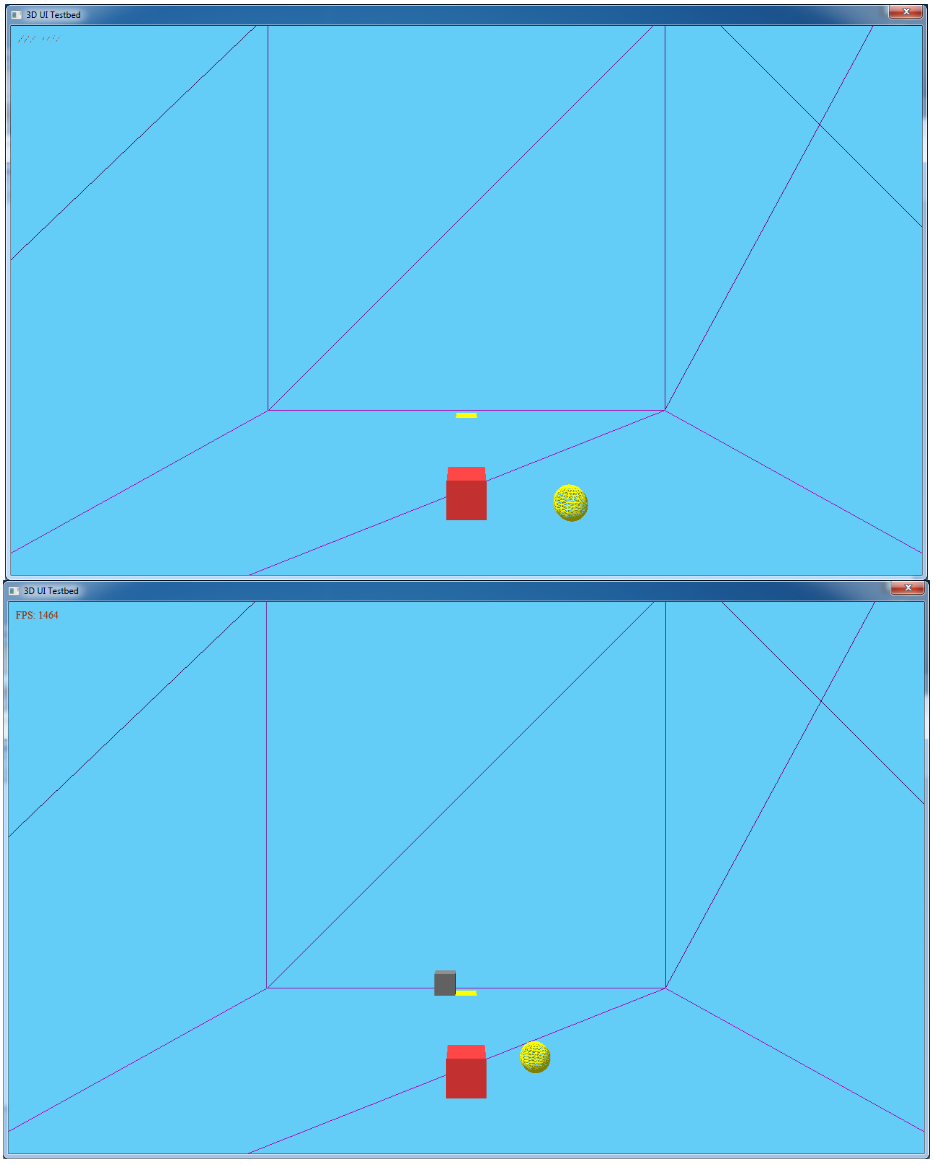

Figure 5. 8: Comparison of Task 1.1 and Task 3.1 scenes displaying clear target versus complex target4 as target complexity. These two scenes above belong to the test-bed experiment set that requires 2D Mouse interaction. Yellow sphere is representing the "virtual hand".

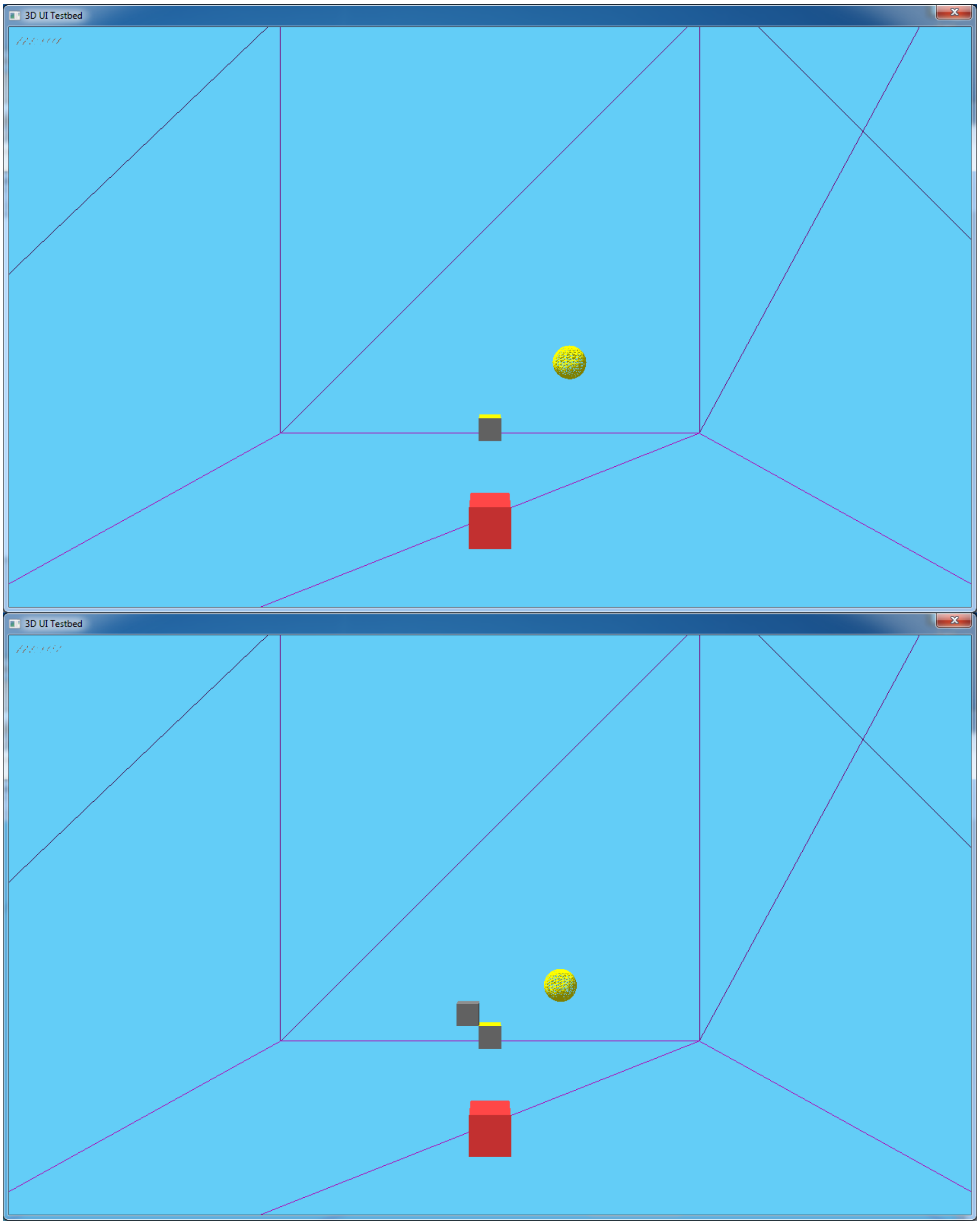

Figure 5. 9: Comparison of Task 5.1 and Task 7.1 scenes displaying clear target versus complex target4 as target complexity. These two scenes above belong to the test-bed experiment set that requires multimodal interaction. Yellow sphere is representing the “virtual hand”.

5.5.4. Example Four

Tasks of Experiment-1 and Experiment-2, similarly Experiment-3 and Experiment-4, can be compared one-to-one based on target complexity as sampled in Example Three, whereas first and second iteration of each task can be compared based on the distance to target location—target location rear or right.

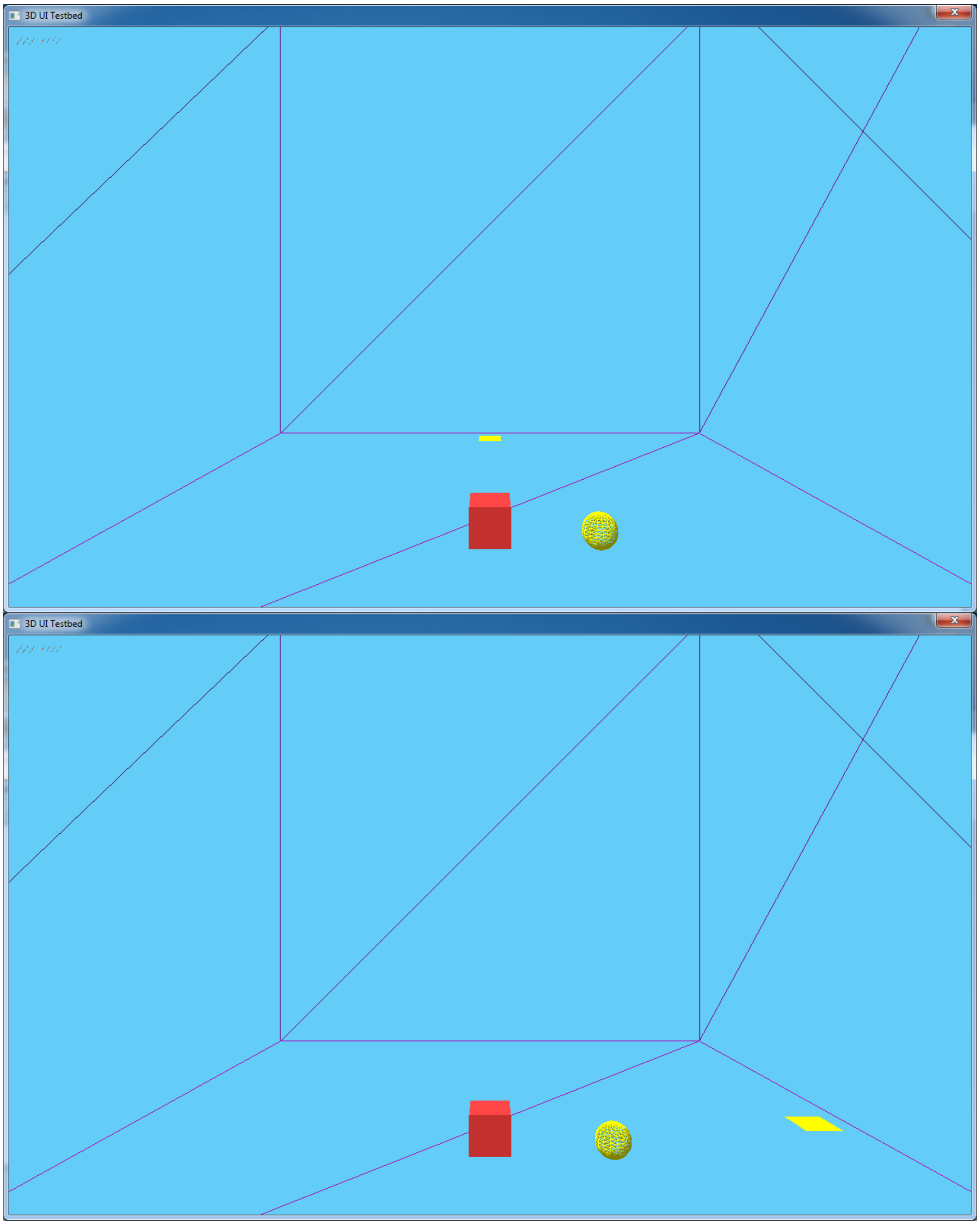

Figure 5. 10: Change in distance to target location between Task 1.1 and Task 1.2. No obstacles in the scene, target is a clear target3, and yellow sphere is representing the "virtual hand".

Completion time complexity between Task 1.1 and Task 1.2 displays the effect of target distance on interaction (See Figure 5.10 above). Same approach can be followed between any first and second iteration. See Figure 5.11 and Figure 5.12 for more samples.

