Mobile Robot Teleoperation through Eye-Gaze

(TeleGaze)

HEMIN OMER LATIF

A thesis submitted in partial fulfilment of the requirements of
Nottingham Trent University for the degree of
Doctor of Philosophy

August 2010

Abstract

In most teleoperation applications the human operator is required to monitor the status of the robot, as well as, issue controlling commands for the whole duration of the operation. Using a vision based feedback system, monitoring the robot requires the operator to look at a continuous stream of images displayed on an interaction screen. The eyes of the operator therefore, are fully engaged in monitoring and the hands in controlling. Since the eyes of the operator are engaged in monitoring anyway, inputs from their gaze can be used to aid in controlling. This frees the hands of the operator, either partially or fully, from controlling which can then be used to perform any other necessary tasks. However, the challenge here lies in distinguishing between the inputs that can be used for controlling and the inputs that can be used for monitoring.

In mobile robot teleoperation, controlling is mainly composed of issuing locomotion commands to drive the robot. Monitoring on the other hand, is looking where the robot goes and looking for any obstacles in the route. Interestingly, there exist a strong correlation between human's gazing behaviours and their moving intentions. This correlation has been exploited in this thesis to investigate novel means for mobile robot teleoperation through eye-gaze, which has been named TeleGaze for short.

The contribution of this thesis is a well designed and extensively evaluated novel interface for TeleGaze, that enables hands-free mobile robot teleoperation. Since the interface is the only part of an interactive system that the remote user comes into direct contact, the thesis covers different phases of design, evaluation, and critical analysis of the TeleGaze interface. Three different prototypes (Native, Multimodal & Refined Multimodal) have been designed and evaluated using observational and task-oriented studies. The result is a novel interface, that interprets the gazing behaviour of the human operator into controlling commands in an intuitive manner. The interface demonstrates a comparable performance to that of a conventional joystick operated system, with the significant advantage of hands free control, for a number of mobile robot teleoperation applications; provided the limitations of calibration and drift are taken into account.

Declaration

Table of Contents

Chapter 5: Multimodal TeleGaze....................106

Chapter 6: Refined Multimodal TeleGaze....................145

List of Figures

List of Tables

List of Abbreviations

BCI	Brain-Computer Interface
CCD	Charged Coupled Device
CI	Centred-Interface
EEG	Electroencephalography
EES	Evolutionary Eye Sensing
EI	Edged-Interface
EMG	Electromyography
EOG	Electrooculography
fps	frames per second
GUI	Graphical User Interface
HBS	Hybrid Bionic System
HCI	Human-Computer Interaction
Hi-Fi	High-Fidelity
HRI	Human-Robot Interaction
IUI	Intelligent User Interface
MAGIC	Manual And Gaze Input Cascaded
MDI	Mouse-Driven-Interface
OpenCV	Open Computer Vision
PC	Personal Computer
POG	Point-Of-Gaze
PTU	Pan/Tilt Unit
PTZ	Pan/Tilt/Zoom
RI	Refined-Interface
RmI	Refined multimodal Interface

List of Publications

The following publications arose from the research during the course of this project:

Hemin Omer Latif, Nasser Sherkat and Ahmad Lotfi, "Teleoperation through Eye Gaze (TeleGaze): A Multimodal Approach", *In Proceedings of IEEE International Conference on Robotics and Biomimetics* (ROBIO), 2009, Guilin, China, pp. 711-716

Hemin Omer Latif, Nasser Sherkat and Ahmad Lotfi, "Information Acquisition Using Eye-Gaze Tracking for Person-Following with Mobile Robots", *International Journal of Information Acquisition*, Vol. 6(3), 2009, pp. 147-157

Hemin Omer Latif, Nasser Sherkat and Ahmad Lotfi, "Fusion of Automation and Teleoperation for Person-Following with Mobile Robots", *In Proceedings of IEEE International Conference on Information and Automation* (ICIA), 2009, Zhuhai/Macau, China, pp. 1240-1245

Hemin Omer Latif, Nasser Sherkat and Ahmad Lotfi, "TeleGaze: Teleoperation through Eye Gaze", *In Proceedings of 7th IEEE International Conference on Cybernetic Intelligent Systems* (CIS), 2008, London, United Kingdom, pp. 6

Hemin Omer Latif, Nasser Sherkat and Ahmad Lotfi, "Remote Control of Mobile Robots through Human Eye Gaze: The Design and Evaluation of an Interface", *In Proceedings of SPIE - The International Society for Optical Engineering. Unmanned/ Unattended Sensors and Sensor Networks V*, 2008, Cardiff, Wales, pp. 71120x-9

Chapter ONE

Introduction

Many researchers in the field of robotics are more interested in controllable agents rather than fully autonomous agents [1]. This, in some cases, is due to the belief that fully autonomous agents within real scenarios are not possible yet. In most cases however, it is due to the importance of the role of human beings in many robotic applications [2], [3]. Therefore, developing the required collaboration between humans and robotic agents, which is known as human-robot interaction (HRI), is one of the remaining challenges in robotics [4].

A wide range of these controllable agents require direct and continuous controlling from a remote location. Using a master-slave mechanical manipulation and video inspection, this controlling is known as teleoperation [3]. Teleoperation therefore, as a means of providing collaboration between humans and robotic agents, remains a widely addressed topic in a variety of robotic applications. More specifically, mobile robot teleoperation is one of the promising application areas in HRI [5].

Since the human operator is located at a remote location from the robotic agent in teleoperation applications, the user interface is the only part of the system where she[1]

---

[1] A female article is used to refer to the human operator throughout this report to avoid using both articles.

comes into direct contact. The user interface includes all parts of the system that the human operator comes into contact physically, perceptually and conceptually. Therefore, a significant amount of effort has been devoted to developing different teleoperation interfaces for different scenarios and robotic applications [6].

Eye tracking, on the other hand, is entering its fourth era with a wide range of applications "*distinguished by the emergence of interactive applications*" [7], [8]. As part of the development of interactive applications, inputs from human eyes have been used in developing a number of user interfaces for human-computer interaction (HCI) [9], [10], [11]. Robotics also has its share in the advancements of this technology and a few attempts in using eye tracking in HRI are reported [12], [13], [14].

However, compared to other input channels, "*eye-gaze is a versatile option that has not been fully explored*" [15]. It is believed that "*novel interactive uses of eye trackers within increasingly complex contextual situations will allow investigation of a broader class of applications than seen in the past*" [7]. Therefore, more research on eye tracking applications for HRI is necessary as real benefits are expected from gaze-based communications. Particularly, when these systems become more able to make decisions about user intentions [16].

1.1 Research Motivations

Most teleoperation applications require the human operator to continuously monitor the status of the robot through some sort of feedback system. The feedback systems are mostly streams of real-time images coming from video cameras mounted on the robotic platform [17]. This is because natural images are believed to act better as an inter-medium between the human operator and the robotic agent [18]. Therefore, the eyes of the operator are engaged in monitoring this stream of real-time images throughout the whole duration of the operation. Meanwhile, the hands of the operator are engaged in controlling the robot using different input devices, such as joysticks. As a result, the eyes and the hands of the human operator are fully engaged in different tasks for the whole duration of the teleoperation, regardless of the complexity of the application or the interaction scenario.

Technologies create new opportunities for people to perform new activities, or to perform the same activities in new ways [19]. Eye-gaze offers the prospect of effortless communication for disabled and mainstream people alike [16]. Usable communication through eye-gaze therefore, has been a goal for many years and advancements in eye tracking have resulted in many interactive applications and novel interfaces [11].

Since in teleoperation, the eyes of the human operator are engaged in the monitoring task already, inputs from their gaze can be used to aid in the controlling task. This frees the hands of the operator, either partially or fully, from the controlling task as both monitoring and controlling are achieved through the eyes of the operator. Furthermore, driving a mobile robot might be the most intuitive task to be implemented through human eyes because "*people mostly look where they want to go*" [20].

Reducing the amount of body engagement in any HRI application, including teleoperation, provides many other opportunities for the human operator to deal with. Hands-free interfaces for robots, in particular mobile and floor mounted robotic systems, are becoming a hot topic of research in the field of robotics [21]. Therefore, the research motivation mainly lies in the correlation between human eye movements and their moving intentions to enable hands-free mobile robot teleoperation. Based on this correlation, natural interpretations of human gazing behaviours into controlling commands can be obtained. If successful, this frees the hands of the operator from the controlling task and enables hands-free mobile robot teleoperation through eye gaze, which has been named *TeleGaze* in this research.

1.2 Problem Statement and Research Challenges

The recognized correlation between human's gazing behaviours and their movement intentions is highly promising for TeleGaze. However, eyes have naturally evolved as input channels and not as output channels. Therefore, they are better functioned in perceiving information and not producing it. Using eyes to perform both monitoring and controlling simultaneously poses a number of challenges. The fact that gaze cannot be reliably controlled by intention in dynamic environments is one of the known challenges [22], [23].

Some natural characteristics of human eyes, such as the one degree pointing precision due to constant micro saccades and the fovea size, shows that eyes have not evolved as manipulation tools [24]. Using them as manipulators is likely to create extra load on the human operator. Therefore, one of the big challenges facing TeleGaze is achieving natural and intuitive interpretations of the gazing behaviours of the human operator without posing additional task load. Also the interpretations of this behaviour must be reliable enough to be used as an alternative controlling mechanism for mobile robot teleoperation.

Furthermore, remote control poses a number of challenges in comparison to the actual physical presence in the scene. That is why if any form of actual presence can be achieved, such as through specially equipped glasses or control rooms, then direct control is preferred over remote control [3]. This is regardless of the nature and the amount of the feedback information that is provided to the human operator. However, in remote control through video cameras, the extra capabilities of some cameras such as pan, tilt and zoom (PTZ) help in reducing the magnitude of the challenges.

On the other hand, the state of the art of the eye tracking technology poses a number of challenges too. Although advancements in the eye tracking technology is ongoing and commercial eye tracking systems are getting more accessible, there are still a number of engineering challenges accompanying such systems. Examples of such engineering challenges are eye tracking failures due to blinks, eye moisture and eye squinting. Also the amount of likely noise that exist in the eye tracking data due to hardware inaccuracies and micro eye movements poses extra challenges [25]. Since the eye tracking system is a main part of the TeleGaze system, the limitations of the eye tracking technology are likely to cause difficulties for TeleGaze.

1.3 Research Question

One of the first considerations in conducting any research is the research question. Formulating the research question properly should lead the research towards a good design [26]. Encouraged by the research motivations, but at the same time, held back by the research challenges, the main research question has been constructed as follows:

Is mobile robot teleoperation through eye-gaze possible?

If yes, then how achievable is it in comparison to conventional means of mobile robot teleoperation considering the research challenges? If not, then what are the reasons that prevent it while a number of motivation factors exist?

As it can be seen, the research question falls into two parts. The first part seeks the answer to the possibility of driving[2] a mobile robot from a remote location using eye-gaze data. This part implicitly declares that no other interaction modes, such as haptic or verbal, are to be used as means for mobile robot teleoperation in this study.

The second part however, depending on the answer to the first part of the question, goes in one of two possible directions. If the answer to the first part of the question proved the possibility of mobile robot teleoperation through eye-gaze, then the second part of the question inquires about the level of this possibility. This can be expressed as the usability of eye-gaze for mobile robot teleoperation. If the answer to the first part of the question showed otherwise however, then the second part of the research question inquires about the reasons behind the impossibilities of using eye-gaze for mobile robot teleoperation.

1.4 Research Aim, Objectives and Target

To answer the first part of the research question, the research aim is to design a novel interactive interface that enables mobile robot teleoperation through eye-gaze (TeleGaze). To achieve this aim, the objectives of the research more specifically are:

- To investigate the natural correlation between humans' gazing behaviours and their moving intentions in order to design an intuitive[3] interface for TeleGaze.
- To design a two-way communication channel between the human operator and the robotic agent that provides the operator with feedback information and the

---

[2] The term "drive" is used in this context to refer to the set of actions that the robotic platform performs in order to move from one point to another. An appealing alternative to this term is "navigate", which has technical meanings that do not fit the context of TeleGaze. Navigation requires three fundamental competences which are: self localization, path planning, and map interpretation (p95, [27]).

[3] In this research "intuitive" is defined as "easy to learn".

robot with necessary commands using the same interface.

- To enhance adequate feedback presentation and access to controlling commands to provide sufficient control over the robot from a remote location that enables sufficient mobile robot teleoperation.

- To provide further control over the on-board cameras and certain elements of the interface itself through eye-gaze in addition to controlling the robotic platform and its locomotion.

- To design a platform-independent and application-independent interface that can be integrated into any eye tracking systems and mobile robot platform in a wide range of teleoperation applications.

- To design an interface that complies with certain HCI and HRI heuristics and design principles without compromising the naturalness, intuitiveness, or the level of control of the interface.

- To design a novel means of HRI that can compete with other means of interactions and is used by people who have the choice of using more than one means of interactions.

Any novel interfaces should be evaluated in comparison to traditional interfaces to see if the level of difference is worth it [28]. Therefore, evaluation of the TeleGaze interface needs to be in comparison to a conventional means of mobile robot teleoperation. The details of this comparison holds the answer to the second part of the research question.

Most of today's robots, including entertainment robots [29], technical robots with highly sophisticated applications [30], and even some research platforms [4], [31] come with joysticks as standard means of remote control. It is arguable whether a joystick is the most natural and favoured means of interaction or not. However, due to its widespread use, it makes a good competitor for TeleGaze. Therefore, in order to address the second part of the research question, a conventional joystick is selected as the target device for TeleGaze to meet. TeleGaze, as a novel means of mobile robot teleoperation, should meet the joystick target in terms of system performance and user satisfaction.

1.5 Research Boundaries

The main focus of this research is on the design and the evaluation of a novel interface that enables mobile robot teleoperation through eye-gaze. In order to keep the research focused mainly on this topic, a number of research boundaries has been set. The interface composes a substantial part of any interactive system and "*is not something that can be plugged in at the last minute*" (p3, [32]). Therefore, setting these boundaries does not affect the quality of the research, but rather narrows the scope of it to keep it more focused on the main topic. In order to avoid misinterpretations of the research approach and achievements, the details of these boundaries are as follows:

1.5.1 Eye Tracking and Robotic Developments

Eye tracking is entering mainstream science "*where the eye tracker is becoming less of a novelty and more of a tool*" [26]. With the advancements in the eye tracking technology, commercial eye tracking systems are becoming accessible and available in the market. The same applies to robotic platforms too, where commercially available research platforms can be found in the market [31]. Therefore, it is outside the scope of this research to develop any eye tracking equipments or robotic platforms. There are known limitations and problems in current robotic platforms and eye tracking equipments. However, these limitations are addressed in this research, as much as possible, through the design of the interface and not eye tracking or robotic developments. This is one of the main differences between this work and other related works in the field of using eye tracking for HRI.