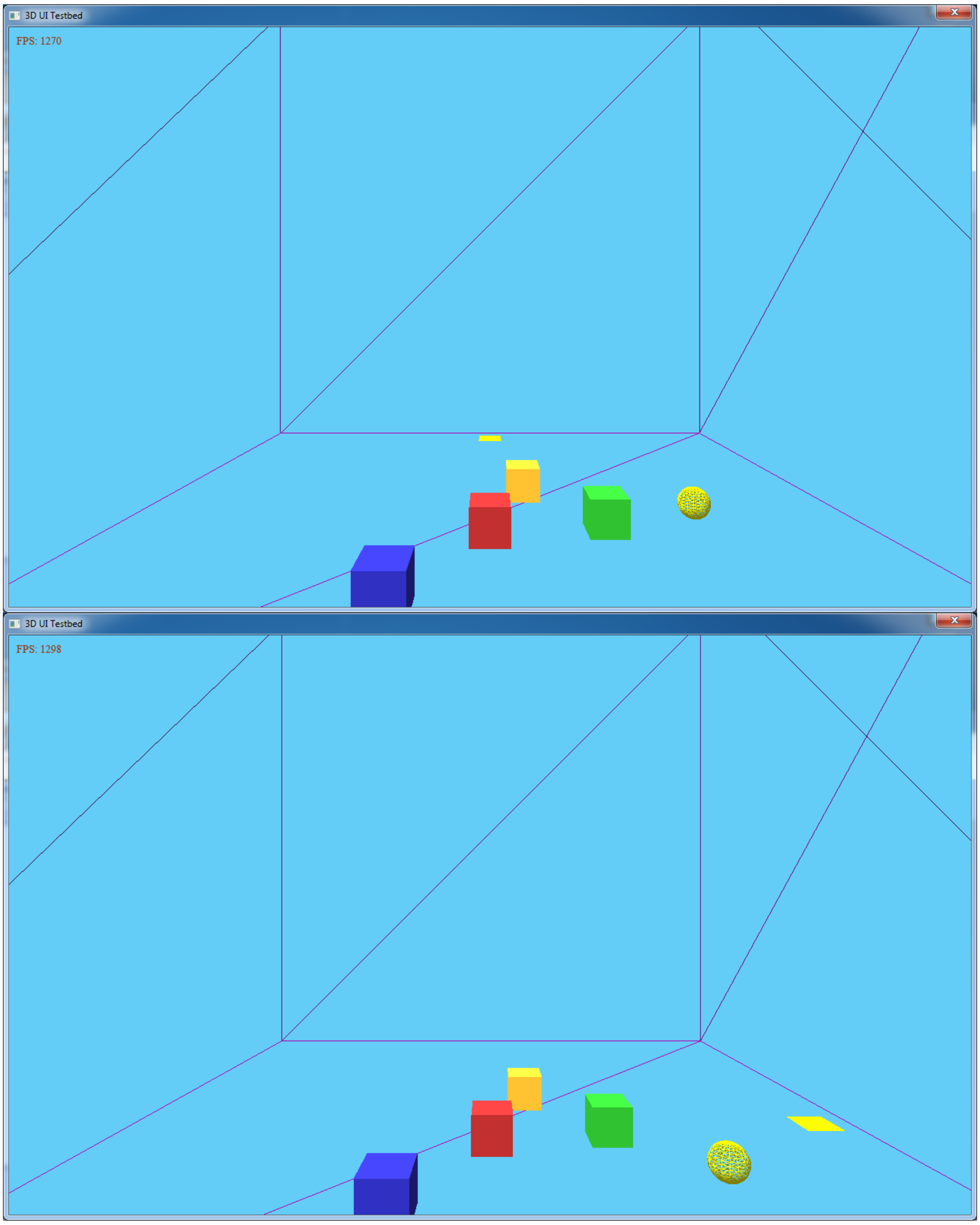

Figure 5. 11: Change in distance to target location between Task 2.1 and Task 2.2. There are obstacles in the scene, target is a clear target3 and yellow sphere is representing the “virtual hand”.

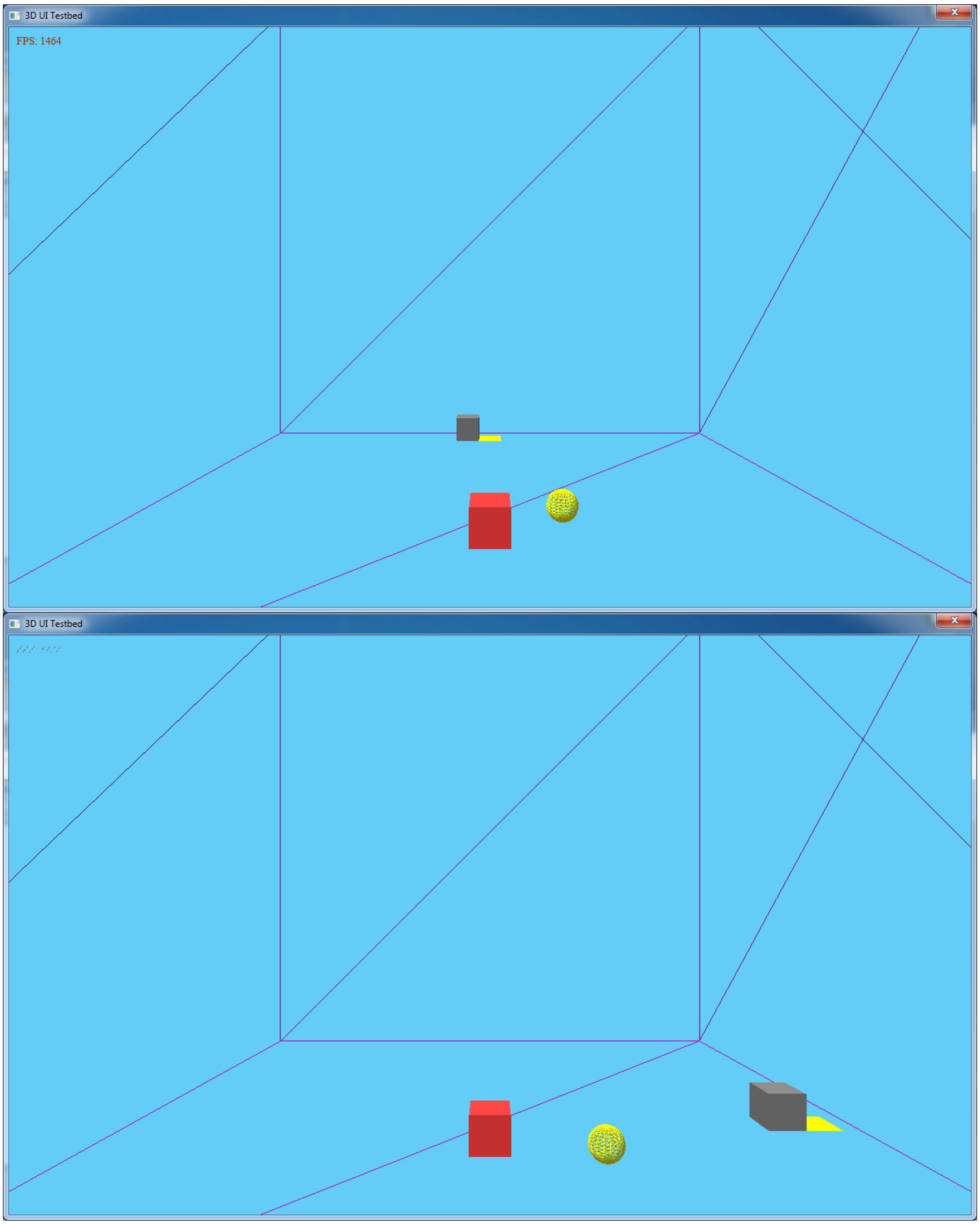

Figure 5. 12: Change in distance to target location between Task 3.1 and Task 3.2. There are no obstacles in the scene, target is a complex target[4], and yellow sphere is representing the "virtual hand".

6. MEASURES and EVALUATION

6.1. Performance Measures

This test-bed is fairly similar to the suggested test bed approach of Bowman, which was used to evaluate 3D interaction metaphors to carry out simple tasks and generic 3D interaction [26]. Even though the success of the tasks and usability are more important for a good interaction rather than the speed of completion, the number of attempts, the progress in every stage of an attempt and the amount of time a task takes would give a good idea of the learning curve by means of monitoring the possible progress of a user who is experiencing such an interaction for the first time. Therefore, time is considered as a measure of performance. Besides, accuracy or granularity of failure is another measure that should be evaluated both by means of task accomplishment and with consideration of different accuracy levels required by different tasks. Even though a primitive task requiring strict accuracy is successful, a relatively simple task expecting the same accuracy might fail just because of the nature of the task. This should be a reminder for the possible ambiguity originating from the nature of the task that might cause difficulty in evaluating the effectiveness of 3D interaction method on different levels of tasks. Therefore, reflection of the user is quite important for comparison of the interactions and tasks.

In conclusion, the intuitiveness, comfort and ease of use of each interaction are evaluated by the questionnaires filled after each task as well as accuracy; completion time and rate of error are used to evaluate the success of the interaction. The metrics collated from the tasks and questionnaires are to be evaluated by empirical evaluation methods.

6.2. Questionnaires

To evaluate the factors that are not directly measurable, a variety of questionnaires are designed to acquire information about abstract performance values such as ease of use, comfort and ease of learning (despite ease of learning metric can be obtained with repetitive testing).

A separate user reaction survey is to be filled for each part of the experiments with the completion of each task. By repeating the questionnaire with each task, evaluation by comparing the task difficulty level becomes possible. Each questionnaire question has e weight on a five point Lickert scale (Strongly disagree, disagree, neutral, agree, and strongly agree) to discover what stands out from the user's experience with the system [32].

Every survey starts with three questionnaire items as listed below:

i. "Interaction with the environment was intuitive."
ii. "Using the mouse to select the object(s) was intuitive."
iii. "I had difficulty in understanding which object I selected."

The rest of the questionnaire is different depending on the experiment task set as listed below:

- First Experiment Set
  - Navigation with 2D Mouse input
    iv. "Moving object(s) in the scene was easy."
    v. "I had difficulty in moving the object(s) around."
    vi. "Moving the objects around in the scene with mouse was intuitive."
    vii. "It was easy to navigate the object towards the designated place."
    viii. "Placing the object at the designated place was tricky."
    ix. "It was not comfortable to use mouse to move the object in two dimensional space."
    x. "It was intuitive to navigate the object with the mouse as right to left and further to closer."
  - Navigation with multimodal input
    iv. "Moving object(s) in the scene was easy."
    v. "I had difficulty in moving the object(s) around."
    vi. "Moving the object(s) with my head was confusing."
    vii. "Combining mouse input with my head motion to navigate the object towards the designated place was very difficult."
    viii. "Placing the object at the designated place with the help of head motion was tricky."
    ix. "It was not comfortable to use my head to move the object up and down in the scene."

x. "It was intuitive to move the object up and down with head motion."

- Second Experiment Set
  - Rotation with head tracking input

iv. "Changing orientation of the object to look like the given sample was easy."

v. "I had difficulty in rotating the object in right-left direction."

vi. "I had difficulty in rotating the object in up-down direction."

vii. "It was easy to rotate the object around itself."

viii. "It was difficult to stop changing the orientation and leave the object in the expected orientation."

ix. "It was not comfortable to use my head to change the orientation of the object in the scene."

x. "While performing rotation, I had difficulty in following the change in orientation."

xi. "Performing rotation with my head made me dizzy."

xii. "Changing the orientation of the object with my head orientation is intuitive."

xiii. "Changing the orientation of the object with my head orientation is a difficult performance to grasp."

  - Manipulation with multimodal input

iv. "Performing the task from start to finish was easy to accomplish."

v. "It was difficult to stop changing the orientation and leave the object in the expected orientation."

vi. "I had difficulty in transition from one part of the task to the other when I was changing the orientation with my head at first and then navigating to the designated spot or vice versa."

vii. "When the parts of the task are performed in selection, rotation, navigation order, it is easier to accomplish."

viii. "When the parts of the task are performed in selection, navigation, rotation order, it is easier to accomplish."

ix. "It is simple to differentiate between navigation and orientation parts of the task performed by head motion."

x. “While performing rotation, I had difficulty in following the change in orientation.”
xi. “Performing changes in orientation with my head made me dizzy.”
xii. “Changing the orientation and navigating the object both by head is confusing.”
xiii. “Using both mouse and my head to manipulate the object is a difficult performance to grasp.”
xiv. “Placing the object at the designated place as ordered in task definition was tricky.”
xv. “While performing the task, stopping at a point and continuing from where I left was not easy to accomplish.”

6.3. Evaluation Approaches

The evaluation paradigms chosen for this study are usability testing and user observation. As the users perform the tasks within test-bed, they are watched, their behaviours are recorded on video and/or are logged according to the metrics explained in Section 6.1. There are many usability methods and techniques available that are suitable for conventional computerized office applications running on desktop computers, most of which are formed around Norman’s “Seven Stages of Action” model [39]. The model characterizes a user’s list of actions aligned with the way that the tasks are constructed and put into order in this study. The steps in Norman’s model [40] are as follows:

- Form the goal;
- Form the intention;
- Identify the action;
- Execute the action;
- Percieve the response;
- Interpret the results;
- Understand the outcome.

Questionnaires and interviews are used to elicit users' opinions to have a better conclusion acknowledging human factors.

Design process of the interaction, interface and test-bed follows iterative and incremental design methodology by means utilising early user input to improve the system. For this purposes informal tests are conducted as well as observations to stir the design towards a more usable direction by obtaining early feedback from the users. Holding the mouse button down during selection was one of the first to come out as a difficulty; therefore, selection is improved to employ select-deselect behaviour with left mouse button down providing a visual feedback (glowing object) to indicate selection. This indication stays visible until the object is de-selected in order to prevent possible confusion of selection with the help of visual feedback.

Early tests for navigation with mouse in 2D (refers to the first part of the first experiment set, 5.3.1. First Experiment Set) displays that performance of the user increases from one task to another showing that learning curve is not steep. Also, the order of the tasks do enable users get trained and enable the researcher decouple the complexity of new task from the already accomplished task.

6.3.1. Test Results and Evaluation

6.3.1.1. Evaluation – Step 1

Task completion time of each participant is displayed in Figure 6.1. Other than drawing information on individual improvement in performance or how individual's habits and familiarity affect their performance, this chart displaying individual completion time is not telling us as much as average completion time about the interaction itself.

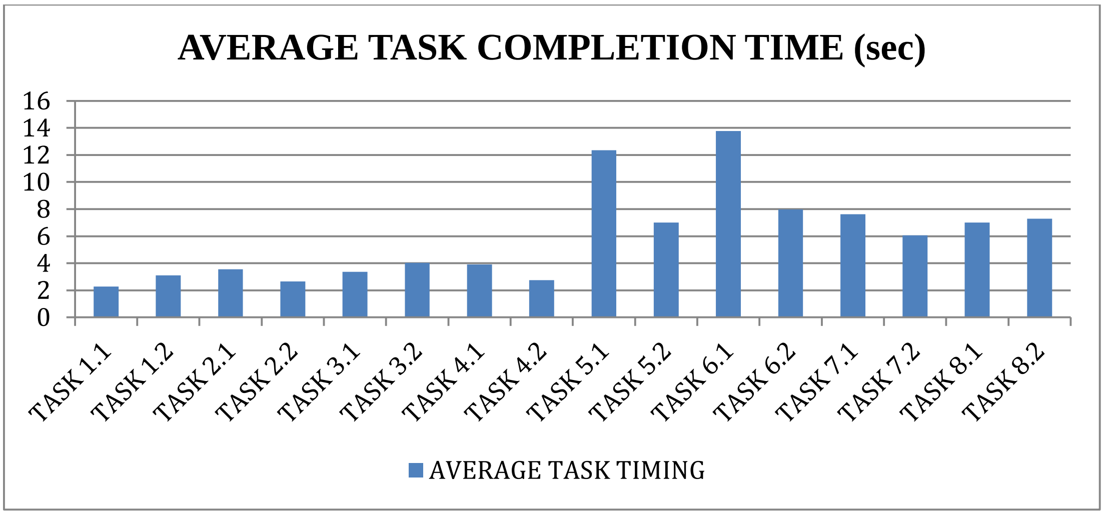

Figure 6.1: Average task completion time of the participants across all tasks are displayed in this chart. First eight tasks are performed via 2D mouse input, whereas second eight tasks are performed via multimodal input. For a detailed explanation of the tasks, their complexity and generation process see Section 5.4 and its subsections.

6.3.1.2. Evaluation - Step 2

Average task completion time is calculated for each task to identify the effect of task complexity on completion time. As seen in Figure 6.2, first eight tasks that are performed via 2D mouse have far lesser completion time compared to the tasks performed via multimodal input. Taking into account Table 5. 2 and Table 5. 3, similarity between the curvature created by tasks from Task 1.1 to Task 2.2 (first four tasks) and the curvature created by tasks from Task 3.1 to Task 4.2 (second four tasks) shows that target complexity has an increasing effect on completion time. On the other hand, when same approach is followed for the second group of tasks that are performed by multimodal input, the completion time drops with the target complexity suggesting that either familiarity to interaction is increased or complex target[4] is helpful in positioning tasks via multimodal input.

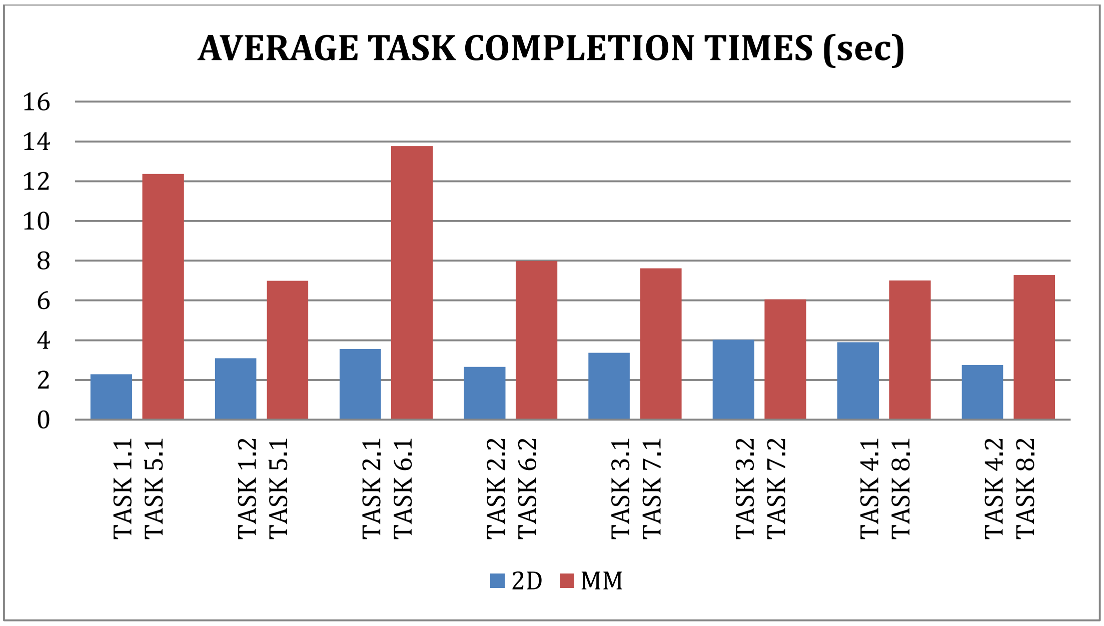

Figure 6.2: Average task completion time calculated from completion times of participants across all tasks are displayed in this chart by bringing 2D mouse and multimodal interactions tasks together for a better comparison. For a detailed explanation of the tasks, their complexity and generation process see Section 5.4 and its subsections.

6.3.1.3. Evaluation - Step 3

The spikes of Task 5.1 and Task 6.1 show the first reaction to the multimodal input suggesting that unfamiliarity to the interaction is resolved quickly. Also, considering that designated target area of both Task 5.1 and Task 6.1 are rear, and watching the video footage of the user performance during these tasks confusion in depth feeling could be the main reason of the spikes. Video footage shows that many users have difficulty in placing the object in the target area, and they either place far behind or before the area leading us to consider the need of improvement in depth perception. Moreover, the drop in completion times of following tasks—Task 5.2 and Task 6.2, leads to conclude that complex target[4] reduces the negative effect of depth field helping in engagement towards the target and understand the position of the moving object.

Another reason that creates the spike of Task 6.1 is probably the bigger size of the obstacles in the scene that prevents the clear vision of the target. That is noted for future run of the test-bed and experiments as all of the objects should be created identical for a more just comparison that is not affected by such change in the environmental variables.

6.3.1.4. Evaluation - Step 4

If we look at Figure 6.2, it is much easier to see the completion time relations of similar tasks suggesting the overhead created by multimodal input. Blue tasks are performed via 2D mouse whereas red tasks are performed via multimodal input.

Task 5.1 and Task 7.1 are identical tasks except the target complexity. However, the completion time of Task 7.1 has drastically dropped even though the complexity it has is higher than Task 5.1 target complexity. This situation can be evaluated in the way either another object adjacent to target eases positioning as we discussed before or improvement in performing via multimodal input which can be evaluated as learnability. Among almost identical tasks, Task 5.1 and Task 7.1, improvement in task completion time is 38%; nonetheless, from Task 6.1 to Task 8.1 improvement in task completion time is even higher, 49%.

Improvement in tasks completion time is calculated based on the equation Eq 6.1, which is derived for a task group of two tasks to find out the selected progression of in completion time from the first task to the second task based on the completion time of the first one. For instance, when the progression from Task 6.1 to Task 8.1 is calculated, Task 6.1 is the first task—t_1 for completion time; Task 8.1 is the second task—t_2 for completion time.

$$\frac{(t_1 - t_2)}{t_1} \qquad (6.1)$$

- t_1: completion time of the first task in the group
- t_2: completion time of the second task in the group

6.3.1.5. Evaluation - Step 4

When we look at individual completion times of each user, test subject three stands out by completing almost all tasks significantly faster than any other users except Task 5.1, which is the very first task performed via multimodal input. By looking at timings performed by mouse input, it is easy to draw a conclusion saying that this user is competant using mouse. Then, he/she spent time to find out how to interact in Task 5.1, which took him/her longer. Nevertheless, he/she performed successfully in the rest of the tasks after trying out and learning the system.

The individual task completion times of the participants are as displayed in Figure 6.3 below.

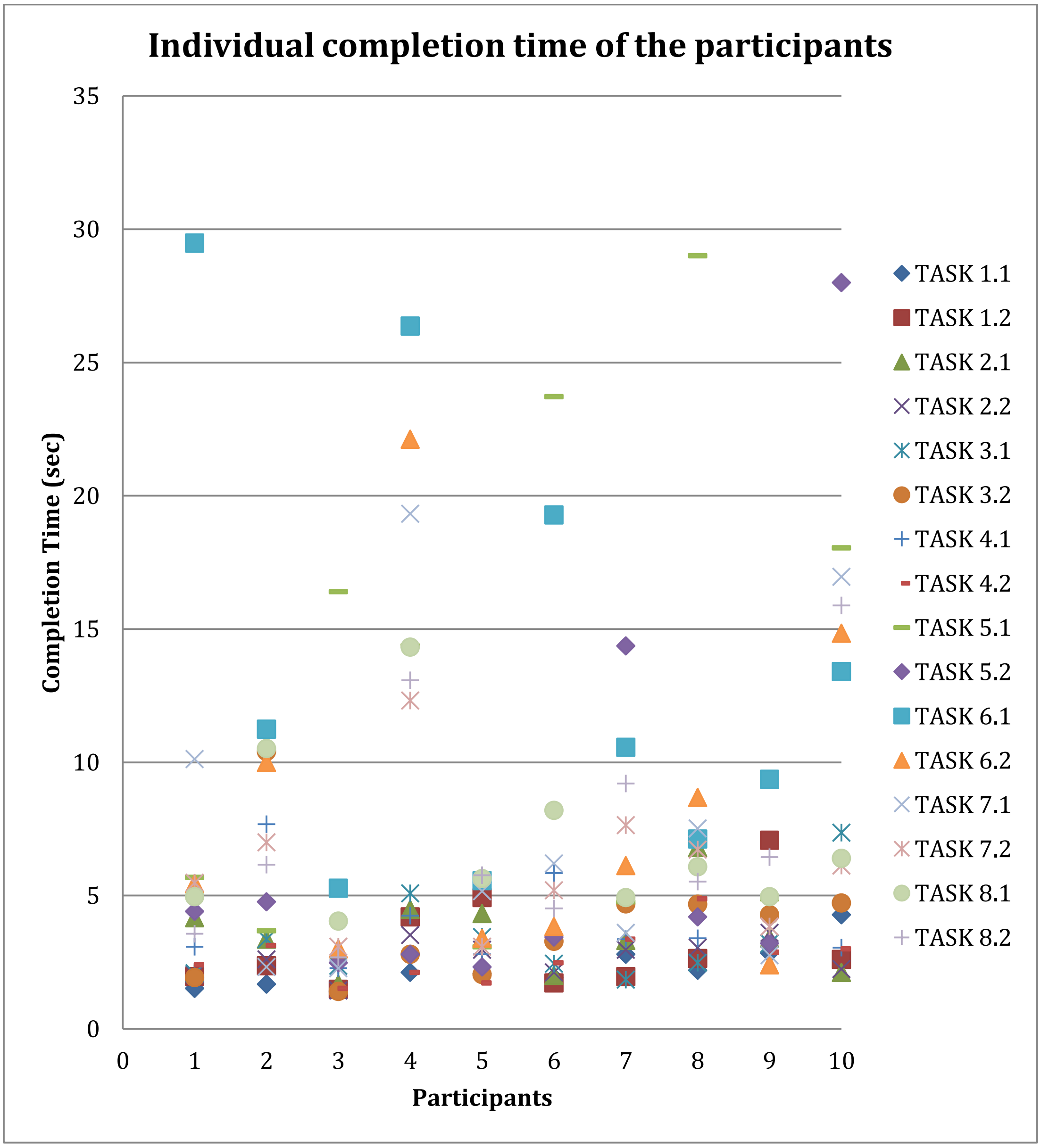

Figure 6.3: Individual completion time of each participant is displayed showing the progress of each participant within the test-bed.

If further evaluations are requried, further detailed charts for individual completion time on each task can be found in Appendix C: Charts section.

6.3.2. Questionnaires and Evaluation

A set of questionnaire questions is created based on the generic questionnaires listed in Section 6.2 based on Lickert scale five. The questionnaire is given to the participants upon their completion of the all experiments in each trial. Three trials are run for multimodal interaction with five participants, and in the last two of them participants are asked to fill in the questionnaire. The reason to employ questionnaires in this study is to understand the perception change of users across the trials in order to accomplish a qualitative evaluation.