1.5.2 Robot Functionalities

In order to test fully the capabilities of the interface, no automated capabilities such as obstacle avoidance or path planning, are integrated into the interface. This is a common approach in testing the usability of novel interfaces, where newly designed capabilities are not mixed with preowned capabilities [4], [33], [34]. Also research shows that the distribution of fixations on video images are affected by differences in the speed of camera movements [35]. To avoid the effect of different speed on the natural gazing behaviours of the human operator, no variations in driving speed are addressed in this work. Instead, a constant driving speed is set in the system and further control over the speed is not provided. These two boundaries might limit the autonomy level of the interface. However, autonomy is only one aspect of HRI [36].

1.5.3 Targeted Users

Although hands-free mobile robot teleoperation looks very promising for people with some sorts of disabilities, such as people who suffer from spinal cord injury [37], TeleGaze is not aimed at disabled people. This is because one of the objectives of the research is to evaluate TeleGaze in comparison with other means of interactions that are not necessarily used by people with disabilities. This approach reflects on the design of the interface and the reported results. It also makes the research more challenging in terms of meeting the target, since people who use TeleGaze have the choice of using other means of interactions too.

Another reason is that most eye tracking applications addressed at people with disabilities are focused on controlling wheelchairs (Chapter 2). TeleGaze requires an interaction screen to be placed in front of the human operator, which is not desired in controlling wheelchairs. The interaction screen limits the situational awareness of the wheelchair operator in comparison with looking around and perceiving the environment more freely and naturally.

1.6 Research Approach and Thesis Organization

A typical scientific research starts by identifying the research question, forming a hypothesis, testing the hypothesis, analysing the results of the test, and modifying the hypothesis based on the findings [27]. Empirical data is the only way to validate the hypothesis on novel interaction techniques [38]. Therefore, a typical design process starts by building a prototype, evaluating the prototype, identify potentials for improvements, and refining the design [27]. This process is repeated in an iterative manner until the design reaches the required level of performance and satisfaction [19].

This work therefore, follows an empirical approach to answer both parts of the research question. Through three different phases of design, evaluation, and refinements a novel interface for TeleGaze is presented. The material presented in this thesis matches the chronological structure of the actual work carried out in all three phases, which are presented in three chapters. Additionally, other materials such as literature survey and conclusions compose individual chapters on their own. To provide better idea on the thesis structure, followings are the details of the coming chapters:

1.6.1 Background and Literature Survey

While surveying the literature for related works, some terminological-inconsistencies have been found in the field of eye tracking. Different eye tracking terms have been used for the same purpose, as well as, for different purposes interchangeably without attention to the technical differences among them. Therefore, prior to reviewing any related works, some terminological-standardisations have been covered and proposed. Clarifications on the technical use of these terms are made in order to categorise the reviewed works later. This is believed to be necessary due to the lack of such information in the literature.

As far as related works are concerned, the multidisciplinary nature of HRI, eye tracking, and interactive systems makes writing a concise literature survey not a trivial task. A literature survey on each of these disciplines is beyond the scope of this thesis. Therefore, only works that are highly related to TeleGaze in that they use eye tracking for HRI, with or without an interface, are covered under the literature survey. This does not include any background information on any of the disciplines mentioned above, except when highly necessary.

Consequently, chapter two starts by covering some terminological-inconsistencies in the field of eye tracking. It moves then to reviewing the highly related works to TeleGaze and placing them in different categories of HRI. Finally, the chapter ends by specifying the gap in the literature and the need for this work in comparison to other available works in this regard.

1.6.2 Native TeleGaze

The first phase of the research is referred to as the *native TeleGaze*[4], because all interface prototypes experimented in this phase depend purely on inputs from human eyes and not any other input modalities. In this phase, two different interface prototypes for TeleGaze are designed with differences in functionalities and their overall layout. Both prototypes however, meet the objectives of the research despite these differences. Then, a group-focused observational study is carried out to find user preferences in terms of level of control, functionality, and the overall layout of each interface. Based on the findings of the observational study, a refined prototype, which combines the preferred features of the two earlier prototypes, is designed. Then, a task-oriented evaluation is carried out to evaluate the refined prototype.

In addition to the physical design, the conceptual design, and the design principles of TeleGaze, chapter three covers the first phase of the research. As a proof-of-concept, this phase mainly addresses the first part of the research question. Therefore, the answer to the first part of the research question is built based on the findings in this phase. This chapter ends with conclusions and directions for future work on TeleGaze towards answering the second part of the research question.

1.6.3 Evaluation Metrics and Experiment Design

The first phase of the research proved the concept of TeleGaze as a means for mobile robot teleoperation. This was concluded based on the findings from the different evaluation techniques that were used in that phase to evaluate the different interface prototypes. However, in order to answer the second part of the research question, more extensive evaluations are required. Therefore, a well designed set of evaluation metrics are selected at this stage of the research. The set includes testing methods, inquiring methods, and inspecting methods to obtain measurements from multiple point of views.

This set of evaluation metrics is used to evaluate TeleGaze against its joystick target in a mock-up application scenario. Therefore, the navigational task used in the usability testing experiment is redesigned at this stage of the research in order to represent better real-life application scenarios. The details of the evaluation metrics and

---

[4] The word "*native*" is used as in "*native C*" for example, which means not mixed with other inputs.

the design of the usability testing experiment are covered in chapter four. The information presented in this chapter is necessary in order to interpret scientifically the results in later chapters.

1.6.4 Multimodal TeleGaze

The results from the first phase of the research showed that depending purely on inputs from human eyes for TeleGaze is not as practical as it was hoped. For TeleGaze to compete with its joystick rival, additional input modalities are needed to increase the system's performance and the level of user satisfaction. Therefore, the second phase of the research integrates an extra input modality to TeleGaze, hence the name *multimodal TeleGaze*. Multimodal TeleGaze still depends mainly on inputs from human eyes; however, augmented by inputs from extra interaction modalities.

The findings from the first phase of the research also suggested some necessary refinements in the design of the interface. Consequently, in addition to experimenting a multimodal approach, this phase of the research carries out some major refinements in the interface. The details of these refinements, the usability testing experiment of the multimodal approach, and data analysis of the results are all covered in chapter five.

The results of the usability testing experiment show that the multimodal TeleGaze statistically meets the joystick rival. Hence, the answer to the second part of the research question is obtained by the end of this phase of the research. However, some very interesting findings from this phase of the research suggested a few more refinements in the design of the interface. Therefore, the chapter concludes by highlighting the findings and suggesting the next phase of the research.

1.6.5 Refined Multimodal TeleGaze

By the end of the second phase of the research, the research target was met and the answers to both parts of the research question were obtained. However, some interesting findings from the second phase showed that TeleGaze has potential for not only meeting, but also exceeding its joystick rival. Therefore, the third phase of the research undertakes some refinements in the design of the multimodal interface, hence the name *refined multimodal TeleGaze*.

Using the same set of evaluation metrics used in the previous phase, the usability testing experiment is carried out for the refined multimodal TeleGaze. This is necessary in order to validate the effects of the refinements on the performance and user satisfaction level of the system. The details of these refinements, the usability testing experiment, and data analysis are all covered in chapter six. Chapter six concludes by statistically ranking TeleGaze in comparison to other interaction modes, such as the joystick target. It also checks the refined multimodal interface against the design principles tailored for TeleGaze. The end of this phase marks the end of the study, as the answers to both parts of the research question are statistically validated.

1.6.6 Conclusions and Future Work

Main conclusions, originality elements of the work, and critical reviews of the work are all covered in chapter seven. Also covered in this chapter are directions for some future work, which are mainly inspired by the findings of the research. The chapter ends with some final thoughts on the research and the proposed interface for mobile robot teleoperation through eye-gaze (TeleGaze).

Chapter TWO

Background and Literature Survey

2.1 Introduction

Future robots are expected to need substantial communication and interaction skills, if they are to share their environment with their human companions. Realising this fact has led vast numbers of researchers to devote their effort to studying many aspects of human-robot interaction (HRI). The most investigated of these aspects is designing and evaluating interfaces that enable this interaction. This is not only studied by engineers and computer scientists, but also people from a variety of other different backgrounds. Therefore, HRI is well known to be a multidisciplinary subject area [39].

Eye tracking on the other hand, has been around for much longer than robotics [40]. With the aid of computers and advancements in technology, eye tracking is rapidly becoming a viable tool for studying and creating different interaction interfaces. Whether this tool is used as a diagnostic or interactive tool, a substantial range of applications can be seen starting from simple typing [41] to aiding in complex surgeries [15], [42]. Therefore, eye tracking is also considered as a multidisciplinary subject.

Due to the multidisciplinary nature of both HRI and eye tracking, writing a literature survey on any subject that combines these two tends to be highly demanding. This is in addition to the fact that the subject of interactive systems is a significantly broad subject with vast numbers of different application contexts. Therefore, the intention here is not to write a literature survey that covers each of these subjects individually and cover all related background information. Instead, only the works that are highly related to TeleGaze are covered. Extensive surveys on the individual subjects can be easily found in the literature [7], [36].

This chapter therefore, reviews all the reported works that have used, or proposed, eye tracking as a tool for HRI. This is not only limited to mobile robot teleoperation though. Prior to the reviews however, a section is devoted to addressing some terminological-inconsistencies in the field of eye tracking. This is believed to be some necessary background information in order to better understand and distinguish between the used approaches in the reviewed works later. All other necessary fundamental and background information is covered throughout the thesis where relevant and needed.

2.2 Eye Tracking Data Types

The diversity of eye tracking algorithms and techniques has resulted in creating different types of eye tracking data. Different types of eye tracking data are obtained differently and therefore, are used differently. Where one type of eye tracking data is useful, another type might be not. Also where one type can be obtained with a specific algorithm or equipment, another type cannot be. The type of eye tracking data obtained has been mainly determined by the different generations of eye tracking equipments. Duchowski [26] classifies eye tracking equipments into four generations as follows:

- **First generation:** consisting of search coil or electro-oculography techniques.

- **Second generation:** consisting of photo- and video-oculography techniques.

- **Third generation:** analog video-based combined pupil/corneal reflection.

- **Fourth generation:** digital video-based combined pupil/corneal reflection.

Different eye tracking data types and different levels of accuracy can be obtained with the different generations. For example, only eye-in-head measurements can be obtained with the first two generations, while line-of-sight measurements can be obtained with the last two. In second generation systems "*eye movement analysis relies on off-line, frame-by-frame visual inspection of photographs or video frames*" and easy calculations of line-of-sight is not allowed [26]. Also higher tracking accuracy can be achieved with the fourth generation than the third generation due to advancements in digital image processing and computation power.

Despite known differences in eye tracking data types, a noticeable terminological-inconsistency can be seen in the literature. In some cases, different terms are used to refer to the same type of eye tracking data. While in other cases, one term is used to refer to different types of eye tracking data. Terms such as *eye-gaze* tracking, *eye-movement* tracking, and *eye-ball* tracking are used in the literature interchangeably. For example, an article that "*deals with the experimental results of the accuracy of the estimation of the rotation angle of the eye ball*", uses "*gaze*" throughout the article to refer to eye-ball tracking [43]. It mentions that "*the gaze direction is expressed by the horizontal angle of the gaze, and this is derived from the triangle formed by the centres of the eyes and the nose*". In this example, the position of the eye ball in the eye socket has been tracked, while it has been misunderstood for eye gaze-tracking and this term is used to refer to the technique.

In another example, the technique is defined as the actual line-of-sight and not the movement of the pupil within the eye, but eye-movement is used throughout the work and not eye-gaze [44]. Elsewhere, eye-gaze and eye-movement are used interchangeably in the same article to refer to the same technique. Quoting from [12] for example, "*turn left: when the user moves his/her eyes towards the left*" and "*turn right: when the user gazes towards the right*". In this context, eye-gaze tracking is the actual data type meant by both quotes. However, eye-movement is used without attention to the technical differences between the two terms.

This terminological-inconsistency might be due to linguistic preferences, or lack of linguistic knowledge. Alternatively it might be due to lack of attention to the differences in the technical meanings of the terms. Regardless of the actual reason, the differences in the used terms require more attention and technical clarifications. To the best of the author's knowledge, such clarifications has not been reported in the literature. Therefore, the followings are the definitions and the technical clarifications of the commonly used terms in the field of eye tracking:

2.2.1 Eye-Ball Tracking

Limited by the equipments available, early stages of eye tracking technology were limited to obtaining eye-in-head measurements using techniques such as scleral contact lens or electrooculography [26]. Eye-in-head measurements provide information on the location of the eye ball within the eye socket, hence the term eye-ball tracking is the best fit. In eye-ball tracking, the eye socket is divided into a number of coarse regions such as up, down, right, and left. The data obtained then, is limited to which region contains the eye ball at any time. The level of details of the data therefore, depends on the number of regions, which is not as detailed as some other types of eye tracking data. However, even with the advancements in the eye tracking equipments, eye-ball tracking is still experimented in research to the present day [45], [46].

2.2.2 Eye-Gaze Tracking

Using more advanced techniques and algorithms than the ones used for eye-ball tracking, this type of data provides information about the projected point-of-gaze (POG) of the subject. The information is provided in the form of (x,y) coordinates of the POG of the subject on the interaction screen. Hence, the term eye-gaze tracking is the best fit for this type of eye tracking data. The first two generations of eye tracking systems do not provide this type of eye tracking data [26]. Therefore, this type of data can only be obtained when the third or fourth generations of eye tracking systems are used. However, with eye-gaze tracking, the level of details expected from eye-ball tracking can also be obtained.

Due to the fact that more detailed information can be obtained from eye-gaze tracking than eye-ball tracking for example, this is the most widely used type of eye tracking data. It is also the most desired type for HCI and usability studies [26]. In most cases, "*our goal is to measure visual line of gaze, that is, the absolute position in space* *at which the user's eyes are pointed, rather than, for example, the position of the eyeball in space or the relative motion of the eye within the head"* [44]. Therefore, eye-gaze tracking is the goal for newly developed eye tracking algorithms and the reason behind the birth of third and fourth generations of eye tracking systems.

2.2.3 Eye-Movement Tracking

The change from one reading to another, whether in eye-ball tracking or eye-gaze tracking, creates a different type of information that is used in some applications [47]. The sequence of the readings and/or what happens between two consequent readings reveal information on the movements of the eye. Therefore, eye-movement tracking is the best term to describe this type of eye tracking data.

Eye-movement tracking depends on either eye-ball or eye-gaze tracking. Therefore, the obtained eye-movement tracking data also includes the eye-ball or eye-gaze tracking data. In order to be more explicit, eye-ball movement tracking or eye-gaze movement tracking can be used to refer to this type of eye tracking data. However, eye-movement tracking is the most widely used term, although used interchangeably to refer to either eye-ball tracking or eye-gaze tracking occasionally.