The questionnaire is composed of the questions listed below:

- *"Moving object(s) in the scene was easy."*
- *"It was easy to navigate the object towards the designated place."*
- *"Placing the object at the designated place was tricky."*
- *"Moving the object(s) with my head was confusing."*
- *"Combining mouse input with my head motion to navigate the object towards the designated place was very easy."*

For a better evaluation, the questionnaire should be given to the participants after each trial, and running many trials would be better for a more appropriate evaluation.

7. CONCLUSION

With this study, a 3D multimodal user interface and multimodal interaction is developed within a 3D DVE. The functionality of this multimodal interaction is evaluated over manipulation task performance in a custom developed test-bed environment. Considering completion time of 2D Mouse tasks as baseline, performance of multimodal interaction is evaluated. The results of this study help us understand the complexity level of head tracking based input for a 3D DVE. Furthermore, results give insight into the effects of target complexity and environmental variables on the interaction in 3D DVE with the fact that 3D perception of the environment is quite valuable for 3D interaction. Also, learnability of this novel interface is tested to some extend revealing that success of interaction increases from trial to trial since the familiarity of the users increase.

For future work, the test-bed is going to be improved with the learnings from the first series of tests and extended to perform the experiments that cover the last canonical task—rotation of manipulation. After the whole manipulation tasks are performed by multimodal interface and evaluated within this test-bed, depending on the results, interaction can be moved to the next stage to be integrated into a CAD-like 3D environment for design purposes.

8. FUTURE WORK

The first step of future work chain is running the tests for the rotation component of canonical tasks and evaluating the data collected across all manipulation tasks within the test-bed for the multimodal interface. The results of the empirical study including all results of manipulation tasks will help evaluating the usability of multimodal interaction more efficiently. After that, multimodal interaction can be switched to a designer environment to find out its usability for such target users. Also, an application utilizing the developed multimodal interface in a game-like test environment to find out the intuitiveness, fun factor and usability of mouse head tracking integration for gaming purposes can be developed.

Besides, based on the user feedbacks, the test-bed can be utilized to improve the interface by employing the selective rotation centre for designing software purposes. After the interface and test-bed is catered for designers, formal tests can be conducted to compare Chen's Virtual Sphere technique to multimodal input with head tracking and evaluate the efficiency of this multimodal interface for designing software purposes.

REFERENCES

[1] Jacob R. J. K., Karn K. S., (2003), "Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises". In: The Mind's eye: Cognitive The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research, 2 (3), 573-603.

[2] Jaimes A., Sebe N., (2007), "Multimodal human-computer interaction: A survey", Computer vision and image understanding, 108 (1), 116-134.

[3] LaViola J. J., (2008), "Bringing VR and Spatial 3D Interaction to the Masses through Video Games", IEEE Computer Graphics and Applications, 28 (5), 10-15.

[4] Chen M., Mountford S. J., Sellen A., (1988), "A study in interactive 3-D rotation using 2-D control devices". In Proceedings of the 15th annual conference on Computer graphics and interactive techniques (SIGGRAPH '88), Richard J. Beach (Ed.) ACM, 121-129, New York, NY, USA, 1-5 August.

[5] Teather R. J., Stuerzlinger W., (2008), "Exaggerated head motions for game viewpoint control." In Proceedings of the 2008 Conference on Future Play: Research, Play, Share ACM, 240-243, New York, NY, USA, 3-5 November.

[6] Nickel K., Stiefelhagen R., (2003), "Pointing gesture recognition based on 3D-tracking of face, hands and head orientation". In Proceedings of the 5th international conference on Multimodal interfaces (ICMI '03) ACM, 140-146, New York, NY, USA, 5-7 November.

[7] Gorodnichy D. O., Malik S., Roth G., (2004), "Nouse 'use your nose as a mouse' perceptual vision technology for hands-free games and interfaces". Image and Vision Computing, 22 (12), 931-942.

[8] Morency L., Rahim, A., Checka N., Darrell T., (2002), "Fast stereo-based head tracking for interactive environments". In Proceedings Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 390-395, Washington, DC, USA, 21 May.

[9] Fu Y., Huang T. S., (2007), "hMouse: Head Tracking Driven Virtual Computer Mouse". In Proceedings of the Eighth IEEE Workshop on Applications of Computer Vision (WACV '07). IEEE Computer Society, Washington, DC, USA, 21-22 February.

[10] Web 1, (2013), http://www2.isr.uc.pt/~pedromartins/, (Retrieved 10/05/2013).

[11] Betke M., Gips J., Fleming P., (2002), "The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access For People with Severe Disabilities." IEEE Transactions on Neural Systems and Rehabilitation Engineering, 10 (1), 1-10.

[12] Web 2, (2010), http://www.cs.bu.edu/faculty/betke/research/hci.html, (Retrieved 01/04/2013).

[13] Ashdown M., Oka K., Sato Y., (2005), “Combining head tracking and mouse input for a GUI on multiple monitors”. In CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05) ACM, 1188-1191, New York, NY, USA, 2-7 April.

[14] LaViola J.J., Forsberg A., Huffman J., Bragdon A., (2008), "Poster: Effects of Head Tracking and Stereo on Non-Isomorphic 3D Rotation", IEEE Symposium on 3D User Interfaces, 3DUI 2008, 155-156, Reno, NE, 8-9 March.

[15] Web 3, (2013), http://www.seeingmachines.com/product/faceapi/, (Retrieved 01/02/2013).

[16] Web 4, (2010), http://facetracknoir.sourceforge.net/home/default.htm, (Retrieved 01/02/2013).

[17] Bowman D. A., Kruijff E., LaViola J.J., Poupyrev I., (2001), “An Introduction to 3-D User Interface Design”. Presence: Teleoper. Virtual Environ. 10 (1), 96-108.

[18] Bowman D. A., Kruijff E., LaViola J.J., Poupyrev I., (2004), “3D User Interfaces: Theory and Practice”, 1st Edition, Addison-Wesley.

[19] Froehlich B., Kitamura Y., Bowman D., (2004), "Workshop: Beyond wand and glove based interaction." Virtual Reality Proceedings IEEE, 268, Chicago, Illinois, USA, 27-31 March.

[20] Schafer W. A., Bowman D. A., (2005), “Integrating 2D and 3D views for spatial collaboration”. In Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work (GROUP '05) ACM, 41-50, New York, NY, USA, 6-9 November.

[21] McMahan R. P., Alon A. J. D., Lazem S., Beaton R. J., Machaj D., Schaefer M., Silva M.G., Leal A., Hagan R., Bowman D., (2010), "Evaluating natural interaction techniques in video games". 3DUI '10 Proceedings of the 2010

IEEE Symposium on 3D User Interfaces, 11-14, Washington, DC, USA, 20-21 March.

[22] Hinckley K., Tullio J., Pausch R., Proffitt D., Kassell N., (1997), "Usability analysis of 3D rotation techniques". In Proceedings of the 10th annual ACM symposium on User interface software and technology (UIST '97) ACM, 1-10, New York, NY, USA, 14-17 October.

[23] Beaton R. J., DeHoff R. J., Weiman N., Hildebrandt P. W., (1987), "An evaluation of input devices for 3-D computer display workstations", True Imaging Techniques and Display Technologies, SPIE Vol. 761, 94-101.

[24] Bowman D. A., Chen J., Wingrave C. A., Lucas J. F., Ray A., Polys N. F., Li Q., Haciahmetoglu Y., Kim J., Kim S., Boehringer R., Ni T., (2006), "New Directions in 3D User Interfaces", IJVR 5 (2), 3-14.

[25] McMahan R. P., Gorton D., Gresock J., McConnell W., Bowman D., (2006), "Separating the effects of level of immersion and 3D interaction techniques". In Proceedings of the ACM symposium on Virtual reality software and technology (VRST '06) ACM, 108-111, New York, NY, USA, 1-3 November.

[26] Bowman D. A., Johnson D. B., Hodges L. F., (2001), "Testbed evaluation of virtual environment interaction techniques". Presence, 10 (1), 75-95.

[27] Poupyrev I., Weghorst S., Billinghurst M., Ichikawa T., (1997), "A framework and testbed for studying manipulation techniques for immersive VR". In Proceedings of the ACM symposium on Virtual reality software and technology (VRST '97) ACM, 21-28, New York, NY, USA, 15-17 September.

[28] Web 5, (2012), http://groups.csail.mit.edu/vision/vip/watson/, (Retrieved 01/10/2012).

[29] Bowman D., Hodges L., (1997), "An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments". In SI3D '97, 35-38, Providence, RI, USA, 27-30 April.

[30] Poupyrev I., Weghorst S., Billinghurst M., Ichikawa T., (1996), "The go-go interaction technique: non-linear mapping for direct manipulation in VR". In Proceedings of the 9th annual ACM symposium on User interface software and technology (UIST '96) ACM, 79-80, New York, NY, USA, 6-8 November.

[31] Bowman D. A., Hodges L. F., (1999), "Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments", Journal of Visual Languages & Computing, 10 (1), 37-53.

[32] Rosson M. B., Carroll J. M., (2002), “Usability Engineering: Scenario Based Development of Human Computer Interaction”, 1st Edition, Morgan Kaufmann Publishers.

[33] Mine M., (1995), "Virtual environment interaction techniques", Technical Report TR95-018, UNC Chapel Hill CS Dept., UNC, USA.

[34] Poupyrev I., Weghorst S., Billinghurst M., Ichikawa T., (1998), “Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques”. In Computer Graphics Forum, 17 (3), 41-52.

[35] Poupyrev I., Weghorst S., Fels S., (2000), “Non-isomorphic 3D Rotational Techniques”. In Proceedings of CHI'2000 ACM, 546-547, Hague, Netherlands, 1-6 April.

[36] Helander M.G., Landauer T.K., Prabhu P.V., (1997), “Handbook of Human-Computer Interaction”, 1st Edition, Elsevier.

[37] Koller D. R., Mine M. R., Hudson S. E., (1996), “Head-tracked orbital viewing: an interaction technique for immersive virtual environments”. In Proceedings of the 9th annual ACM symposium on User interface software and technology (UIST '96) ACM, 81-82, New York, NY, USA, 6-8 November.

[38] Stoakley R., Conway M., Pausch R., (1995), "Virtual reality on a WIM: interactive worlds in miniature", Proceedings of CHI'95, 265-272, Denver, Colorado, USA, 7-11 May.

[39] A. Dix, J. Finlay, G. Abowd, R. Beale, (1998), “Human Computer Interaction”, 2nd Edition, Prentice Hall.

[40] Norman D. A., (2002), “The Design of Everyday Things”, Reprint edition, Basic Books.

[41] Hinckley K., (2003), “The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies”, 2nd Edition, Lawrence Erlbaum Associates.

[42] Oviatt S., (2003), “The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies”, 2nd Edition, Lawrence Erlbaum Associates.

[43] Oh J. Y., Stuerzlinger W., (2005), “Moving objects with 2D input devices in CAD systems and Desktop Virtual Environments”. In Proceedings of Graphics Interface 2005 (GI '05), 195-202, Waterloo, Ontario, Canada, 9-11 May.

[44] Strauss P. S., Issacs P., Shrag J., (2002), “The design and implementation of direct manipulation in 3D”, SIGGRAPH 2002 Course Notes, San Antonio, Texas, USA, 23-26 July.

BIOGRAPHY

Aslıhan Tece was born in Mersin Turkiye, graduated from Istanbul Technical University with a Bachelor of Science degree in Computer Engineering. She was employed as software engineer at an IT company serving in defense industry, developing simulations and applications for command and control systems of navy ships. Afterwards, she decided to take a break and see the world with her husband. This journey took them to New Zealand and they settled down in Auckland. Since 2012, she has been employed as a lecturer at Game Development Faculty in Media Design School, teaching game development in Bachelor of Software Engineering-Game Programming degree. Her focus of interest is game design and development including but not limited to human computer interaction, computer graphics, physics programming, real time systems, parallel and distributed programming.

APPENDICES

Appendix A: Task Organization Tables

Table 5. 4.Task table for 1st Experiment Set utilising 2D Mouse Input

Experiment	**Task id**	**Task Definition in Test-bed**	**Task Target**	**Helper Text on screen**	**Task Image**
Experiment-1	Task 1	Grab the red cube object and carry it to the target area marked with yellow square. No other objects are present in the scene.	Place the object in yellow square. This task is composed of two stages: 1. Target area is at the back of the room. 2. Target area is at either side of the room. When first stage is completed, test-bed automatically continues to the second stage. Once two stages are completed, this task is completed.	***Task Definition:*** Take the red cube to the target area marked with yellow square. ***How:*** Grab the cube by bringing the hand to the cube and clicking with left mouse button when the hand is on the cube. Once the cube is grabbed, it is attached to the hand and can be moved by moving the mouse. Left click again to drop the cube.	Figure 11. 1 Figure 11. 2
	Task 2	Grab the red cube (there are other cubes that user can grab) and carry it to the target area marked with yellow square.			Figure 11. 3 Figure 11. 4

Table 5.5. continued.

Experiment-2	Task 3	Grab the red cube and carry it to the target area marked with yellow square. No other objects are present in the scene as obstacles other than the one right next to the target area.	Place the object in yellow square (target area is next to another cube object). This task is composed of two stages: 1. Target area is at the back of the room. Target and object next to it are aligned on x-axis. 2. Target area is at either side of the room. Target and object next to it are aligned on z-axis. When first stage is completed, test-bed automatically continues to the second stage. Once two stages are completed, this task is completed.	***Task Definition:*** Take the red cube to the target area marked with yellow square. ***How:*** Grab the cube by bringing the hand to the cube and clicking with left mouse button when the hand is on the cube. Once the cube is grabbed, it is attached to the hand and can be moved by moving the mouse. Left click again to drop the cube.	Figure 11. 5 Figure 11. 6
	Task 4	Grab the red cube (there are other cubes that user can grab) and carry it to the target area marked with yellow square.			Figure 11. 7 Figure 11. 8

Table 5. 5. Task table for First Experiment Set utilising Multimodal Input

Experiment	**Task id**	**Task Definition in Test-bed**	**Task Target**	**Helper Text on screen**	**Task Image**
Experiment-3	Task 5	Grab the cube object and carry it to the target area marked with yellow square (Yellow square is above another cube). No other objects are present in the scene (other than the one underneath the target).	Place the object in yellow square. This task is composed of two stages: 1. Target area is at the back of the room. 2. Target area is at either side of the room. When first stage is completed, test-bed automatically continues to the second stage. Once two stages are completed, this task is completed.	***Task Definition:*** Take the red cube to the target area marked with yellow square. ***How:*** Grab the cube by bringing the hand to the cube and clicking with left mouse button when the hand is on the cube. Once the cube is grabbed, it is attached to the hand and can be moved by moving the mouse and shifting the head up/down. Left click again to drop the cube.	Figure 11. 9 Figure 11. 10
	Task 6	Grab the red cube (there are other objects that are blocking the way, user needs to carry above and around those objects) and carry it to the target area marked with yellow square. Target area is on another cube object.			Figure 11. 11 Figure 11. 12

Table 5.6. continued.

Experiment-4	Task 7	Grab the cube object and carry it to the target area marked with yellow square. No other objects are present in the scene as obstacles other than the one right next to the target area and below it.	Place the object in yellow square (target area is next to another cube object). This task is composed of two stages: 1. Target area is at the back of the room. Target and object next to it are aligned on x-axis. 2. Target area is at either side of the room. Target and object next to it are aligned on z-axis. When first stage is completed, test-bed automatically continues to the second stage. Once two stages are completed, this task is completed.	***Task Definition:*** Take the red cube to the target area marked with yellow square. ***How:*** Grab the cube by bringing the hand to the cube and clicking with left mouse button when the hand is on the cube. Once the cube is grabbed, it is attached to the hand and can be moved by moving the mouse and shifting the head up/down. Left click again to drop the cube.	Figure 11. 13 Figure 11. 14
	Task 8	Grab the red cube (there are other objects that are blocking the way, user needs to carry above or around those objects) and carry it to the target area marked with yellow square.			Figure 11. 15 Figure 11. 16

Appendix B: Screenshots

Below are the screenshots from Experiment-1.

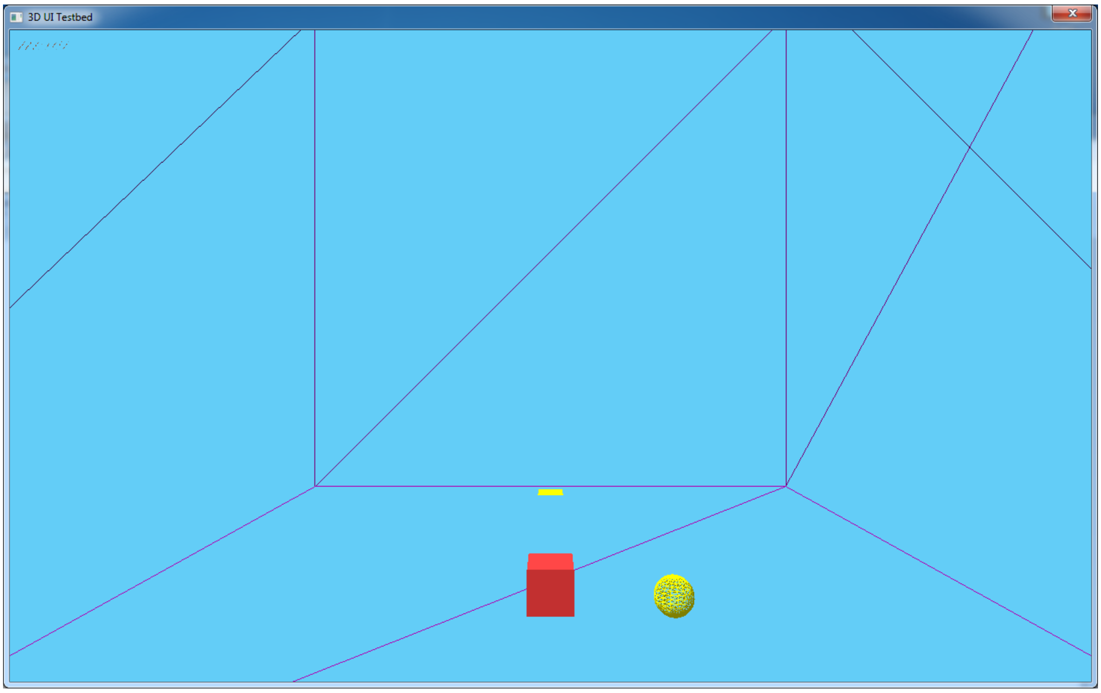

Figure 11. 1: Screenshot for Task 1.1 of Experiment-1.

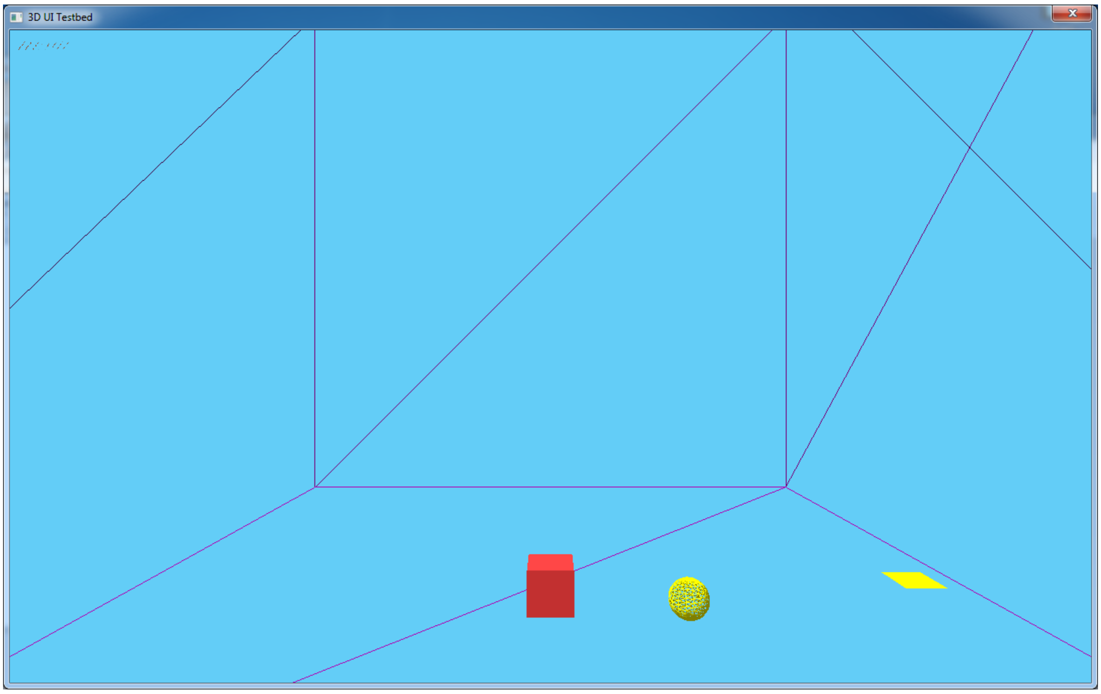

Figure 11. 2: Screenshot for Task 1.2 of Experiment-1.

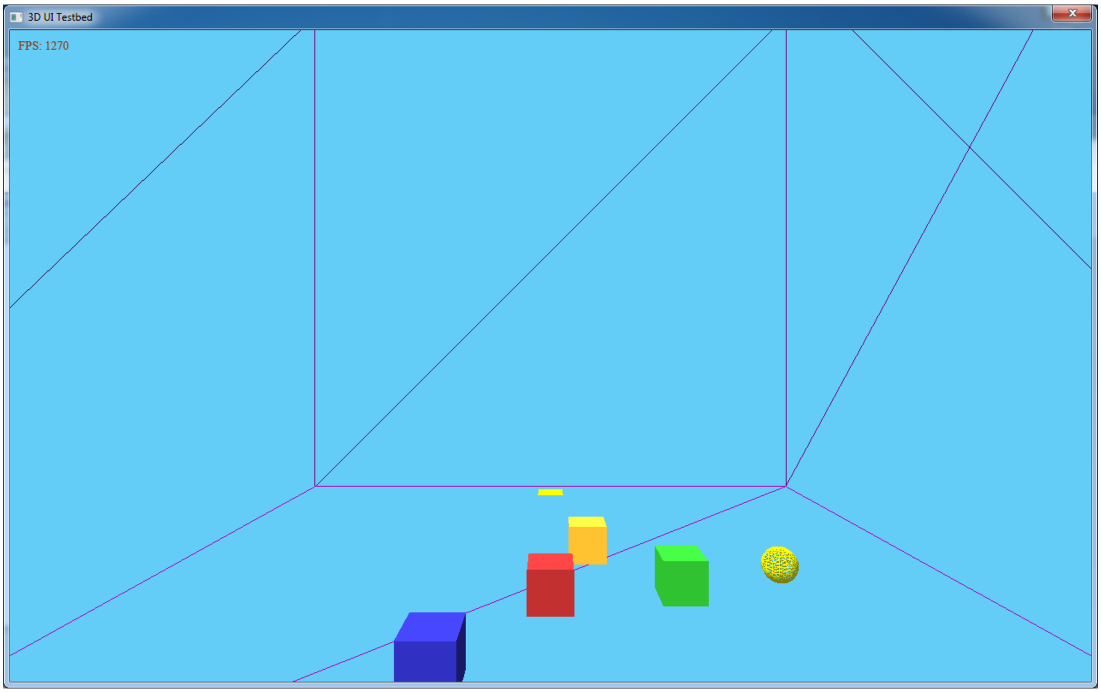

Figure 11. 3: Screenshot for Task 2.1 of Experiment-1.

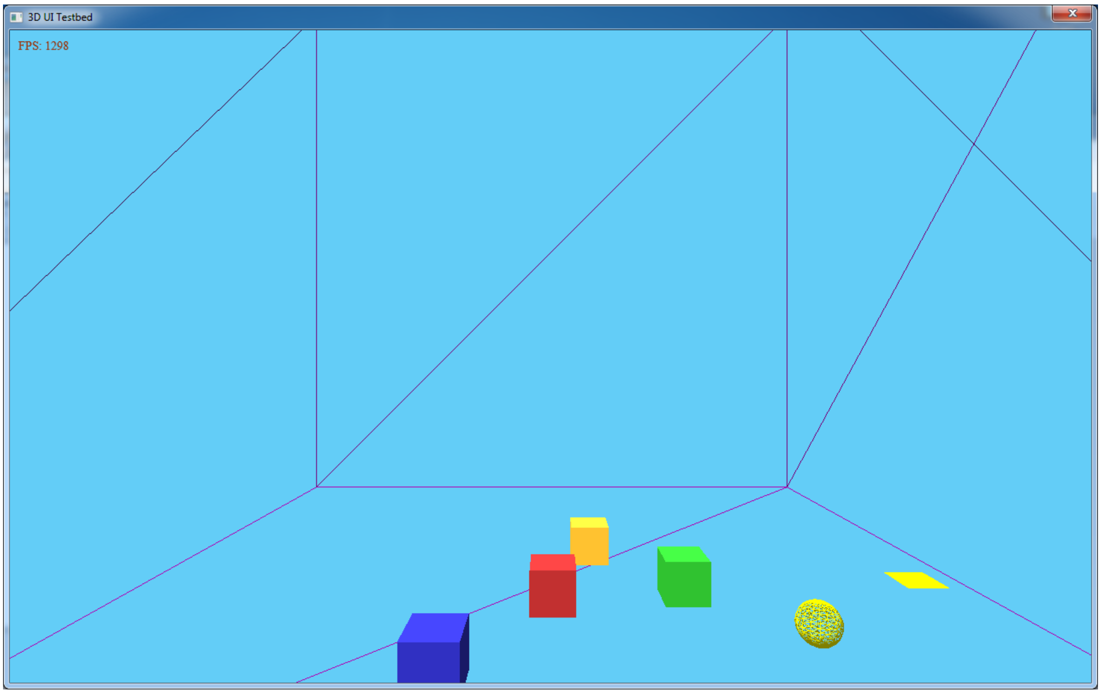

Figure 11. 4: Screenshot for Task 2.2 of Experiment-1.