2.2.4 Eye-Gesture Tracking

The work reported in [11] introduces "*gaze gestures"* as a novel way to direct computers by eye-gaze. Gaze gestures are based on eye motions instead of fixations and dwell-time. The gestures consist of a sequence of strokes that are performed in a sequential time order. The claimed advantage of gaze gestures is that the gestures are immune against calibration shifts and insensitive to accuracy problems. This is because the gestures are not used for pointing to a particular region or zone for example. Also another mentioned advantage is that the number of commands can be increased by increasing the list of gestures by designing new ones.

The main question in using gaze gestures is the level of complexity of the gestures that people can perform. To separate between the gestures and the natural movements of the eyes, the gestures are distinguished based on time elements. Due to the novelty of the concept, more evaluations and experiments are needed to obtain a clear idea about the usability of this type of eye tracking data. Therefore, eye-gesture tracking is not as common as eye-ball, eye-gaze, or eye-movement tracking data types.

2.2.5 Eye Tracking

Similar to the use of the other terms, eye tracking is used interchangeably with the above terms to refer to the same eye tracking data types in the same context. However, unlike mixing between the other terms, using eye tracking to refer to any particular eye tracking data type is accepted, both linguistically and technically. Therefore, eye tracking can be used as a general term to refer to any one of the eye tracking data types. In order to indicate the nature of the data obtained and processed, using more specific terms such as eye-ball tracking or eye-gaze tracking is proposed rather than using general terms such as eye tracking.

2.2.6 Eye Tracking for TeleGaze

Considering the nature of TeleGaze and the requirements of the interface, eye-gaze tracking is believed to be the best choice. With eye-gaze tracking, the projected POG of the subject on the interaction screen can be obtained. The interaction screen in TeleGaze is where the interface displayed to the user and the interaction takes place. The requirements in this case, is the projected POG of the subject on the video streams displayed through the interface. This information is only obtainable using eye-gaze tracking and not other types of eye tracking data. Eye-ball tracking for example, does not provide sufficient details and accuracy in terms of the POG on the TeleGaze interface. Eye-movement tracking on the other hand, is rather too much detail that is not necessary for TeleGaze. Also eye-gesture tracking is too complicated for TeleGaze and is likely to affect the naturalness and intuitiveness of the interface.

Throughout this thesis, eye-gaze tracking and eye tracking are used interchangeably to refer to the data type used in TeleGaze. However, other terms are used when referring to related works, such as in the coming sections of this chapter. Regardless of the used term in the original resource, the terms that best describe the eye tracking data type in the cited work are used. For example, if a work actually uses eye-ball tracking data type, then eye-ball tracking is used whether this term, or other terms such as eye-gaze or eye-movement tracking, are used by the authors.

2.3 Eye Tracking for Robotic Applications

Eye tracking applications, whether in HCI or HRI, can be categorized as either *diagnostic* or *interactive* [26]. In diagnostic applications, the eye tracking data is used to obtain objective metrics of the visual attention processes of the subject. In interactive applications however, the applications are expected to change or respond to the user's gaze. Interactive applications can be categorized as either *gaze-contingent* or *selective* applications [26]. Gaze-contingent applications manipulate the display depending on the eye tracking data, for example to solve bandwidth or resolution problems. Selective applications use the eye tracking data as an input device, for example similar to the conventional computer mouse. TeleGaze lies in this category in the sense that it uses the eye tracking data to substitute a conventional input device namely the joystick.

Eye-gaze as an input control device has been explored extensively within the fields of assistive technology and alternative interface design in HCI [22]. In assistive technology, eye-gaze has been mostly studied to provide real-time communication and interaction for disabled people [40]. In HRI, the role of eye-gaze has been investigated widely, with overall better achievements when gaze included in the communication in addition to other modalities [48]. Similar to HCI, a significant amount of these works in HRI is devoted to disabled people. Very few works which are not addressed to disabled people, such as using eye-gaze in robotic surgery to help surgeons, can be seen.

Followings are the reviews of the works that use eye tracking in *interactive* HRI applications, which are most related to TeleGaze:

2.3.1 Eye Tracking as a Controlling Tool in Local HRI

Eye tracking as a controlling tool in local HRI interactions is mostly investigated in wheelchair controlling applications. However, other controlling applications are reported too. The following section reviews the works that investigate eye tracking in wheelchair controlling applications. All other works that use eye tracking in local HRI are reviewed in the section after.

a. Wheelchair Controlling Applications

In an attempt to develop interfaces for wheelchair users, variety of approaches have been applied and experimented, such as using forehead bio-signals [37] or electrooculography (EOG) [49]. Using forehead bio-signals in [37], a custom-build sensory head band with embedded three electrophysiology sensors for data acquisition has been used. Using these sensors, the head band provide five distinctive face movements, including eye closing as a face movement. In this case, eye tracking information is used partially due to the fact that the information is limited to whether the eyes are closed or open.

In [33], a CyberLink system is used in order to generate control commands. The system is a small wearable device that acquires electromyography (EMG) and electrooculography (EOG) signals from three sensors on a headband. Those signals also are used to switch between control and non-control modes of the system. The user's eye movement and head movement constraints have been freed by limiting the use of EOG to periods of moving forward only. EMG-click, which is used to refer to frequent EMG on/off signals, is used mainly for directional control. EOG detection on the other hand, is used for speed limitation only because it is less responsive in comparison with EMG due to limitations in the placement of the electrodes. Thresholds for both signals are set based on individual subjects to achieve good, as it is reported, performance.

To switch between different control states, a non-intuitive algorithm based on the number and sequence of EMG-clicks for each particular control command has been used. For example, to switch from the “stop” states to the “left” control states the user has to perform one EMG-click to get a command window, and then two other EMG-clicks to choose the “left” states. To switch back to the “stop” states the user has to perform one EMG-click. It is even more complicated to switch between control and non-control states as the user has to enter a three digit password to switch from the non-control states, which has been called password states, to the control states. Digit entries for the password window is performed through the same EMG-clicks in a similar sort of technique to switch between control states explained previously. Very interestingly, in addition to the complexity of entering the password, it has been reported that still there is a possibility that the user accidentally inputs all password digits, total of three, without intention to do so.

Some experimental results limited to time-to-complete a task have been included in the above report. However, no further evaluation of the system in comparison with alternatives or details of the experiment in terms of evaluation metrics and details of participants have been reported. It is not mentioned either whether the participants who performed the task had any disabilities or not. It is claimed that the system is "*easy to setup and easy to use*" but reported that "*a new user might require about half an hour of practice with a simulator*" before actually trying the system. It has been reported that the system is not as natural as other means of control such as eye movement. The advantage of the system however, has been highlighted as allowing the user to look around and observe the surroundings while driving the wheelchair.

Presenting an EOG based wheelchair control system with an active obstacle avoidance capability for hands free control, a novel way of mounting electrodes for detecting EOG signals is used in [49]. EOG is a popular solution for detecting eye directions, which are measured based on the steady corneal-retinal potential. Instead of mounting sensors around the forehead and/or parts of the face of the user, an eyeglasses is used. The idea behind using the eyeglasses is to simplify the use of the system and to increase the willingness of potential users to use the system. The used EOG eyeglasses module is reported to be more convenient to setup and more compact when compared to the cutting strips of adhesive tape-holding solution.

The above proposed EOG control algorithm is evaluated in a 4.5m x 3.0m indoor area. A marker used to draw the real path of the wheelchair while being driven. The reported results show that training increases the performance of a junior volunteer to generate smoother path and driving experience with obstacle avoidance. No further evaluation results or usability experiments are reported in this paper. Also, it is mentioned in this paper that EOG signals are dependent on individual subjects. However, the reported results are collected from only one subject.

The novelty of the above work is in the way of mounting the electrodes using the eyeglasses approach rather than conventional approachs. In terms of the interface, it is similar to other reported works in this regard. Also it is one of the very few works that uses automated obstacle avoidance while testing a controlling interface. This automated capability is not preferred in most cases as it overrides the functionalities of the interface. Hence, the evaluation of the interface might not lead to deep results as far as the usability is concerned.

To develop wheelchair guidance strategies for assisted mobility, Barea and colleagues use EOG to obtain eye tracking information [45], [46]. Using eye-ball tracking, "*where the control is actually affected by eye movements within the socket*", different controlling strategies are proposed and commented. Two different approaches are experimented, where in one of the approaches an interaction screen is placed in front of the user, while in the other approach there is no interaction screen. Two different commanding strategies are experimented in the first approach, which are named "*direct access guidance*" and "*scanning guidance*". Regardless of the complexity of these strategies, particularly the scanning guidance strategy, only the basic forward, backward, right, and left driving commands are provided. This is mainly due to the limited accuracy of eye-ball tracking in comparison to eye-gaze tracking as covered earlier in section 2.2.

In the above approach, the screen placed in front of the user blocks the user's vicinity since it is not a transparent screen. Therefore, in the second approach, where the screen is removed, the commands are extracted based on the position of the eye-ball in the eye socket. In this approach, looking up moves the wheelchair forward, looking down moves the wheelchair backward, and looking right/left moves the wheelchair to the right/left. Despite the unnatural interpretations of the eye movements, such as looking down to move backward, only the four basic driving commands are provided in this approach too. A very basic navigational task has been carried out with all the different approaches. No detailed evaluation results have been reported though. However, the researchers conclude that when the users have the option of using a joystick, they prefer it more than using eye tracking to control the wheelchair. This conclusion is not based on comparative experiments for both the joystick and the eye-driven interfaces, but based on users' previous experience in using joysticks. Also no results of any objective metrics have been reported.

The work in [50] presents another wheelchair controlling interface, where eye movement is detected by processing EOG signals. To detect vertical and horizontal eye movements, electrodes are placed around the eyes in order to create micro potentials which are known as EOG signals. This signal varies for different individuals. Therefore, a trainer module to learn EOG signal level for each individual has been used. The fact that it has been reported that even without training for specific users, the system performed well, eliminates the need for the trainer module. It is claimed that simple pattern matching is used to detect and classify the eye movements. Associations of eye movements to certain robotic commands however, are far from being simple. To decide the command given by the user, a module called processor module, is used to identify the order of consecutive positive and negative pulses. For instance, to issue a turn left command for example, the user has to move the eye from the centre position to extreme left position and return it to the centre position without delaying. To go forward, the user has to move the eye from the centre to extreme upward position and bring it back.

An error probability of 1% and a command missing probability of 3% has been reported when testing the above system for more than one thousand eye movements on aged healthy people. However, no navigational tasks performed using this system have been reported. No comparison evaluation results have been reported either. In addition to the fact that the system has been developed for disabled people, but it has been tested on people without any disabilities.

To control a powered wheelchair, an optical-type eye tracking system is used in [51]. Pupil-tracking goggles equipped with a video CCD camera and a frame grabber is used to analyse a series of human pupil images when the user is gazing at an interaction screen. A graphical user interface (GUI), which is displayed on the interaction screen, is divided into nine command zones. Only four zones out of the nine command zones on the interface generate motion commands. The other five command zones are called "*idle*", where gazing at them does not generate any commands. The distribution of the commands on the interface are similar to [45], in the sense that moving forward is by looking upward, moving backward is by looking downward, and turning right/left are by looking right/left. The eye-gaze tracking data is used to place the cursor in the desired command zone on the GUI instead of the computer mouse. No evaluation or navigational-task experiments are reported. However, the researchers have concluded that "*the vision-controlled wheelchair is not easy to control*".

In an interesting attempt to implicitly distinguish between intentional and non-intentional gaze behaviours, Bartolein and colleagues have used a set of Hidden-Markov-Models (HMM) to estimate the user's current gaze state [52]. The complexity of their algorithm seemed necessary due to the lack of an interaction screen, and hence lack of regions of interest, to distinguish between the gazing behaviours based on being inside or outside those regions. Based on some physiological findings and previously recorded gaze data [53], a set of distinctive gaze patterns occurring during wheelchair navigation has been identified. Then, the current user's gaze state together with user's input states to generate the active motion state of the wheelchair have been used.

In the above works, a trial run of the proposed approach has been reported. It has been claimed that, due to the considerably reduced handling effort compared to traditional wheelchair control, the presented approach should find high acceptance rate among potential users. The reports do not include any evaluation results comparing the approach with the conventional wheelchair control, however. Also no details regarding the calculation of fixations and the other types of eye data have been reported. More interestingly, implicitly distinguishing between intentional and non-intentional gazing behaviour contradicts with the findings in this work, which are covered later in chapter three. However, using inputs both from the subject's eyes and other input devices, a sip-puff device, matches the findings of this work in terms of user preferences and system performances.

In [54], an "*eye-mouse*" interface is developed to control a robotic arm called KARES II, which is mounted on a wheelchair. A number of necessary tasks "*according to extensive interviews and questionnaires*" are predefined. Then eye-mouse, shoulder/head interface, and EMG signal based control sub-systems are used to perform the predefined tasks and issue necessary commands. It is reported that an intention reading experiment by "*utilizing the visual images obtained through visual servoing*" has been performed. It is assumed that "*one can show his/her intention to drink or not to drink by opening or closing one's mouth*". Users can indicate the position of an object that they want to grab and give commands to the robotic arm through the computer that is mounted on the wheelchair. A menu-driven interface has been developed that enables selecting the appropriate command from a drop down menu. The menus contain also commands for controlling a pan/tilt unit on which a stereo camera system is attached. The image taken by the camera, which is mounted on the wheelchair, is displayed to the user to select the object of interest. Once the object of interest is placed in the centre of the scene, the 3D position of the object is calculated and used by the robotic arm to locate and grab the object. Some evaluation results are reported for the interface and the hardware kit, without any particular task that has been performed with the system in a usability testing experiment.

b. Other Controlling Applications

To allow a surgeon to perform a minimally invasive surgery procedure as normal, while having access to an additional tool when required, a gaze contingent control system is developed in [42]. The system, which is a binocular eye tracking unit integrated into the stereoscopic console of a daVinci surgical robot[5], allows control of an articulated robotic device through the eyes of the surgeon. The desired location of the robot probe is set by the surgeon's fixations in 3D where the necessary inverse kinematics are calculated to direct the robot tip. This saves the need for a set of robotic actions in order to control the robotic probe. Instead, the robotic probe is directed to a goal point with coordinates equal to the coordinates of the surgeon's fixation.

Some experiments of the developed system have been reported where four markers have been easily identified by both the fixations and the robot tip. Also results from a proof-of-concept task, which a trajectory for the robot tip to follow in real time has been defined using the operator's eyes, are reported. The results show that the eye is capable of finer motions that the robot as the robot is limited by its mechanical resolution. It has been mentioned that currently it is infeasible to envisage a device which would perform direct tissue interactions. However, it has also been mentioned that there is a niche to develop instruments which would operate on a non-contact basis.

The above system also is used in [15] to prescribe 3D paths on tissue surfaces for ablation using focused energy delivery for enhancing robotic control in Atrial Fibrillation surgery. In this work, with the 3D fixation points, the surgeon is able to pinpoint specific locations on the soft-tissue surface. When the gaze-contingent control

---

[5] The daVinci surgical robot is an example of existing minimally invasive surgery (MIS) systems which allow a surgeon to interact with the operative environment through teleoperation.

system is used to prescribe a desired ablation path, a final path optimization is performed before focused energy delivery. This was found necessary because the 3D fixations collected during the path prescription also contain noise due to the natural behaviour of the eyes.

The above framework has been implemented using the same daVinci surgical robot used in [42]. The work reports experimental results for eight subjects using the framework to prescribe an ablation path. The results of the experiments are presented without further discussion on efficiency. The authors mention that, to their knowledge, this work is the first attempt in fusing human and machine vision for robotic surgery.

In a different application context, a guide system for daily life, called GazeRoboard, is proposed in by adopting a gaze-communication stuffed-toy robot and a gaze-interactive display board [14], [55]. In addition to providing voice guidance, GazeRoboard provides joint attention and eye-contact reactions based on ambient gaze tracking. Using a rather customized eye tracking algorithm, the stuffed-toy robot knows the element on the interactive board that the user is interested in. Accordingly, the robot provides guidance information on that point in addition to building eye-contacts with the user. The only robotic action driven by the gaze information then is determining the point of interest of the user on the interactive board. The work is more focused on building mutual gaze-communication between the robot and the user. The motivation for mutual gaze-communication is driven by regarding the robot's gazing behaviour as a "*kind of persuasive power*". Results of evaluations, as reported, show that "*eye contact brings the user a favourable feeling for the robot*" and "*this feeling is enhanced when eye contact is used in combination with joint attention*".

Titled "*human-robotics interface for the interaction with cognitive and emotional human domains*", the work in [56] uses eye tracking to observe the subject's line of gaze for active interaction with the cognitive and emotional human domains. The application of the human-robot interface is also presented for preliminary studies concerning new cognitive rehabilitation strategies in depression. Little details regarding the interface is included in the report. Most of the snapshot figures are of neutral, sad or happy human faces which the human companion interacts with. The work as reported, assumes "*that* *the permanence of the subject's gaze over happy or sad faces could reflect their empathy towards happiness or sadness feelings*". Experimental results are presented as far as the algorithm and the base principle are concerned. However, no evaluation results of the interface is included, neither details of the design of the interface.

In [57] and [58], a robotic system is presented that identifies and picks up an arbitrary object in 3D space based on the gaze direction of the human companion. The gaze direction of the person is determined in 3D space, which is used to identify the object of interest. A robotic arm then "*responds by picking up this object and handing it over to the person*". As reported, "*by utilizing the gaze information provided, the active vision detects when a person is staring at an object and searches the gaze line to find this object of unknown shape, size and colour*". In terms of evaluation or experiments, "*a demonstration on how the active vision gains attention through a waving hand and then continuously tracks the user's face is shown*". No further user studies or interface evaluations are reported as part of this work. The gaze tracking system used in this work is developed by the researchers for HRI applications earlier in [59].

Similar to the above gaze tracking system, in [60] a camera orientation device has been developed that can be mounted on the head of a humanoid robot to track the eye gaze of the robot's human companion. As reported, "*the long-term aim is to integrate eye tracking capabilities into the vision system that will equip the humanoid with the ability to infer the target of gaze of a human in human-machine cooperation scenarios*". The camera orientation device is developed to cope with the head motions of the human companion which helps obtaining better eye gaze information and readings. Also in another line, the work included the design of a new eye tracker. An eye tracker "*that can operate from a distance and that does not require any head mounted device*", as reported. The focus of this work is mainly the development of the camera orientation device and not a HRI interface based on inputs from human eyes. No experiments have been reported as far as HRI is concerned.

Another example of the works that mainly focus on developing an eye tracking system rather than trying an eye tracking algorithm to develop and interface and perform real tasks is the work of [61]. Using evolutionary eye sensing (EES) method, an interactive interface to operate a welfare apparatus, such as feeding device for an orthopedically-impaired individuals has been developed. However, the focus of the work is on developing different algorithms rather than actually implementing an algorithm and developing the interface. The design of an interface that is composed of nine focus zones has been reported without any experiments or evaluations.

2.3.2 Eye Tracking as a Controlling Tool in Remote HRI

In order to explore a proactive use of the gaze as part of the control interface in goal directed tasks for a future hybrid bionic system (HBS), the work in [62] has developed a gaze based algorithm to send commands to a robot. The robot tracks the gaze behaviour of a human actor and uses these observations to select which action to execute among a set of stored programs. Overlaid on a video display from the application's camera, the gaze fixations and trails are obtained using faceLAB. If the subject's gaze is in one of the landmark zones and the gaze velocity is within a threshold, then a transition on the robot state machine will be triggered. Four subjects are asked to perform four different tasks for experimental purposes. The fact that the algorithm can only be used to predict actions from a set of predefined list is a highlighted weakness. Implementation of the prediction algorithm on an actual HBS robot is not included in the work, but has been mentioned as part of future plans.

In an attempt to develop a non-intrusive gaze-driven interface for man-machine interaction based on vision techniques some preliminary work has been described in [8]. Using a simple deformable template for eye-iris, pupil localization in the image has been achieved in a decoupled fashion at a high cycle rate. Dividing the computer screen into a number of windows where each corresponding to a specific action, an action is issued if the persistence of gaze in the window exceeds a given threshold. The main focus of the work is the development of an eye tracking algorithm with the aim of tracking the position of the eye pupil in the eye socket in real-time for man-machine interaction interfaces. Although this aim is set clear in the report, no actual experiments on man-machine interaction have been included in the report. The developed interface however, has been experimented with the developed algorithm to determine the performance of the algorithm in terms of gazing at the different windows on the proposed interface.

Despite the relatively high error rate in the above work, the commands to the interface were correctly interpreted by the system in the experiment. The conditions and parameters of this experiment in terms of the number of subjects and number of trials have not been included in the report. It has been reported that "*much work has still to be done, both in the theoretical developments and in the experimental aspects*". However, to the best of the authors knowledge, no further works have been published since then that are highly related to this context and can be included here.

In [12] and [63], a real-time vision-based eye tracking system for HRI has been presented. It has been reported that the "*depending on the position and movement of the eyes, the system determines where on the display the user is looking*". However, no command buttons are placed on the GUI that is, supposedly, the interaction interface. Instead, the robot starts walking "*when the user changes his/her gaze direction from the normal position to the upward direction*" and turns left "*when the user moves his/her eyes towards the left*". Similar techniques have been used for turning right and stopping the robot. These techniques does not require a GUI, since the information is based on eye-ball tracking and not eye-gaze tracking. Although it has been mentioned that "*the main objective of this research is to establish a human-robot symbiotic community*", the focus of the work is mainly on developing the eye tracking algorithm and not any HRI interfaces. The objectives of the research are stated as "*detection of human faces*" and "*localizing the eyes*". Using vision based techniques at different stages, face detection, eye localization, and gaze estimation have been calculated and extracted. Therefore, no interface design details or usability testing experiments of the interface have been included in this work.

Due to the belief that better eye tracking algorithms and systems are needed, many works exist where the focus is on developing an eye tracking system rather than using one. The work in [64] is similar to the work mentioned above in the sense that the focus is on developing an eye tracking system, with the aim of controlling a robotic arm. A button based GUI has been presented where the user can select a number of commands to control the robotic arm. Although a GUI is presented, the experiments and evaluations are for the developed eye tracking system and not for the interface. Therefore, no real experiments have been reported that test the HRI interface. Also no comparison results for the developed system and other available systems have been reported in this work.

For the human operator to gain a sense of the remote environment surrounding the robot, a "*method of grasping visual information from the robot using 3D images*" has been presented in [65]. The method changes the line-of-sight of the humanoid robot in conjunction with the line-of-sight of the human operator, using eye tracking systems. As a result "*reduced eye fatigue when viewing 3D images*" are demonstrated. To gain a 3 dimensional sense of the remote environment, two cameras have been installed on the robot. The manipulator on the robot is divided into controlling the robot body parts and the robot locomotion. The robot body parts are controlled by the body and arms of the human operator, while the locomotion is controlled by the feet of the operator.

Eye tracking in the above application has been used instead of using a computer mouse to specify the object of interest in order to obtain the 3D image of it. Using the mouse, the power of the manipulator needs to be turned off for safety reasons. This is necessary because when the hand is controlling the mouse, the manipulator might move unpredictably. Using eye tracking to perform this selection, the hand stays free from using the mouse and hence, no need to turn the power of the manipulator off. In addition to this advantage, slight reduction in time-to-complete a task is measured when eye tracking is used in comparison with using the mouse. It has been reported that "*usage of eye tracking device will simplify the operation process and improve safety in the operation of the robot*". The three authors themselves have participated in this usability testing experiment without including the task of using the manipulator, which is the situation where the advantage of using eye tracking should appear.

Very recently in [23] the researchers developed a gaze-controlled driving interface which enables controlling a mobile robot from a remote location using inputs from human eyes in comparison to other modes of inputs such as computer mouse. The robotic platform has been built around a plastic frame using some Lego Midnstorms NXT component equipped with a webcam. The proposed interface has no visible components because the direction and the speed is calculated by the distance of the POG from the centre of the monitor. A task-oriented evaluation with five participants which each completed the task using only one input device has been conducted. The results have shown that the mouse scored the highest efficiency in terms of time-to-complete-task and highest accuracy in terms of error rate.

On the other hand, in a rather technical-report like paper, the progress and future plans of an integrated complex robotic platform where "*a lot has still to be done*" have been reported [66]. As part of integrating many individual components to this future robotic platform, the same eye tracking system developed in earlier research [20] is going to be integrated as well. This is in order to build a multi-model spatial and transactional intelligence system. The aim of the robot is to help elderly and disabled people "*cope with their living environment in an assistive technology context*". There is no clear functionality of the eye tracking in the system and in the context of the application as yet. However, apparently it is going to function in 3D environment directly with the robot and not from a remote location as it was the case in [20].

In a different publication addressing the same platform, eye tracking has been proposed to be used in combination with the head direction and body pose for a multimodal gesture recognition system [67]. The main focus of this work is the body posture, especially the upper part, recognition. But eye tracking is going to be used to determine the direction of the visual attention of the human companion when interacting with the robot. It has also been mentioned that the work, as far as this part is concerned, is still in its early stages and experiments will be carried out in the future.

Similar to these works "*a system that utilizes gaze tracking for real time robotic teleoperation that can be extended to a variety of technical disciplines*" has been proposed in [13]. No actual developments of teleoperation interfaces have been reported though. The work is mainly a review of previous works and proposing a gaze-driven interface, focusing on the advantages that eye tracking is likely to bring to HRI.

2.3.3 Eye Tracking as a Diagnostic Tool in HRI

Although different from the category of TeleGaze, eye tracking is used also in some HRI applications as a diagnostic tool and not as a controlling tool. In some applications, eye tracking information is combined with information from other sources to learn about certain behaviours of human beings. For example, in [68]. eye tracking has been used as a diagnostic tool and not as a selective tool. Elsewhere, eye tracking data combined with head tracking has been used in the making of an "*object tracking model*" [47]. In the making of this model, the human action of tracking an object by the eyes and the head is analysed. Accordingly, a model for "*Humanoid Vision*" has been developed, which implements the features of the tracking actions of the human.

The above model is implemented on YAMATO, a humanoid robot that "detects an object and determines its speed", for object tracking applications. The aim of this work is to develop an oculomotor control system for a humanoid robot that implements human object tracking behaviours. The same problem has been addressed elsewhere without directly using eye tracking information [69].

In a similar application to the object tracking model mentioned above, the use of eye tracking information has been investigated in developing an intelligent prosthetic hand [70]. The proposed hand will "*eventually guess the user's intentions and correctly grasp a series of different objects*", which are placed in front of it and are visible to the user through a monitor. The eye tracking information, together with hand position information obtained from the magnetic glove used to control the hand, has been used to know which object the user intends to grasp. The system is expected to eventually learn that "*gazing at an object and moving the hands towards that object means: I want to grasp that object*". The reported results show that "*gaze tracking significantly improves both the accuracy and compactness of the obtained models, if compared with the use of the hand position alone*" [71].

2.4 Conclusions

Eye tracking has been used and investigated in a number of application contexts for HRI, as it can be seen from the reviewed works above. It has been used as the only mode of interaction, as well as, in addition to other modes of interactions. It has also been used for controlling, as well as, diagnostic purposes to learn certain behaviours of people. More specifically, the following key conclusions can be drawn from the works reviewed above:

● People who suffer from certain disabilities, such as people who cannot use their hands, have limited options in terms of using controlling devices. In this case, providing any means of interaction is considered a significant achievement. Therefore, eye tracking as a means of HRI has been mostly addressed towards disabled people, which is also the case in HCI [40]. The dominant application in HRI has been wheelchair controlling applications. Despite the potential advantage of using eye tracking as an input device to control motorized wheelchair, only basic driving commands have been experimented and addressed. Even in the cases where eye tracking is the only means available for controlling, many difficulties have been faced and highlighted, such as the screen placed in front of the subject which blocks the vicinity of the subject [45].

● Due to the belief that better or simpler eye tracking algorithms and systems are needed, most of the works are eye tracking focused and not HRI focused. In these works, eye tracking algorithms or systems have been developed and HRI has been used as a testing bed. Therefore, the same eye tracking algorithm, or system, cannot be seen across different works. Instead, a custom eye tracking system has been developed in each work. In some cases, although aimed for HRI, no robotic experiments have been reported at all [72].

● The design of an interface that enables natural and intuitive HRI has not been studied in these works. Even when a GUI interface is presented, it is either not functioning or not designed thoughtfully. Alternatively, the interface is too complicated for the intended purpose and does not create any natural HRI. This is due to the fact that originally the works are aimed at developing an eye tracking system, or creating any means of interactions, as mentioned above. Therefore, very limited information and attention can be seen as far as the interface is concerned.

● No extensive evaluations or usability testing experiments have been reported. Even when a developed system has been evaluated, it has not been evaluated against other means of interactions. Also, very unrealistic tasks have been used to test the developed system. No HCI, HRI, or interactive systems heuristics and design principles have been considered in designing and evaluating the developed systems. In most cases, no evaluation results have been reported at all.

● Finally, eye tracking has not been studied, or at least considered, for mobile robot teleoperation, except in [20], where the intention of using eye tracking information to aid in the teleoperation of mobile robots has been mentioned. However, after personal communications with the author, it was concluded that the work has not been developed any further.

From the conclusions above, a significant gap in the literature has been identified. This work has been shaped by the need to fill this gap. Therefore, this work is different from the related works in the following points:

● Using eye-gaze tracking as a means for mobile robot teleoperation. The use is not only limited to issuing basic driving commands, but also controlling the pan/tilt unit (PTU) of the on-board camera. All necessary teleoperation commands are issued by the means of eye-gaze tracking.

● Focusing on the design of an intuitive novel interface and not on developing eye tracking systems or robotic platforms. This work depends on well established HRI heuristics and design principles in designing the interface, as "*the interface is not something that can be plugged in at the last minute*" (p3, [32]).