Below are the screenshots from Experiment-2.

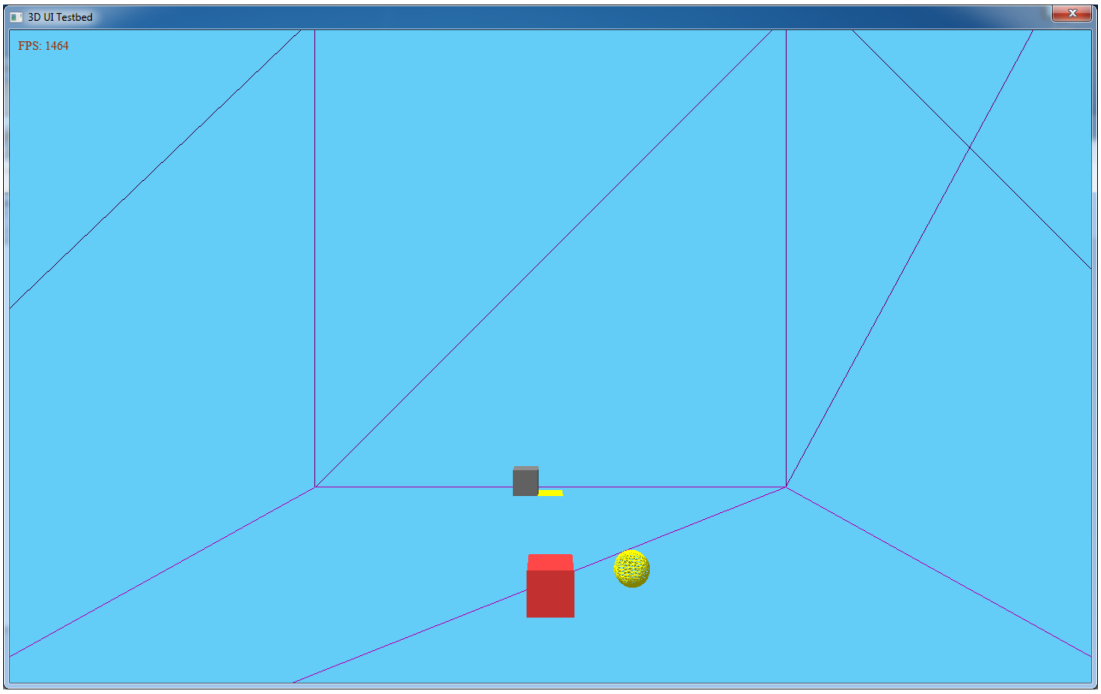

Figure 11. 5: Screenshot for Task 3.1 of Experiment-2.

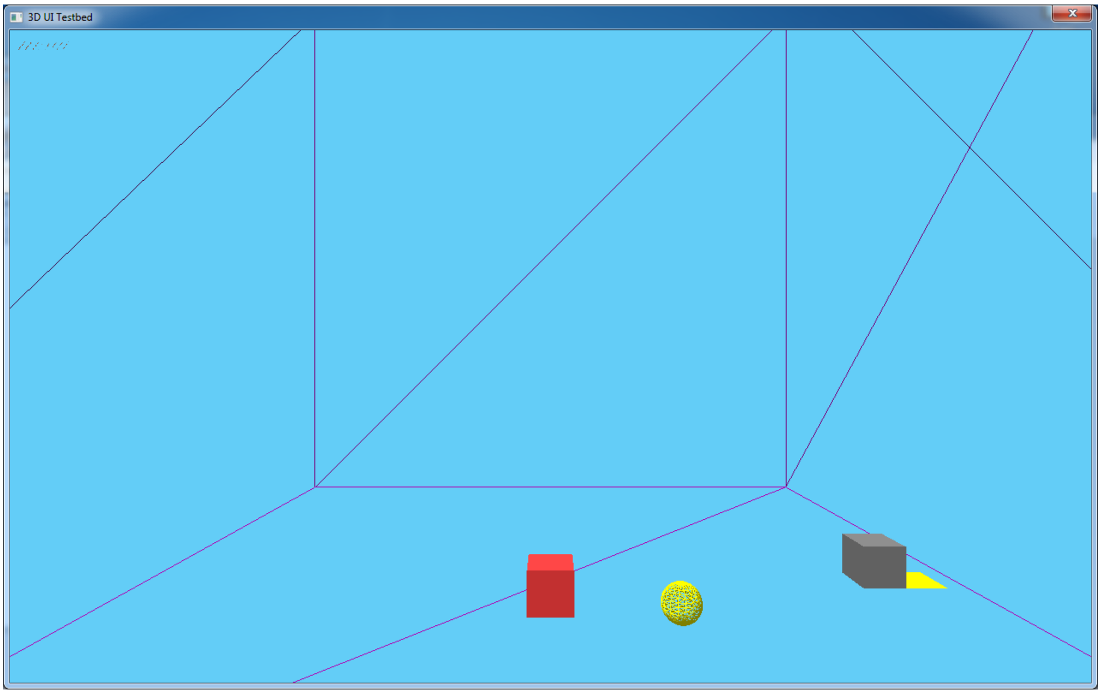

Figure 11. 6: Screenshot for Task 3.2 of Experiment-2.

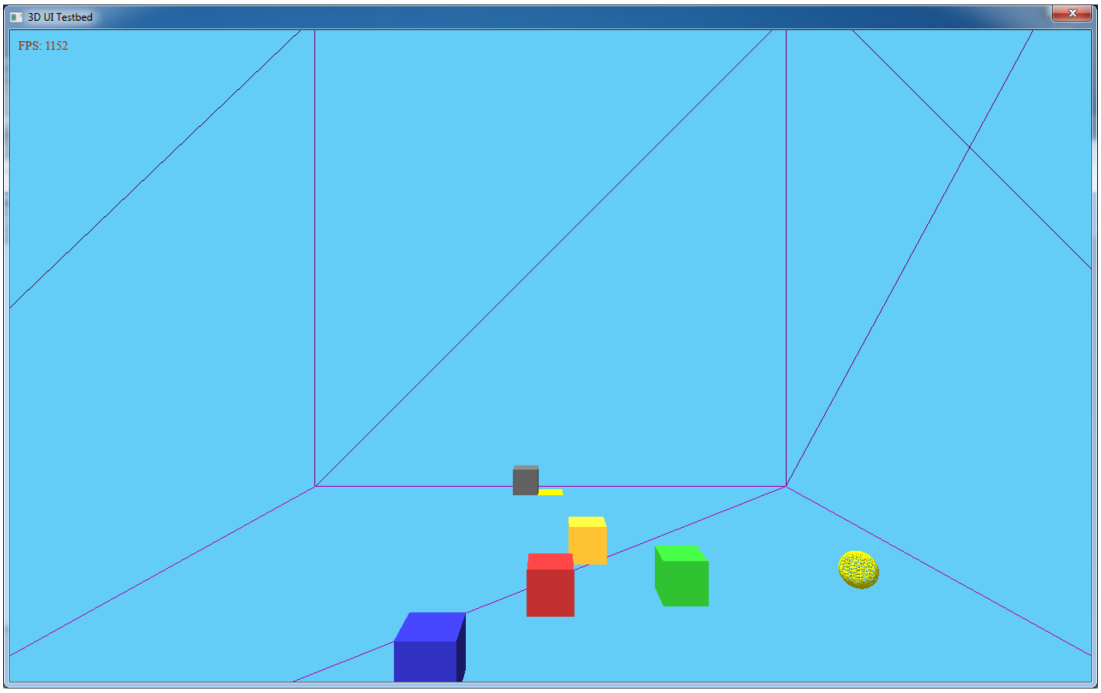

Figure 11. 7: Screenshot for Task 4.1 of Experiment-2.

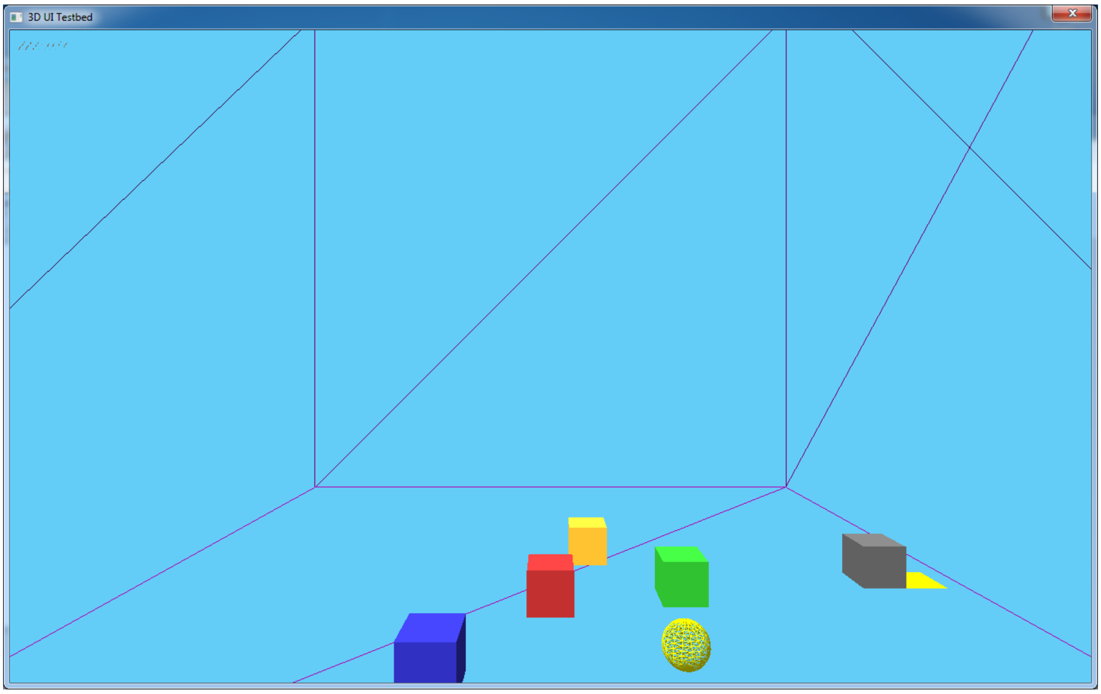

Figure 11. 8: Screenshot for Task 4.2 of Experiment-2.

Below are the screenshots from Experiment-3.

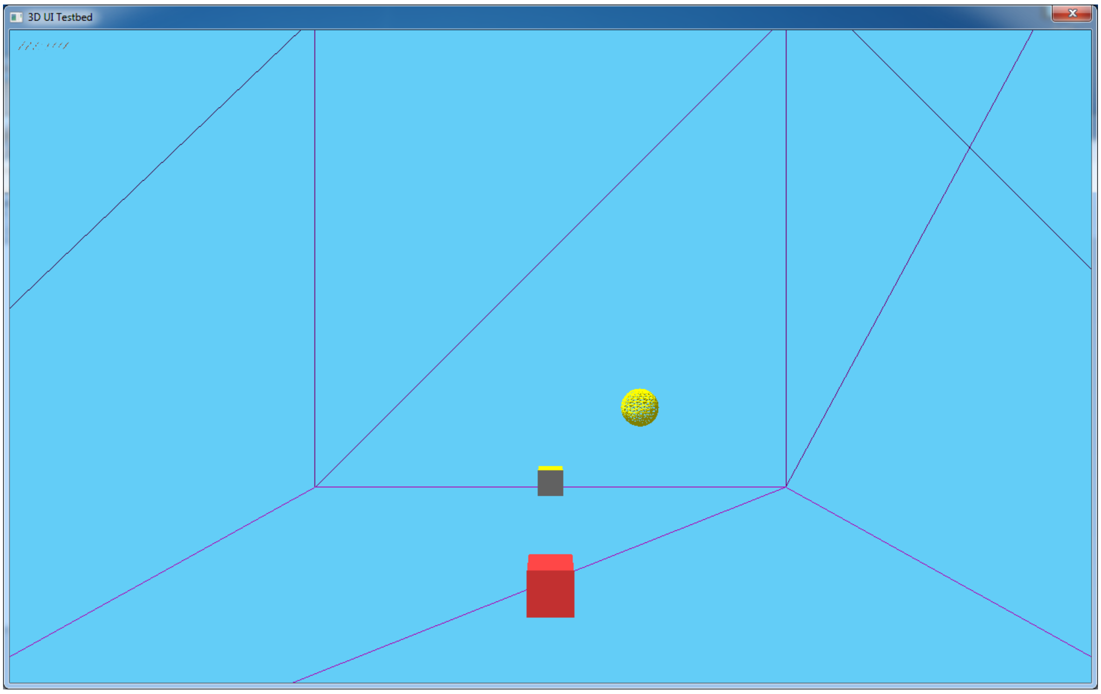

Figure 11. 9: Screenshot for Task 5.1 of Experiment-3.