● Conducting extensive evaluations and usability testing experiments for the designed interface. To better quantify and standardize the evaluation results, this work evaluates the interface against conventional means of mobile robot teleoperation in task-oriented evaluations.

● Addressing main stream people and not only disabled people. Although the interface has great potentials for disabled people, this work addresses main stream people. This adds to the challenge of competing with conventional means of interactions, since potential users have more options in using different input devices.

Finally, to the best knowledge of the researcher, these elements have not been combined in any previous works. Some researchers have shown interest in this work since the start of it, who are in personal contacts with the researcher. For example, Zaheer Ahmad, who is a master student at Blekinge Institute of Technology, Sweden, is working on trying TeleGaze for different interaction scenarios, based on the recommendations of the researcher. Also the work in [23], has started after TeleGaze and has cited earlier publications on TeleGaze. These works however, are still in very early stages of design and evaluations. Therefore, at the present time, it is not clear which direction specifically they will follow.

Chapter THREE

Native TeleGaze

3.1 Introduction

In human-computer interaction (HCI) and human-robot interaction (HRI), quick prototyping techniques are used to start the design of novel interfaces [39]. It is mainly adopted for proof-of-concept and initial evaluations of novel systems. An integrated system that complies to a clear conceptual design, and a running interface that meets certain objectives can be used for this purpose. Some design work is necessary to acquire thorough understandings of the requirements of any system [19]. With clear research aim and clear research objectives, this stage of the research therefore, aimed at proofing the concept of TeleGaze. It started by integrating the necessary hardware components to build the TeleGaze system. It moved then, to experimenting few interface prototypes, and ended with some clear directions for redesigns and further evaluations. In this stage, all interactions with the interface are done through eye tracking only. Hence, this stage is referred to as *native TeleGaze*[6].

This chapter therefore, starts with the early design of the TeleGaze system including the TeleGaze conceptual design and the hardware components. The design

---

[6] See footnote 4.

principles of the TeleGaze interface are covered, which then are followed by a detailed explanation of the initial prototypes that comply with these principles. In order to determine user preference in terms of overall layout and design, an observational study is then described. Some initial usability testing that was carried out at this stage of the study is also covered together with the usability testing experiment design. An initial set of evaluation metrics, that produced some evaluation results, were used in the usability testing experiment of TeleGaze. Prior to the end of the chapter, these results are analysed and discussed. The conclusions at the end of the chapter clarifies some necessary research directions for the next stages of the work. These research directions are concluded based on some observed limitations of the system components at this stage including the interface.

3.2 TeleGaze System Components

The novelty of TeleGaze meant that no ready systems could be accessed and used to test the concept and conduct usability testing for the interface. Therefore, the research required intensive system development and component integration prior to trying any interface prototypes. This required a clear conceptual design that works as the base for the TeleGaze system throughout the different stages of the design and usability testing. The TeleGaze conceptual design, the hardware components and the system data manipulations are covered in the following sections.

3.2.1 TeleGaze Conceptual Design

A clear conceptual design, for both the user and the designer of any interactive system is vital for the success of the system in delivering the required interaction results [19]. The level of clarity of the conceptual design plays a significant role in building the appropriate mental model for the user of the system. The ideal mental model that the user can have is the one that matches the design model, which is the same as the designer's mental model. Both the user's and the designer's mental models should match the model used in the conceptual design.

The best mental model, as far as the user is concerned, is the one that is simple and easy to learn. This model can also be applied in storing and retrieving the required relationships between the system's components. On the other hand, from the designer's point of view the best mental model is the one that is simple to implement and simple to interpret [3]. Therefore, the conceptual design of TeleGaze was developed taking both the user and the designer into considerations such that it meets the criteria mentioned above within possibilities. Also the conceptual design meets one of the objectives of TeleGaze that is a platform-independent system. The conceptual design of the TeleGaze system is shown in Figure 3.1.

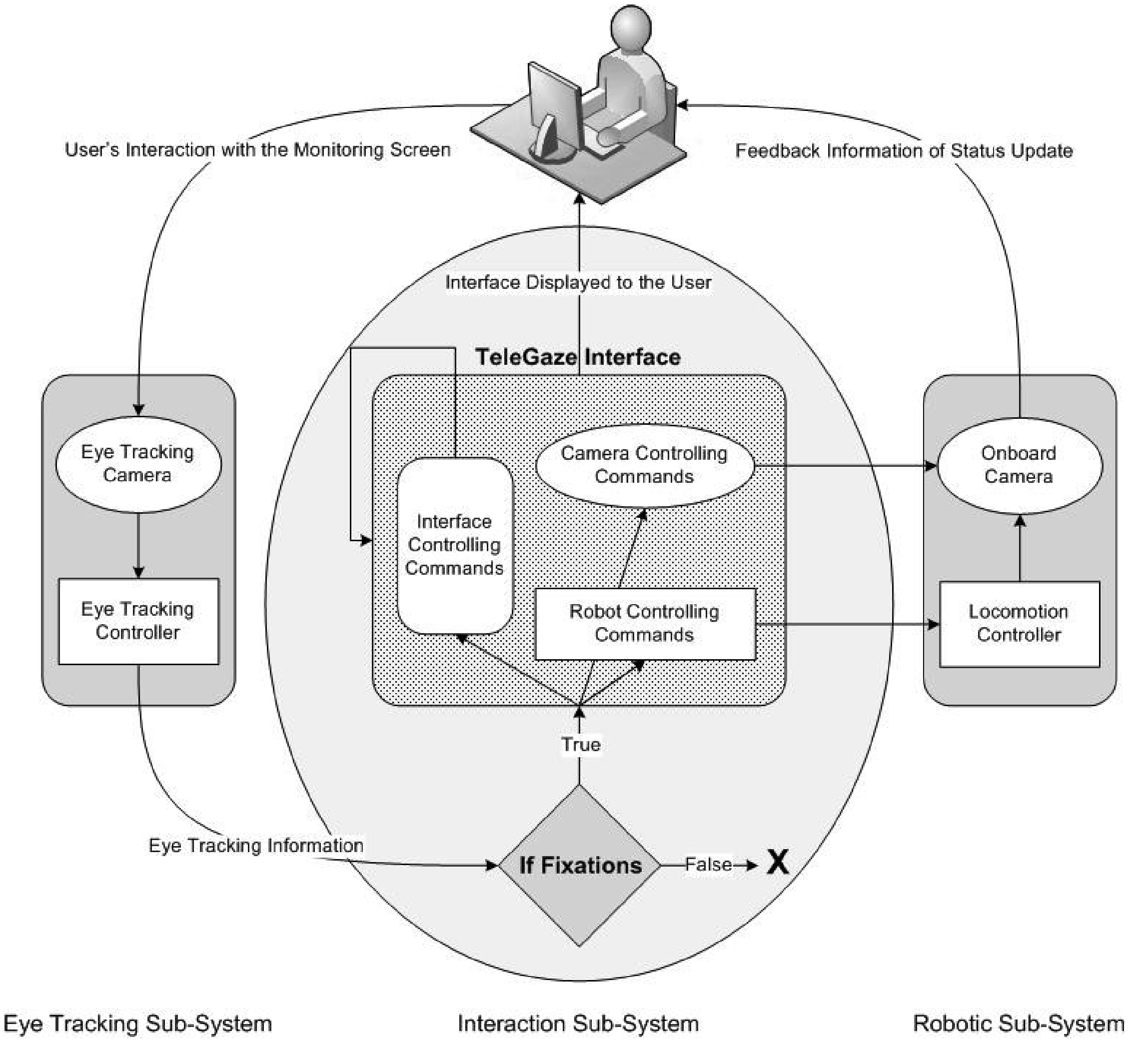


Figure 3.1: TeleGaze conceptual design.

The TeleGaze system consists of three integrated sub-systems; namely robotic, eye tracking and interaction sub-systems. The robotic and the eye tracking sub-systems are two components linked with the interaction software mainly running the TeleGaze interface. The human operator interacts with the interaction sub-system and not with the robotic sub-system directly[7]. Both the robotic and the eye tracking sub-systems interact with the interaction sub-system and do not interact with each other directly.

Following is the execution cycle that TeleGaze implements in order to enable teleoperation through eye gaze:

- The eye tracking sub-system provides the interaction sub-system with the eye gaze information of the human operator. This information is provided in accordance to the monitoring screen that displays the robot status and that the human operator interacts with.

- The interaction sub-system interprets this information into necessary commands and feeds them to the robotic sub-system. Only information on eye fixations are considered in this interpretation since fixations are considered the most common modality for gaze contingent interfaces [25].

- The robotic sub-system reacts to this information by making necessary changes in its status. This is implemented through executing driving actions and camera actions at a predefined speed.

- This change then is transformed back to the human operator through the feedback system. The feedback information consists of streams of images that are displayed to the human operator in real-time through the interaction medium.

- The human operator reacts to this information and behaves based on her will to produce the next action. This behaviour can be extracted through her gazing behaviour which is tracked by the eye tracking sub-system.

- The eye tracking sub-system once again reads the gazing behaviour and feeds the information to the interaction sub-system.

- ... and the cycle continues.

---

[7] This is important to bear in mind when classifying TeleGaze either as Human-Robot Interaction (HRI), Human-Computer Interaction (HCI), or both (Chapter 4).

From the execution cycle above, it can be seen that the interaction sub-system works as a meeting point for the eye tracking sub-system, the robotic sub-system, and the human operator as well. This shows the importance of the interaction sub-system in the TeleGaze system. The contact point between the interaction sub-system and the human operator is the TeleGaze interface which shows the importance of the TeleGaze interface in the TeleGaze system. This has dictated the direction of the research to focus on the design and evaluation of an interface and not other components of the system. The focus of this work sets it apart from other works in the field and is part of the main contribution to the knowledge (Chapter 2).

3.2.2 TeleGaze Physical Design

One of the main objectives of this work is to develop a platform-independent interface. A platform-independent interface can be integrated into any robotic and eye tracking sub-systems as long as they comply with the conceptual design. To comply with the conceptual design, the experimental platform used in this work consists of an eye tracking equipment, a mobile robot platform and a teleoperation station. The experimental platform used to design and evaluate the TeleGaze interface at this stage of the work is presented in Figure 3.2.

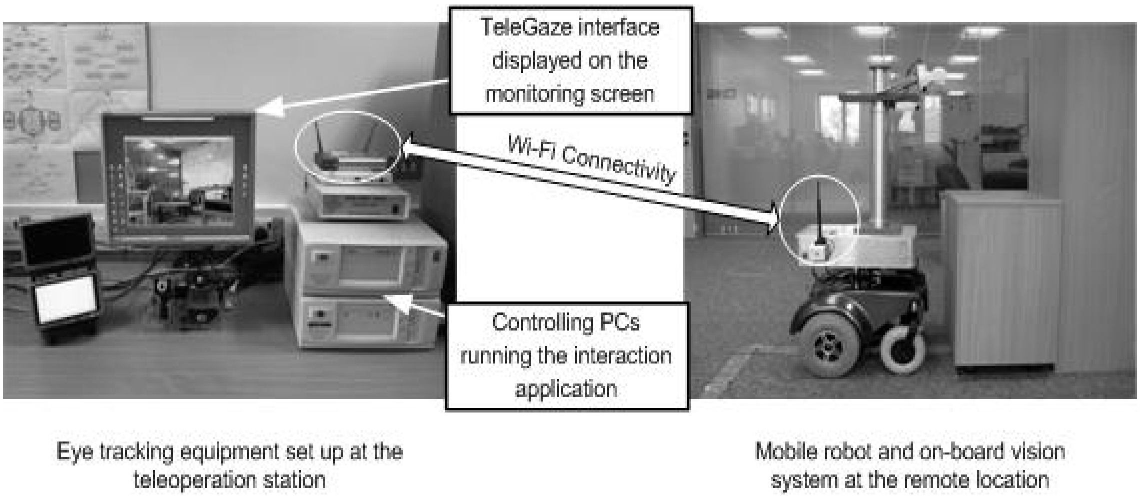


Figure 3.2: TeleGaze hardware components.

It is beyond the scope of this work to go deep into the details of the hardware components used in the experimental platform to design and evaluate the interface. However, "*performance of any interaction technique is the product of both its software and hardware*" [73]. Therefore, for the sake of information integrity with the empirical results presented in the coming sections, brief descriptions of the components are covered in this section:

Eye tracking equipment: A commercial eye tracking equipment from the Applied Science Laboratories (ASL)[8] has been selected for the purpose of this experiment. In addition to a set of interface software, the eye tracking equipment consists of three main units which are: the tracking camera, the tracking controller and the monitoring screens. A special camera, that is equipped with near infrared light emitters directed to the subject's eye, works as the tracking camera. The camera reads the reflections of these infrared lights from the subject's eye and feeds this information into the tracking controller.

The tracking controller extracts digital information of the subject's eye, such as the x- and y- coordinates of the line of the gaze on the interaction screen and the pupil diameter. These information can be obtained as superimposed crosshairs on the image that is viewed by the subject. It can also be fed to the connected PC in order to be used in customized applications. The monitoring screens are used for calibration purposes and monitoring the status of the system. The eye tracking equipment and its components are shown in Figure 3.3.

Eye Tracking Camera with Infrared Transmitters

Eye Tracking Controllers that Feed the PCs with the Readings

Eye tracking Monitoring Screens

Figure 3.3: The eye tracking equipment and its components.

This particular type of eye tracking equipment has been selected because it uses techniques based on reflected light from the operator’s eyes which is believed to be appropriate for this kind of interaction applications [9]. Although these systems are non-invasive and reasonably accurate, there are some known drawbacks. Among these

---

[8] The website address of the Applied Science Laboratory (ASL) is http://asleyetracking.com/Site/

drawbacks are the requirements to keep the head still and also the difficulty to keep a good contrast image.

To overcome these drawbacks up to a certain point, continuous parameters adjustments are required during the use of the system. Therefore, the system is designed so that a second person is necessary to supervise the system while the first person's eye is being tracked. The supervisor needs to keep monitoring the status of the system through the monitoring screens and adjusting a number of thresholds in real time. Furthermore, a nine point calibration for each tracking session is required prior to commencing any experiments. This type of calibration is required in order to obtain more accuracy[9] from the system [72].

Mobile robot platform: The mobile robot that was integrated into the system at this stage of the work is a modified wheelchair base equipped with on-board vision systems, Wi-Fi connectivity and necessary controllers. The vision system is composed of two network based pan/tilt/zoom (PTZ) cameras from VIVOTEK[10]. Only one main camera is required for TeleGaze. However, a second camera is added to increase the quality of feedback information provided to the human operator.

Controlling a robot from a remote location requires awareness of any obstacles in the close surroundings of the robot. To gain this awareness, one of the cameras is mounted looking downwards. This helps the operator to get an idea of the distance between the robot and any obstacles in the close surroundings of the robot. The physical setup of the cameras is meant to provide the user with sufficient situational awareness about the remote location. The robot and the camera setup are shown in Figure 3.4.

The mobile robot platform is equipped with two differential controllers mounted to the rear wheels. The two front wheels are caster wheels that help steering the robot. Differential steering is used to steer the robot depending on the angular and linear velocity values that the on-board controller receives from the teleoperation station. Wi-Fi connectivity is used to establish connections between the mobile robot platform and the teleoperation station where the interaction takes place.

---

[9] The average spatial error of the used equipment is under 1 degree. This error increases in cases where the subject's head moves from the initial position.

[10] The website address for the type of cameras composed the vision system of the mobile robot is: http://www.vivotek.com/products/model.php?network_camera=pz61x2

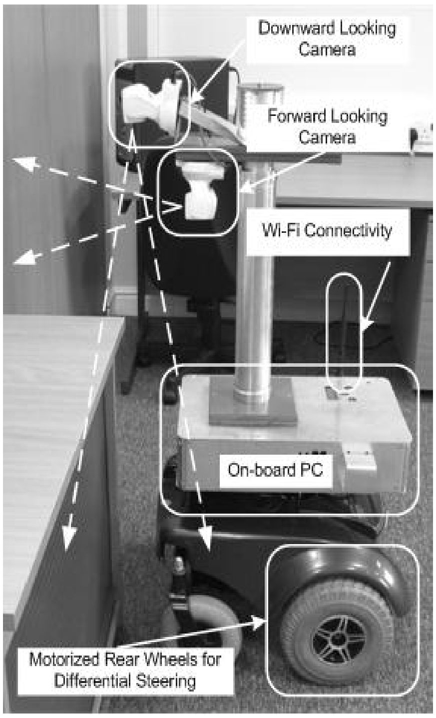


Figure 3.4: The mobile robot platform with the onboard vision system.

The teleoperation station: The teleoperation station is where the human operator located and the actual interaction between her and the TeleGaze system takes place. The teleoperation station is the interaction sub-system that displays the TeleGaze interface and performs most data interpretations. This is a conventional desktop PC with a 19” flat screen that is located above the eye tracking camera. The screen at this stage was set to a resolution of 1024x768 pixels. The interface is the only part of the screen that the subject interacts with during teleoperation. Controlling software and other application components are accessed from a second screen that is displayed to the supervisor and not the user. The screen setup and the teleoperation station is shown in Figure 3.5.

A *position-position* command strategy is used to generate controlling commands. Unlike *position-speed* command strategy, position-position is more accurate and allows the operator to move the robot to the desired location [5]. Therefore, sequences of discrete commands are generated in the remote station and set over ethernet to the robot. Any break in the connectivity between the remote station and the robot results in stopping the robot, since the robot stops receiving any more commands to execute. This is considered a safety precaution as well. The robot does not execute any commands if it is not connected to the remote station and controlled as well as monitored in real-time.

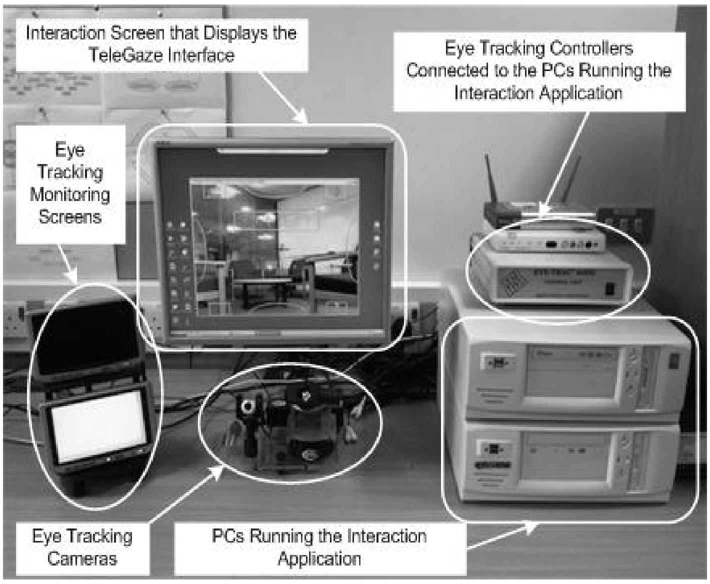


Figure 3.5: The screen setup in the teleoperation station.

3.2.3 Fixation Models

Raw eye tracking data might seem informative. However, to gain better insight into the subject's intentions further data classifications are required. As part of these classifications, mostly fixations are calculated from the raw eye tracking data. Fixations can be defined as fixing the gaze on a particular point for acquiring the information behind that point [73]. They are considered the most common modality for gaze contingent interfaces. Whereas very few works exist that use other modalities of eye tracking data [25]. Two main approaches exist for calculating fixations out of raw eye tracking data: *position-variance* method and *velocity-variance* method or what might be referred to as *summation* and *differentiation*, respectively (p138, [26]).

In the former method, averaging the signal over a time window, which is mostly known as dwell-time [44], is used. The variance in the signal is then compared against a predefined threshold. Fixations are registered if the variance calculated for the time window is less than the threshold. In the latter method, the velocity of the signal is calculated between two consequent points of the gaze. If the velocity is less than a predefined threshold, then the point is considered to be part of a potential fixation. This calculation continues for all consequent points that stay below the threshold and the average value is considered the fixation value at any time. Different works suggest different values for these thresholds [25]. However, most resources suggest that the values of these thresholds need to be found empirically [26].

Due to the need for temporal frequency matching, the position-variance method is used in TeleGaze. Based on this method, for any x_t and y_t as coordinates of the point-of-gaze (POG) at time, fixations are calculated using the following expressions:

$$\bar{x}=\frac{1}{n}\sum_{t=1}^{n} x_t \quad \text{and} \quad \bar{y}=\frac{1}{n}\sum_{t=1}^{n} y_t \qquad 3.1$$

$$\sigma_x=\sqrt{\frac{1}{n}\sum_{t=1}^{n} (\bar{x}-x_t)^2} \quad \text{and} \quad \sigma_y=\sqrt{\frac{1}{n}\sum_{t=1}^{n} (\bar{y}-y_t)^2} \qquad 3.2$$

$$f_n(\bar{x},\bar{y}) \ \textit{is Fixation if} \ \sigma_x<x_c \ \textit{and} \ \sigma_y<y_c \ \textit{for} \ t=n \qquad 3.3$$

Where n is the span of the time window (dwell-time)[11]. f_n is the calculated average of POG over time for $t=n$. $\bar{x}$ and $\bar{y}$ are averages of both x and y, respectively for $t[1,n]$. σ_x and σ_y are standard deviations of x- and y-coordinates of POG for $t[1,n]$, respectively[12]. x_c and y_c are predefined thresholds for x- and y-coordinates variance, respectively.

Recommended values in the literature can be used for the thresholds (x_c and y_c) and the time window (n) [26]. Alternatively, they can be determined empirically based on the context and the application. Due to the fact that a complicated data flow takes place in the TeleGaze system, different sets of data gets exchanged between the different components of the system.

The different sets of data gets produced at different frequencies depending on the frequencies of the individual components. Therefore, the value of the window time (n) is determined using a hybrid approach. It is calculated not only depending on recommended values, but also on some necessary calculations of frequency matching. The data flow of the TeleGaze system is illustrated in Figure 3.6.

---

[11] Time is a continuous variable. However, n is determined by the number of frames. That is why it is treated as a discrete variable and not continuous.

[12] Since the standard deviations are calculated for the sample and not for any populations, n is used and not $n-1$ in the equations.

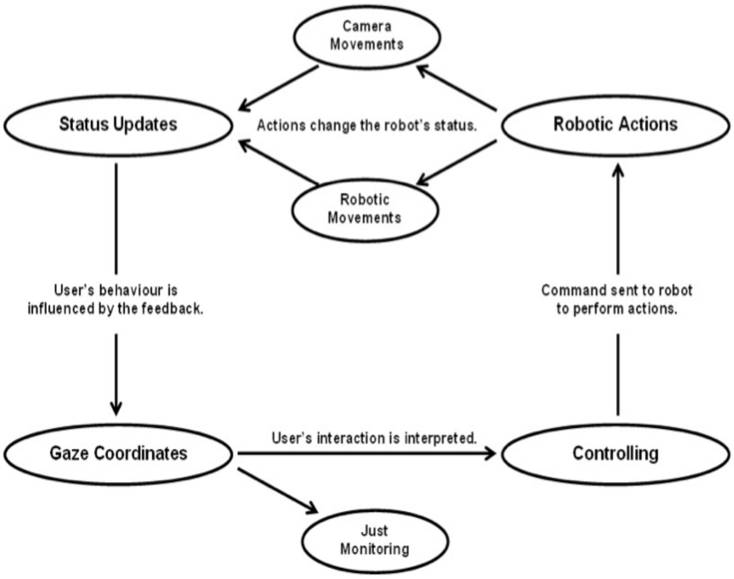


Figure 3.6: Data flow in the TeleGaze system.

3.2.4 Frequency Matching

The sub-system that manipulates most of the data of the other components is the interaction sub-system running on the remote PC. This sub-system is controlled by the interaction application that runs the interface and performs all necessary data interpretations for TeleGaze to work. The application behind the interaction sub-system, sitting in the middle of the TeleGaze system, receives data from both ends of the system and feeds back both ends of the system with data.

The robotic sub-system deals with two different sets of data. It receives the action commands from the interaction sub-system that has to be executed by both the robot and the camera. In the meantime, it sends real-time video images back to the interaction sub-system. Each of these two different sets of data is produced at different frequencies. The eye tracking sub-system, at the other end of the whole TeleGaze system, produces data at a completely different frequency. For the system integration to succeed as aimed for, some data tradeoff is necessary in order to match these different frequencies. This also aids in decreasing the bandwidth demand for the Wi-Fi connectivity to run. For these reasons, a range of data filtering is used in the interaction application in addition to running the interface.

The on-board controller of the robotic platform is designed such that it gets a sequence of discrete commands in forms of *forward*, *backward*, *left*, and *right*. The robot executes actions in a continuous manner only if a continuous sequence of commands is received. The idea behind this design is to stop the robot from executing any actions if the Wi-Fi connectivity with the interaction sub-system is lost. In the current form, any single command moves the robot only a certain distance and stops unless another command is received. The frequency of commands that the robot can handle is 3Hz (3commands/second). This determines the highest frequency that the robotic platform can run at. Receiving commands at higher frequencies than this, results in queueing commands and unpredicted behaviours when executed [21].

The maximum frequency that the on-board cameras can run at is 25Hz (25fps). The frequency of the on-board cameras in TeleGaze however, is set to 15Hz (15 fps). This is again to minimize bandwidth demand to a reasonable limit since the same available bandwidth is shared between all the components of the system. In addition to decreased frequencies, other approaches are used to minimize the bandwidth demand. For example, transmitting the images at a low resolution and decompressing them in the interaction application. Therefore, two out of the three known variables[13] for bandwidth (bits per second), are modified in TeleGaze to achieve real-time interactions. In addition to limiting the frequencies of the on-board cameras to only 15Hz (15fps), lower frequencies used for one of the cameras at different times. This is determined by the capability of TeleGaze in switching between the views of the two cameras. More details on this frequency differences is covered in the coming section (Section 3.5.1).

The eye tracking equipment on the other hand, works at a frequency of 50Hz (50fps). This frequency is the number of readings of the eye coordinates that the equipment produces in a second (50readings/second). Since the projection of the gaze needs to be superimposed on the video images displayed to the subject, this frequency needs to match the frequency of the video images. Therefore, the frequency of the readings has been averaged and reduced to only 15Hz (15readings/second). This is the same as the frequency of the on-board cameras. The actual frequency of the eye tracker

---

[13] The three known variables for bandwidth (bits per second) are: frame rate (frames per second), resolution (pixel per second), and grey scale (bits per pixel) (p157, [3]).

(50Hz) however, is used in extracting the fixations from the raw eye data and calculating the equations 3.1, 3.2, and 3.3 above.

The value of the window time (*n*) in the equations, can be determined using the hybrid approach mentioned above depending on both empirical and recommended values. Recommended values in the literature range from 50ms to 600ms [26] and [74]. Considering these recommendations and considering the necessary frequency matching in TeleGaze, a value of 330ms was selected for the window time (*n*) at this stage of the research. This value was selected because it lies in the recommended range and it helps in the necessary frequency matching. With this value, a maximum of only 3 consequent fixations could be registered in a second (3fixations/second), which matches the frequency of the commands that the robot can handle (3commands/second). It also matches one third of the frequency of the video images that are displayed to the user in the interface (15frames/second).

The selected value at this stage was not selected as a final value that would be used for all the future experiments of the research. This is an experimental value that was formed to fit the purpose of the experiments at this stage of the research with potential amendments based on empirical findings. Altering this value inline with experiments is highly recommended and is the most common approach used by researchers in the field [26].

3.2.5 Software Development and Programming

Integrating all the different components of the TeleGaze system to communicate and exchange data required a substantial amount of software development and programming. In addition to this, a significant amount of software development has gone to producing the different prototypes of the interface at different stages of the study. All data communications have to be done in real-time since TeleGaze is an interactive system that is used for real-time human-robot interaction (HRI). Therefore, different advanced programming techniques have been used to produce working prototypes that fit the purpose of TeleGaze. Also the necessary frequency matching and fixation calculations have been achieved through advanced software development.

In terms of programming languages, mainly C/C++ programming language has been used for all the different components of the software application. Data communication between the different components has been achieved using TCP/IP sockets where ethernet is used to connect the components. On the other hand, mail slots are used where other forms of connections have been used, such as serial or USB. The software diagram presented in Figure 3.7 shows more details on the different components of the software application and the connectivity technologies used.

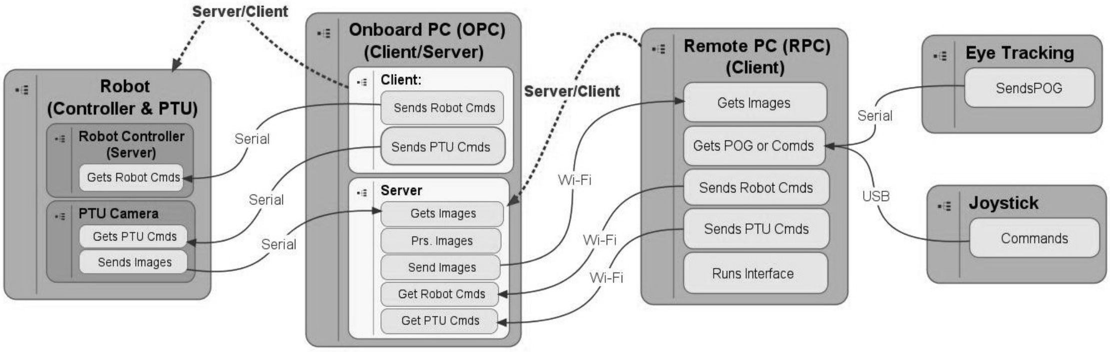


Figure 3.7: Different components of the software application.