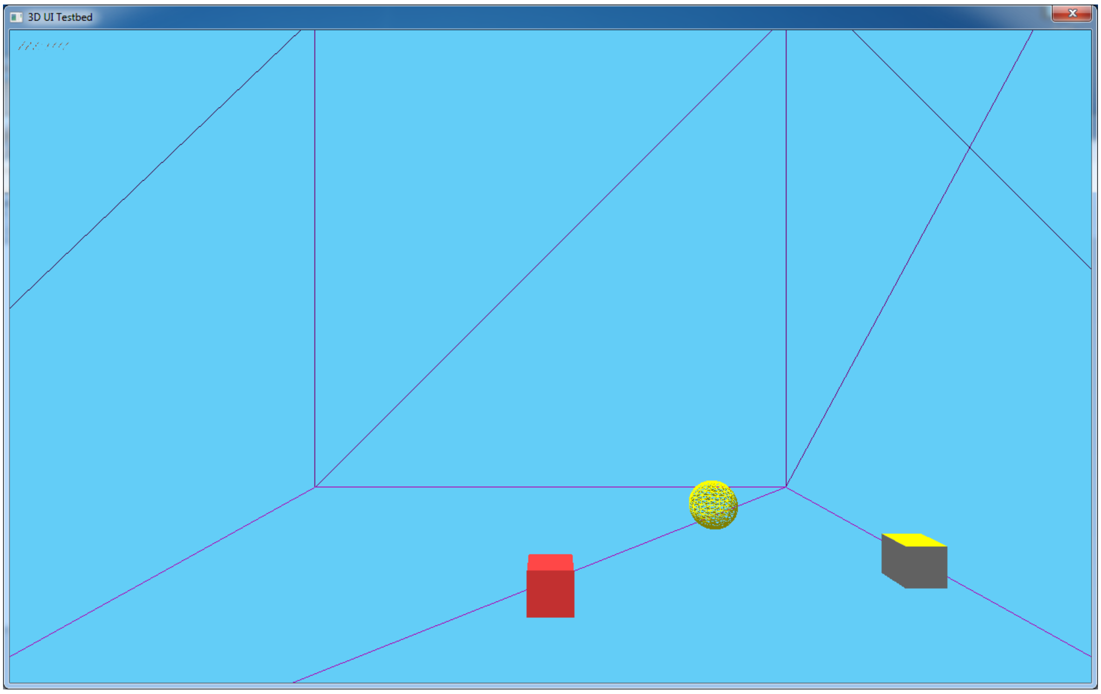

Figure 11. 10: Screenshot for Task 5.2 of Experiment-3.

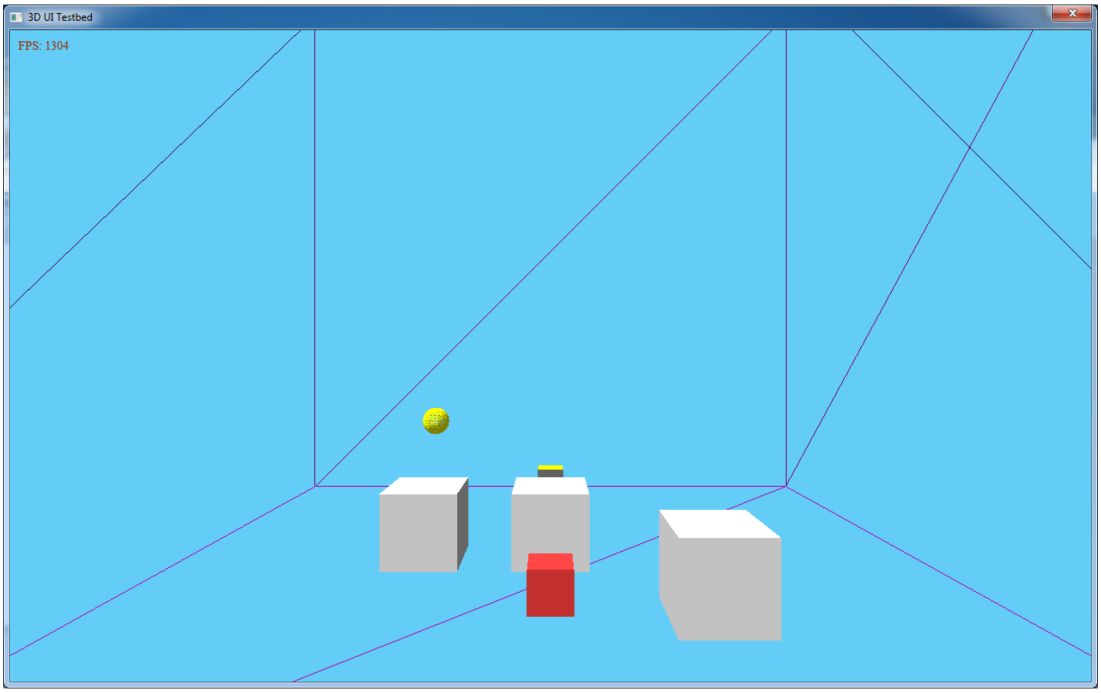

Figure 11. 11: Screenshot for Task 6.1 of Experiment-3.

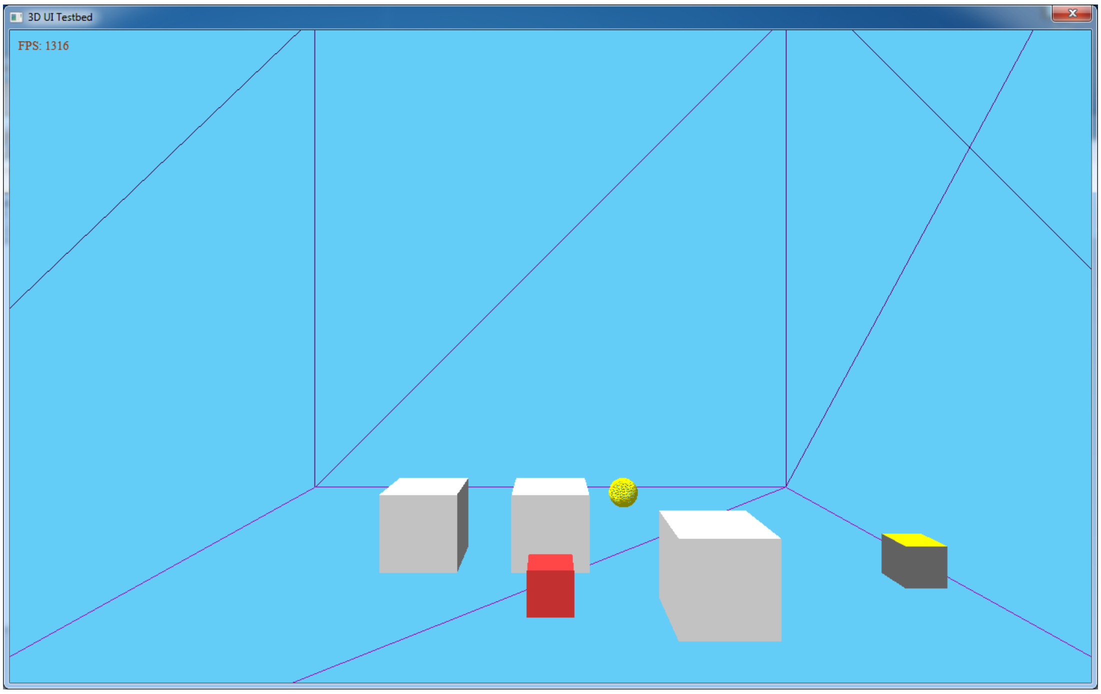

Figure 11. 12: Screenshot for Task 6.2 of Experiment-3.

Below are the screenshots from Experiment-4.

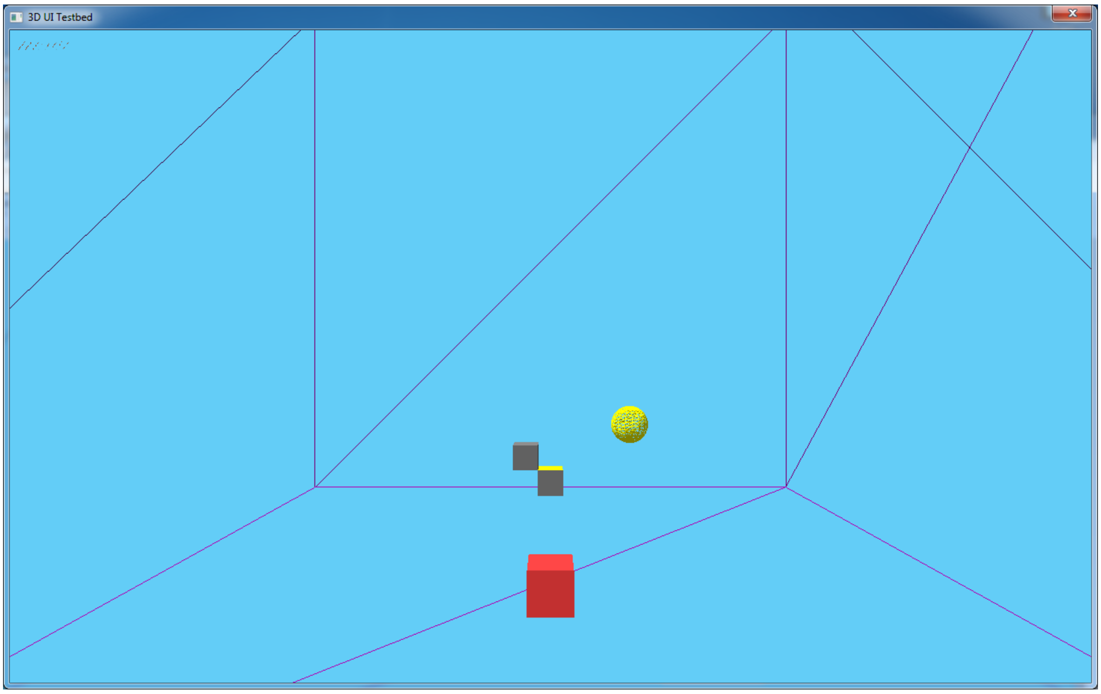

Figure 11. 13: Screenshot for Task 7.1 of Experiment-4.

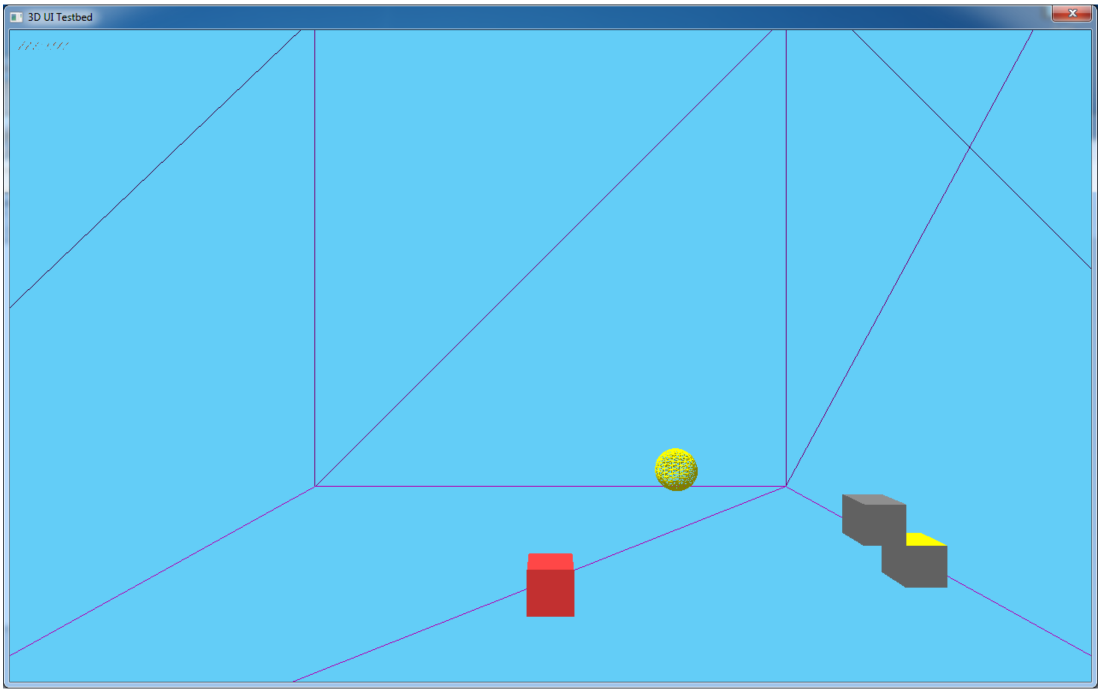

Figure 11. 14: Screenshot for Task 7.2 of Experiment-4.