As far as producing the layout of the prototypes is concerned, the OpenCV library has been used. OpenCV is an open source, C/C++ based computer vision library that is used for real-time vision based applications[14]. It includes basic drawing functions that have been used to superimpose the action regions on the images received from the camera in real-time. The OpenCV library is used also to transfer the images over ethernet in a compressed form in order to minimize bandwidth demands. Other libraries such as software development kits (SDK) of the cameras, the robot and multithreading libraries have been used when necessary to produce the software application.

3.3 TeleGaze Interface Design Principles

A well-known design principle that is implemented in most gaze driven interfaces is based on "*what you look at is what you get*" [8]. According to this principle, specific actions are triggered based on the current direction of the user's line of gaze. The same design principle has been implemented in the design of the TeleGaze interface. However, in addition to this specific design principle for gaze driven interfaces, a

---

[14] The latest version of OpenCV can be downloaded from http://sourceforge.net/projects/opencvlibrary/

number of rather generic design principles exist for any interactive interface. For instance, Benyon and colleagues [19] present an extensive set of design principles for interactive systems that can be tailored for TeleGaze.

Some design principles have evolved from human-computer interaction (HCI) systems into human-robot interaction (HRI) systems. Drury [75] for example, slightly adapted ten heuristics for HRI that are originally presented by Nielsen [76] for HCI. These heuristics can be further adapted to any interface in concern as ten "*rules of thumb*" [4]. Similar principles are addressed in [32], [77]. In this work, these rules of thumb and principles have been crosschecked, and applied to TeleGaze. Table 3.1 presents the original heuristics from Nielsen [75], the adapted heuristics from Drury [75], the principles from Benyon [19] and the final adaptation for TeleGaze.

The design principles in the table, are closely taken into considerations in developing any prototypes of the TeleGaze interface. The more the interface prototype comply to these principles, the better it complies to the common heuristics of HCI, HRI, and interactive systems. Therefore, each developed prototype of the interface is checked against these principles [77]. This checklist can also be considered as a form of evaluation of any of the developed prototypes prior to experimenting the prototype.

3.4 The Design of the TeleGaze Interface

Interfaces can be considered as tools to perceive the environment, make informed decisions and generate necessary commands to perform certain tasks [6]. Therefore, three main challenges can be found in designing any gaze directed interfaces, which are: the layout design, the size of individual components, and the visual feedback [25]. The layout design is challenging because the interface behaviour is associated with the, rather complicated, gazing behaviour of the human subject. The size of the individual components is challenging due to the nature of the data obtained from human eyes and due to the limited accuracy of current eye tracking equipments. The visual feedback is challenging because the information contained in the feedback affects the natural behaviour of the subject's eyes [35]. In addition to these general challenges that exist in most gaze contingent interfaces, specific challenges exist in TeleGaze. For instance, the

TeleGaze interface being a vehicle teleoperation interface requires sufficient command generation and feedback representation [6]. Poor feedback representation directly affects the quality of command generation. This is due to the fact that in teleoperation applications, the human operator relies totally on feedback for command generation.

Table 3.1 : Heuristics from HCI, HRI, and Interactive Systems adapted for TeleGaze [15]

Nielsen's heuristics proposed for HCI [75]	Drury's heuristics adapted for HRI [75]	Benyon's principles for interactive systems [19]	Heuristics adapted for TeleGaze
Does the program speak the user's language?	In the robot's information presented in a way that makes sense to human controllers?	Is the system status known to the user in real-time?	Is the interface interactive?
Does the program minimize the user's memory load?	Can the human(s) control the robot(s) without having to remember information presented in various parts of the interface?	Does the user feel being in control, and knows what to do and how to do it?	Is the interface responsive?
Is the program consistent?	In the interface consistent? Is the resulting robot behaviour consistent with what humans have been led to believe based on the interface?	Does the interface use the same design language?	Is the interface consistent?
Does the program provide feedback?	Does the interface provide feedback?	Does the interface provide feedback?	Is the interface informative?
Does the program have aesthetic integrity (e.g., a simple design)?	Does the interface have a clear and simple design?	Does the interface have a design style of its own?	Is the interface intuitive?
Does the program help prevent, and recover from, errors?	Does the interface help prevent, and recover from, errors made by the human or the robot?	Is there any chances of recovery, in case something goes wrong?	Is the interface elegant?
Does the program follow real-world conventions?	Does the interface follow real-world conventions, e.g., for how error messages are presented in other applications?	Does the design of the interface consider the user's familiarity?	Is the interface familiar?
Is the program forgiving; does it allow for reversible actions?	Is the interface forgiving, does it allow for reversible actions on the part of the human or the robot?	Does the interface provide any flexibility in performing the tasks?	Is the interface flexible?
Does the program make repertoire of available actions salient?	Does the interface make it obvious what actions are available at any given point?	Does the interface have good visibility in terms of available actions?	Is the interface clear?
Does the program provide shortcuts and accelerators?	Does the interface provide shortcuts and accelerators?	Does the interface have clear navigation amongst its commands?	Is the interface user friendly?

---

[15] The first and second columns are quoted exactly from the cited sources. The third column has been paraphrased from the cited source to fit the context.

In order to comply with the design principles and in order to tackle the challenges, a number of crucial points have been taken into considerations in the design of the interface. These points have been taken into considerations from functional, practical, and technical points of view. Figure 3.8 Shows the layout of one of the very first prototypes. Figure 3.9 shows an actual snapshot of the same prototype in action with the background view included in the scene.

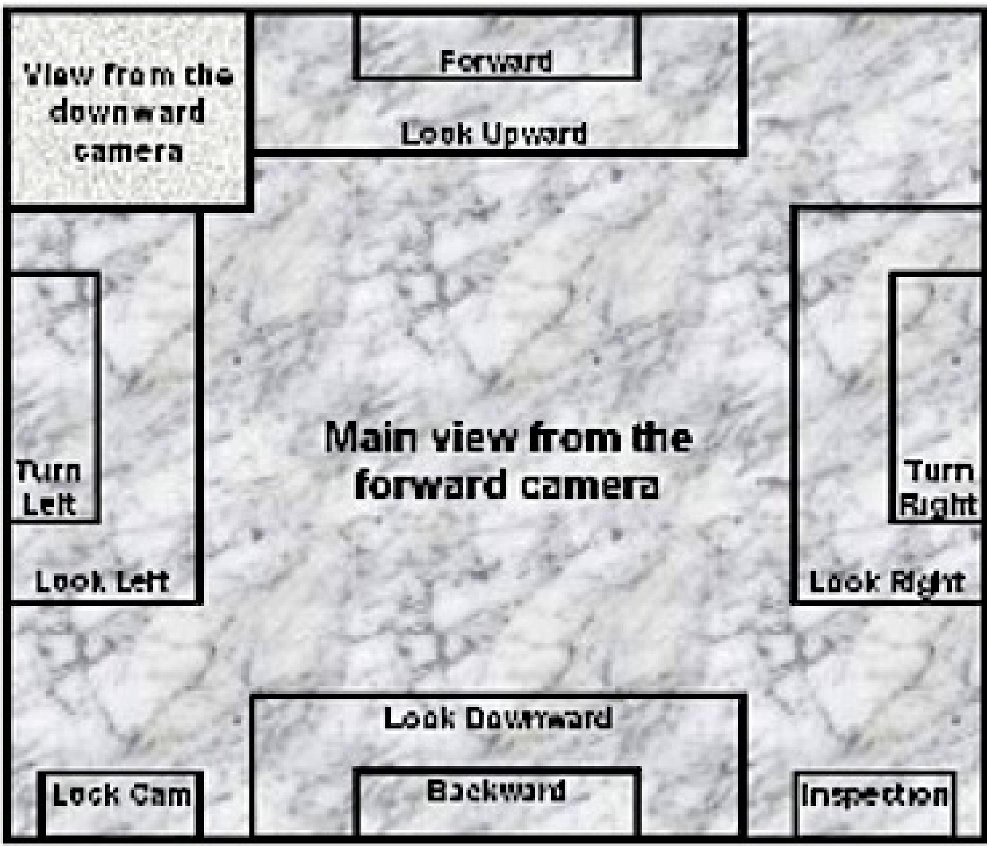


Figure 3.8: Layout of one of the first prototypes of the TeleGaze interface.

Following, is the details of the design considerations and the ideas behind the followed approaches:

3.4.1 Action Regions versus No-Action Regions:

One of the main objectives of the TeleGaze interface is to design a two way communication channel. This is to enable monitoring as well as controlling through the same interface using the same available space. Therefore, the concept of focus areas [19] has been used to augment the visual feedback on the interface by creating, what have been called in this work, *action regions*. The action regions are transparent regions on top of the video images that are displayed to the subject, rather than button-like shapes placed by sides of the interface. The rest of the display area therefore, has been called *no-action regions*. Having the action regions superimposed explicitly on the video images is necessary because "*understanding what a human intends is often fraught with considerable ambiguity*" [66]. Irrespective of action or no-action regions, all points on the interface can be potentially fixation points [78]. This is summarized in the following expression:

$$I_x = A_x \cup N_x \quad \textit{and} \quad I_y = A_y \cup N_y \tag{3.4}$$

Where I_x and I_y are the sets of the pixel values along the *x*- and *y*-coordinates of the whole interface, respectively. A_x and A_y are the set of the pixel values along the *x*- and *y*-coordinates that are considered as action regions, respectively. N_x and N_y are the set of the pixel values along the *x*- and *y*-coordinates that are considered as no-action regions, respectively.

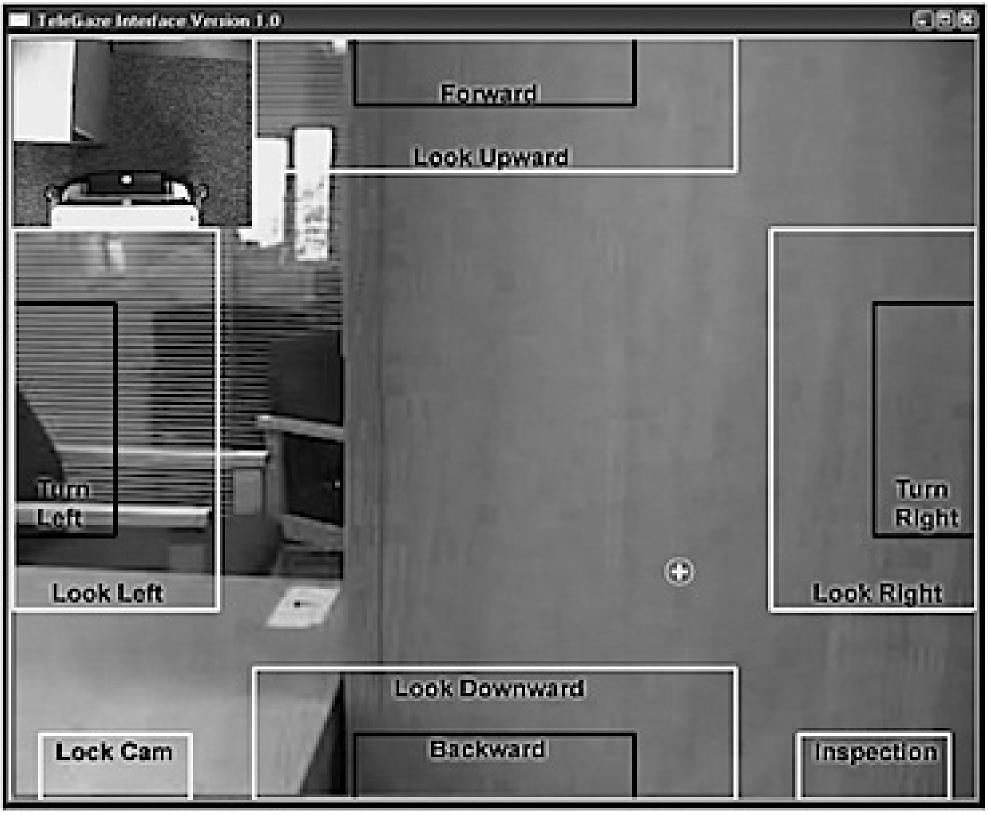


Figure 3.9: A snapshot of one of the first prototypes of the TeleGaze interface.

As mentioned above, one of the main principles in the design of any gaze contingent interface is "*what you look at is what you get*" [8]. Therefore, each action region is associated with an action that is believed to be necessary for TeleGaze. All the actions are necessary to achieve the required level of performance in any mobile robot teleoperation application. However, the action regions can be classified into three different categories: action regions that control the robot locomotions, action regions that control the on-board cameras, and action regions that control the interface itself. Despite the difference in the categories of the action regions, the same design language is used for all the action regions to comply with one of the design principles (Table 3.1). This is believed to increases the level of learnability of the interface regardless of the categories of the actions [19].

3.4.2 Size of Individual Action Regions

In general, very limited information is reported in the literature regarding the criteria for selecting a particular size for a particular interface. Even when the selected size is explicitly reported, the criteria for the selection is not available [8]. However, in designing gaze contingent interfaces, usually a relatively big size for the commands is chosen in comparison with conventional command buttons. This is because pointing in small areas with high resolution is not possible given the status of the current eye tracking technology [9], [79].

The size of the action regions can be chosen empirically based on the performance of the eye tracking equipment for each individual user [54]. This however, requires a complicated calibration process to find out the optimal size of action region for each particular user in a particular application context. Due to unavailable proven methods for this sort of calibrations, some rather generic approaches were followed for TeleGaze at this stage of the research. Using top-down and bottom-up approaches, an experimental size for the action regions was selected at this stage.

The top-down approach requires action regions as big as possible in order to ease the act of keeping the gaze in any of the action regions. This is an important consideration as focusing at smaller regions tend to be harder due to the jittery nature of human eyes. The bottom-up approach, on the other hand, requires action regions as small as possible. This is in order to increase the amount of no-action regions available for resting the subject's eyes. To arrive at a compromise solution for this stage of the research, an initial proportion of 1/4 of the interface area was selected as the total area for the action regions. This was believed to give the subject enough space to focus on the action regions, and enough space to rest the eyes. Evaluations of the interface at later stages of the research, determines the suitability of this initial value.

3.4.3 The Midas-Touch Problem

The nature of the inputs from the human eyes differs from that of any other input devices. Eyes are always engaged and therefore, a non-stopping stream of inputs needs to be dealt with during the course of the interaction. Picking what should generate actions out of a continuous stream of inputs is one of the biggest challenges in any gaze contingent interface. This problem is well known as the *Midas-Touch* problem [44]. Although a number of different solutions to this problem are recommended in the literature, one of the most practiced solutions is dwell-time [24] and [61]. Dwell time is keeping the gaze within a particular area continuously for a period of time to substitute the act of clicking in a conventional mouse.