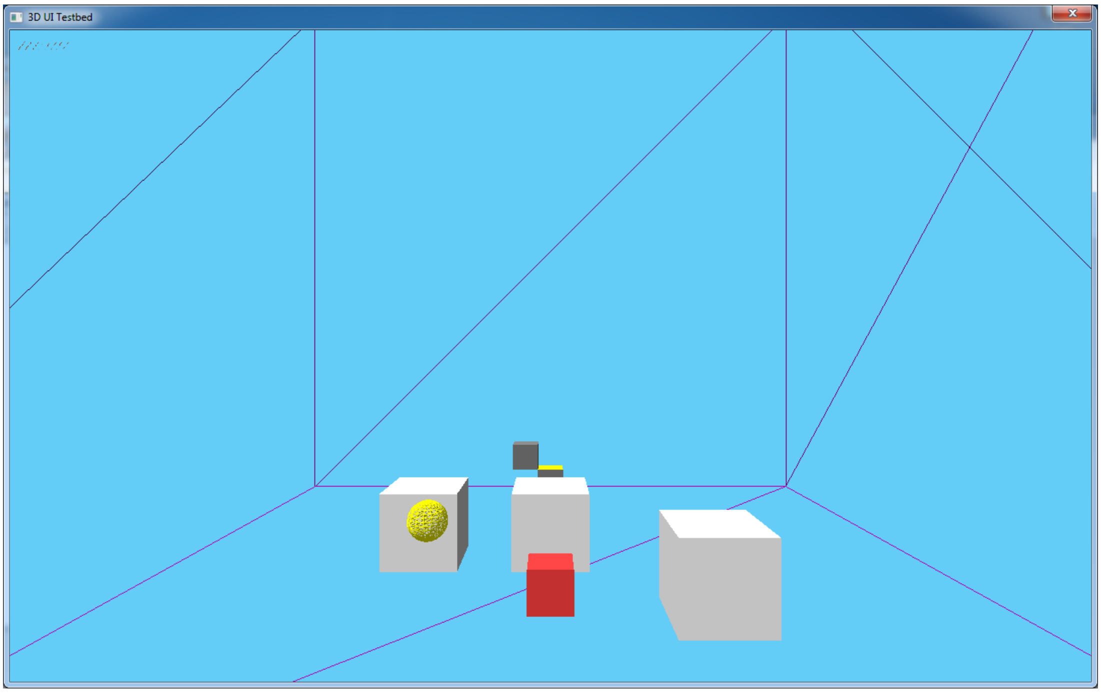

Figure 11. 15: Screenshot for Task 8.1 of Experiment-4.

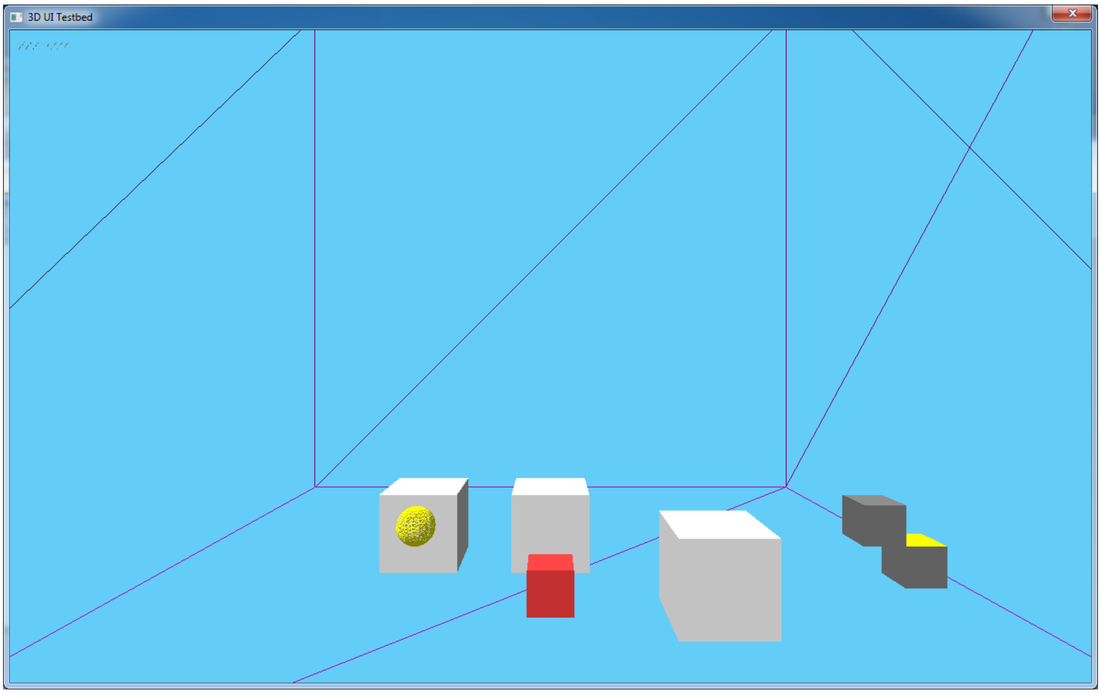

Figure 11. 16: Screenshot for Task 8.2 of Experiment-4.

Appendix C: Charts

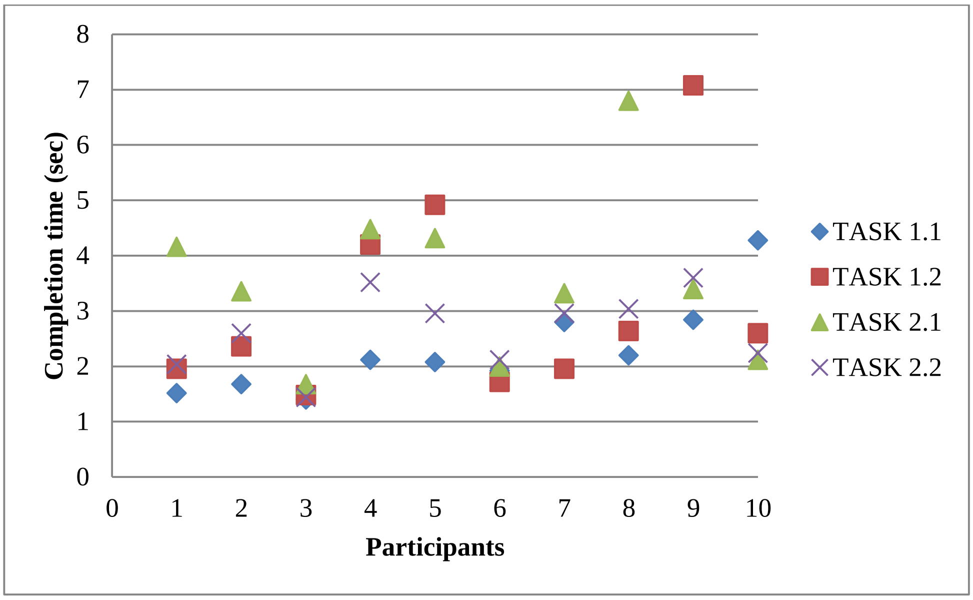

Figure 12. 1: Individual task completion times of ten participants in Experiment-1. These tasks are performed by 2D mouse input.

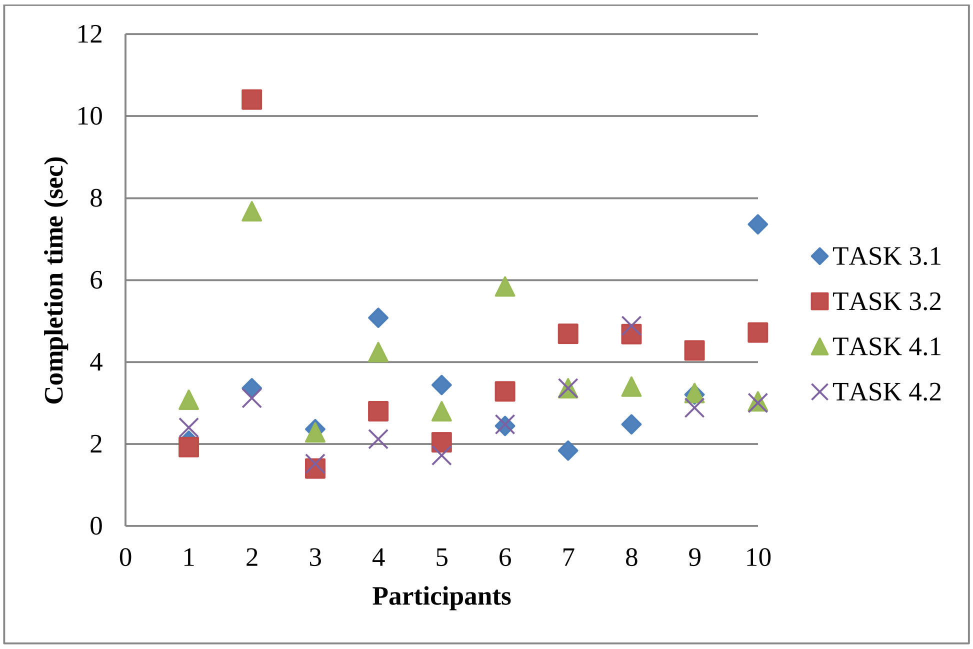

Figure 12. 2: Individual task completion times of ten participants in Experiment-2. These tasks are performed by 2D mouse input.

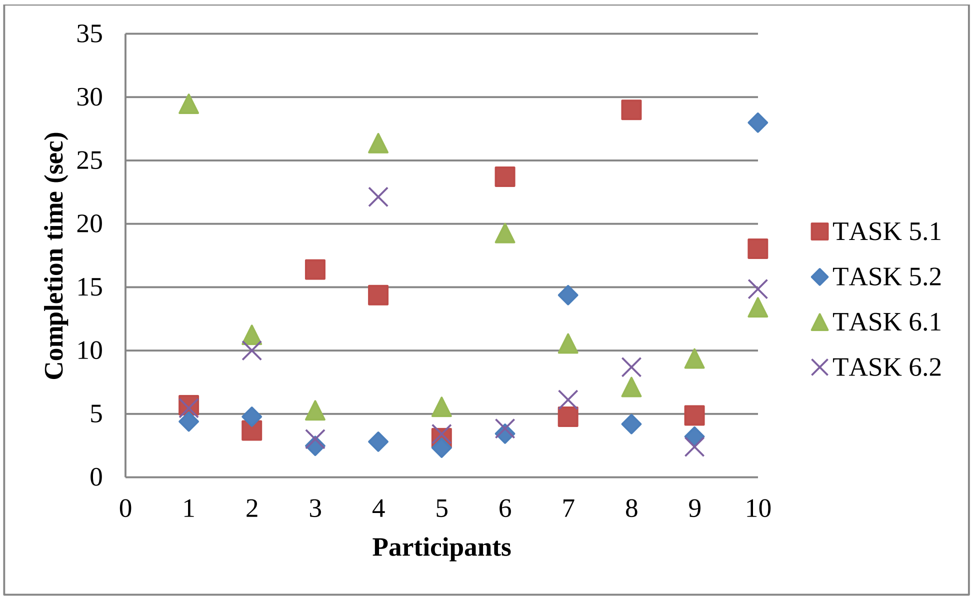

Figure 12. 3: Individual task completion times of ten participants in Experiment-3. These tasks are performed by multimodal input.

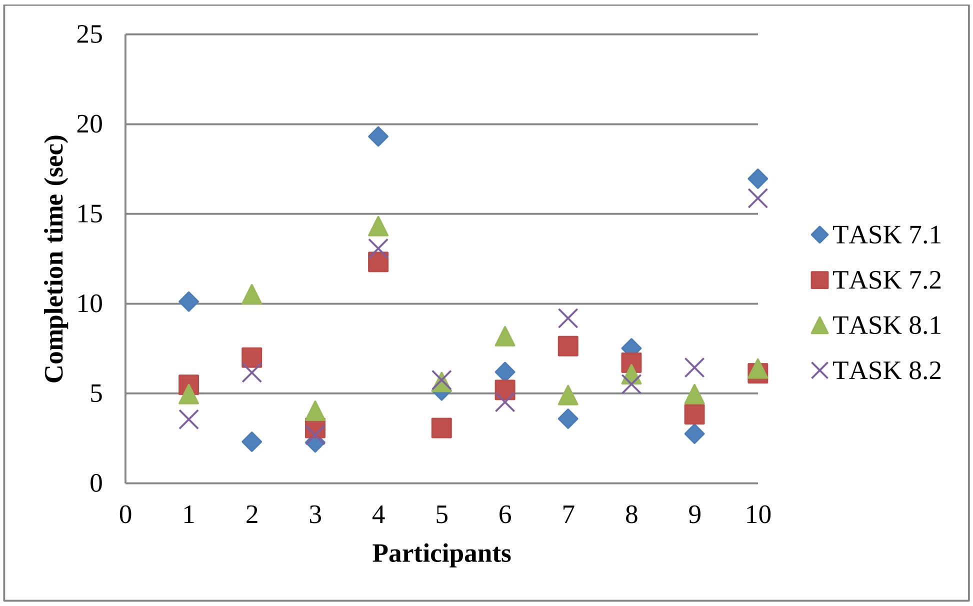

Figure 12. 4: Individual task completion times of ten participants in Experiment-4. These tasks are performed by multimodal input.