As covered earlier in section 3.2.3 , eye tracking data can be classified into eye movements and fixations. Limited information can be acquired from some types of eye movements due to their high speed [13]. It is the fixations therefore, that most eye tracking applications depend on [25]. In TeleGaze, fixations that occur inside the action regions are highly considered as intentions for actions. As a confirmation for the intended action, and to experiment the recommended solution for the midas-touch problem, dwell-time is used at this stage of the research. The same value of the window time used to register fixations (the value of n in equations 3.1, 3.1, and 3.3) is used as the value of the dwell-time. As a result, a fixation that happens inside an action region issues the command that is associated to that particular region. Adopting a confirmation mechanism, such as the dwell-time, minimizes the likelihood of triggering any actions if not intended.

For safety purposes, fixations that occur outside the action regions do not trigger any actions. This fact is used as a mechanism to stop the robotic platform from executing any actions. Although a mechanical switch is recommended as an emergency stop, the followed approach has been experimented too [43].

Finding and triggering the intended action happens as modelled in the following expression:

$$\mathit{if}\ f_n(\bar{x},\bar{y})\ \mathit{is}\ \mathit{Fixation},\ \mathit{and}\ (\bar{x},\bar{y})\in A(x,y),\ \mathit{then}\ f_n(\bar{x},\bar{y})\,\mathit{is}\,\mathit{cmd}(A) \qquad 3.5$$

Where $f_n(\bar{x},\bar{y})$, x, y, $\bar{x}$ and $\bar{y}$ are the same as in 3.3. *A(x,y)* is the same as in 3.4. *cmd (A)* is the command associated with the action region A.

Based on the necessary frequency matching and on the value of *n* in expressions 3.1, 3.2 and 3.3 a maximum of three commands can be registered in a second. To keep the robotic platform execute the same command continuously for a period of time, the values of each *x* and *y* should satisfy the above expression for that period. Therefore, both acts of selecting and confirming any actions is achieved through gazing only.

3.4.4 Independent Camera Control

The act of associating robotic head movements with operator head movements has proven to show significant effects on the telepresence of the user [3], [80]. Research shows that even independently controlling a camera mounted on a mobile robot helps in teleoperation [17]. Therefore, the operator is provided with action regions to control the on-board camera separately from the robot base. This helps in minimizing the need to move the robot when the required angle of view can be achieved through moving the camera alone.

Although some of these actions can be achieved through the robotic actions as well, having extra options adds to the flexibility of the interface. Flexibility is one of the design principles that has been implemented in designing the TeleGaze interfaces. Furthermore, this extra level of controlling the camera independently is most useful when the operator is interested in different vertical angles of views to explore the scene vertically. Other practical uses also exist for this level of control. For instance, it helps in minimizing power consumption as the camera consumes less power than the robot base to achieve the same angle of view. Also the camera is more responsive, which increases the speed of achieving the desired angle of view in comparison with moving the robot base.

3.4.5 Extra View for Obstacle Detection

Research shows that having parts of the robot's body visible in the display helps in teleoperation [4]. This was difficult to achieve because the forward camera is mounted relatively high relative to the robot's body (Figure 3.4). In general, a limited field of view is achieved when teleoperating from the egocentric perspective of the on-board camera, which causes disorientation in teleoperation [81]. Therefore, a second camera looking downward was added to the vision system in order to achieve a top view of the robot's body. This camera provides a clear view of the close surroundings of the robot, which helps detecting obstacles and easing teleoperation. Although a PTZ camera is used for this purpose too, no control over the PTZ capabilities is provided within the TeleGaze interface. This is because, for the purpose of detecting obstacles, sufficient view can be achieved with a pre-fixed view angle which eliminates the need for further PTZ control.

Having two views from two different angles is an advantage [18]. However, having one being more predominantly displayed than the other is recommended. Based on some recommendations in [17], a screen-in-screen technique may be the most appropriate for this purpose. Therefore, the video display from the forward camera is displayed predominantly occupying 7/8 of the interface. While the video display from the downward camera is displayed at the upper-left corner of the interface occupying only 1/8 of the interface. This arrangement of displays was believed to provide feedback sufficient to both teleoperate the robot and inspect the scene, efficiently.

3.4.6 Interaction Mode vs. Inspection Mode

The eyes of the operator require rest from time to time [50]. In addition to the need for resting, the operator might need to inspect the scene more freely and closely from time to time. Therefore the no-action regions on the interface provide the operator with rest for the eyes. They also provide the opportunity to inspect parts of the scene that are not covered with action regions. However, to provide the user with a greater opportunity to inspect the scene and rest the eyes, a more radical solution is used in TeleGaze.

The operator is provided with the option of using the interface either to interact with the system or to inspect the scene only. What has been called the *inspection mode* of the TeleGaze allows the user to disable and remove all action regions from the interface. The only action region available in the inspection mode is one that is used to switch back to the interaction mode and reactivate the rest of the action regions. A practical use of this functionality is when reading a sign or inspecting a poster requires more space than what the no-action regions provide. Interacting with the interface in the inspection mode is, almost, completely free from the fear of issuing a non-intended command. Switching between the interaction mode and the inspection mode is also performed using inputs from the eyes. Figure 3.10 shows an actual snapshot of the inspection mode of the TeleGaze interface where only one action region can be seen.

Figure 3.10: A snapshot of the interface in the inspection mode.

3.5 Observational Study

During the process of designing a first prototype for the TeleGaze interface, it was realized that there is more than one alternative for achieving the same level of functionality. Making a final decision as far as the layout design is concerned, turned out to be difficult in terms of personal preferences. Choosing the right interface design depends highly on *user research* [82]. Therefore, unlike the previous related works (Chapter 2), this work extended the layout design alternatives to actual users of the interface. To find out the most preferred layout design for the TeleGaze interface, prior to any task oriented evaluation, an *observational study* was carried out. Observational studies, or what might be referred to as *formative designs* [19], are recommended and practiced in the design of many interactive systems [4], [26].

Two different High-Fidelity (Hi-Fi)[16] prototypes of the TeleGaze interface were initially designed for this observational study. The prototypes had differences not only in the layout design, but also in some functionalities. However, the major differences were in the layout of the action regions. Actual snapshots of both prototypes are shown in Figure 3.11. To easily distinguish between the two prototypes they are named the Edged-Interface (EI) and the Centred-Interface (CI).

(a) (b)

Figure 3.11: Snapshots of initial prototypes. a)- The Edged-Interface (EI), and b)- The Centred-Interface (CI).

3.5.1 The Differences in the Prototypes

Following are the key differences between the two prototypes that were used in the observational study:

---

[16] High-Fidelity (Hi-Fi) prototypes are similar in look and feel to the final product that can be used in usability studies. Therefore, they are time and effort consuming and users believe them [19].

a. Overall Layout

In the Edged-Interface (EI) the action regions are distributed more towards the edges of the interface. Whereas, in the Centred-Interface (CI) the action regions are concentrated in the centre of the interface. The idea behind the design of EI is to keep the centre of the interface free from action regions. This distribution of the action regions helps inspecting through the centre of the interface, which is likely to be more comfortable. Moreover, having the action regions that control the camera around the edge of the interface are more intuitive than having them in the centre. This is because the subject's line of gaze is more likely to move towards the sides when she intends to inspect more of the scene behind the sides.

The tradeoff in EI however, is the distance between the action regions, especially when the intention is to move the robot along a curvature. Research has shown that completion time increases as the distance between targets increase [38]. An additional tradeoff is the non-intuitive positioning of some of the action regions, which leads to non-intuitive gazing behaviours such as looking upward to drive the robot forward. On the other hand, the intuitive positioning of the action region “forward” in CI leads to more intuitive gazing behaviours such as looking forward to move forward. However, the centre of CI is packed with action regions which makes inspecting the scene harder.

b. Action Region Captions

Although the positioning of the action regions are meant to be as intuitive as possible, having captions was initially thought to be helpful. Therefore, another point of difference between the two prototypes is in the way the captions are displayed. In EI for example, the captions on the action regions are displayed continuously in a static manner. In contrast, the captions in CI are not displayed statically and they are moving along with the POG. They are displayed in the form of *tool tips*[17] rather than captions. In this design, the text is changing to the caption that is associated with the current action region, that is the one containing the POG. The caption style used in EI has been called *static captioning*, while the caption style used in CI has been called *dynamic captioning* due to its contextual changing behavior.

---

[17] Tool tips are dynamic information that is usually moved with the cursor. It can be considered as a common graphical user interface element.

The purpose of using dynamic captioning was to free the action regions from writings and increase the level of inspection that can be achieved through the action regions. However, “studies have shown that interrupting a user's task at random moments can cause decreased performance on the main task” [83]. Therefore, this is one of the differences that was necessary to be studied in the observational study.

c. Top-View Display

As it was mentioned earlier in section 3.4.5 , the display from the downward looking camera (Figure 3.4) is positioned at the top-left corner with a size of only 1/8 of the interface. The video images of this display are streamed at 1Hz to minimize bandwidth demands. Higher frame rates than 1Hz is not thought to be necessary as this display considered a secondary display to aid the main display from the forward camera. This view is thought to be necessary only when obstacles need to be detected and monitored in the close surroundings of the robot. Although, the mentioned displays arrangement (Section 3.4.5 and Figure 3.8) is though to be sufficient for teleoperation, further control over the arrangement was believed to be useful.

As a difference in the functionalities of the interfaces, EI does not provide any control over the display from the downward camera. However, CI provides the opportunity to enlarge and centre this display in order to have better view of the obstacles in the close surroundings of the robot. This was thought to be very useful when precise manoeuvrability is necessary, or the environment is cluttered with obstacles. Also, in addition to enlarging and centring the display, the frequency of streaming the images is increased to 15Hz. This is equivalent to the frequency that the main display in the interface runs at (Section 3.2.2).

To provide this level of control, an action region labelled “swap cam” is added to CI. Swap cam provides the user with the capability of switching the main display between the forward camera and the downward camera. Hence, the user has the capability to set the view from the forward camera as the dominant display similar to EI. Alternatively, she can set the view from the downward camera as the dominant display when required. This control over the display arrangement is one of the key differences between both prototypes, as it is not available in EI. Changes in the display arrangement does not affect the layout of the interface and the action regions. Figure 3.12 shows a snapshot of the CI with the display from the downward camera enlarged and centred.

d. Camera Realignment

Since the camera is mounted on the robot and not globally in the environment, the horizontal alignment of the camera is difficult to notice or remember. The user might change the pan angle of the camera at one moment, then start moving forward at a later moment without being aware that the camera is not aligned with the robot. This makes the robot's forward movement look like sideways or at a different angle. To avoid this situation, it is necessary for the camera to realign with the robot alignment whenever the robot starts to move. This might not be desired at all times however, especially when the operator remembers the fact that she has aligned the camera differently from the robot.

Figure 3.12: The CI with the top-view enlarged and centred.

Therefore, another point of difference in the functionalities of EI and CI is the automatic realignment of the camera once a moving command is issued to the robot. In

EI, the operator is presented with the capability of locking the camera at any angle and moving the robot, which means disabling the automatic realignment. In CI however, the camera is realigned automatically whenever a moving command is issued to the robot. The operator has more options in EI, but the extra options require extra care and the responsibility to use it appropriately. Research shows that the option of locking/ unlocking some degrees of freedom might be useful in some situations [81]. Therefore, the option of enabling, or disabling, the automatic realignment is provided in EI through an action region labelled "lock cam".

Four key differences can be identified between EI and CI. It was decided that both prototypes should be tested by potential users of the interface and their preferences should be taken into considerations. Therefore, the aim of the observational study was to find out the most preferred option for each of the four differences.

3.5.2 Participants[18]

A group of ten participants volunteered to participate in the TeleGaze observational study. The age of the participants ranged from first year students to some senior members of staff (22 to 43 years old), including 2 females and 8 males. Different levels of familiarity with using computers could be noticed among them. As far as the number is concerned, ten was selected as an initial number for this stage of the research. Although this was an experimental number, it is a recommended number for similar user studies [19], [84].

3.5.3 Design of the Observational Study

At this stage of the work and for the purpose of this observational study, the participants were not asked to perform any specific tasks with the interface. Instead, they were left free to explore the functionalities and study the layouts of both prototypes. Unlike [24], where only a couple of users were left to try the system, in this work all ten participants tried both prototypes of the interface. To obtain comparable results however, the participants were left to try the prototypes for an equal period of time (4 minutes[19] precisely for each prototype). This approach, although different from

---

[18] Since the participants are not the subject of the test and the eye tracking equipment is non-intrusive, this study has been considered as ethically approved.

[19] This was equal to the estimated time that the intended navigational tasks in future experiments were believed to take. Eye tracking experiments are normally short to avoid the effects of fatigue and accumulated drifting errors, except when other

traditional scientific experiments, it reveals interesting findings in some occasions and can lead to building necessary hypotheses [86]. Therefore, it is useful when observations are necessary for building initial ideas, such as at this stage of designing the TeleGaze interface.

Before people set about performing tasks, they need to establish an understanding of the system abilities and functionalities [19]. Therefore, at the beginning of any session, the participants were given a brief verbal description of the idea of the study including descriptions of how the interface works. They were asked later to explore and try both prototypes of the interface with clear explanation of the aim of the experiment [77]. To avoid any bias towards any particular prototype, half of the participants were presented with EI first and then CI. This was turned around for the other half of the participants. This presentation in a systematic rotation was necessary to counter-balance for any likely practice and boredom effects. However, hard copies of both interfaces were given to the participants prior to commencing with the actual use of the prototypes. This is to help the verbal description and get them prepared to meet the prototypes of the interface.

After trying both prototypes for an equal amount of time, the participants answered a predesigned questionnaire for both prototypes. The questionnaire was designed to examine the participants interaction experience and their observations on both prototypes. The questionnaire was specially designed to address the four key differences between both prototypes (Section 3.5.1) in addition to some other related questions. A full version of the questionnaire is included in Appendix A (Section A.1).

3.5.4 Results and Data Analysis

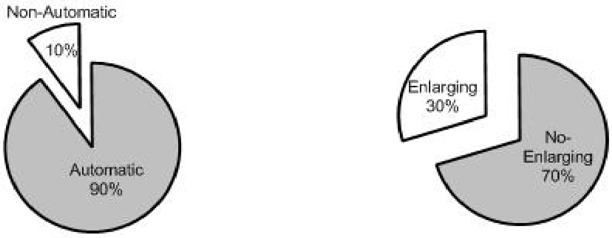
With two key differences in the layout and two key differences in the functionalities of the prototypes, the participants preferences can easily be figured out from the answers to the questionnaire. The questionnaire, in its simplest function, gave the participants the chance to vote for their preferred option after trying and exploring both prototypes. The results of the voting are presented in Figure 3.13.

---

modes are used to assist such as in [85].

Overall Layout of the Interface
- a -

Captioning the Action Regions
- b -



Camera Realignment with the Robot
- c -

Top-View Display
- d -

Figure 3.13: The results of the voting for ten participants on the four key differences in both prototypes. a- Overall layout (EI vs. CI), b- Captioning (static vs. dynamic), c- Camera realignment (automatic vs. non-automatic), d- Top-view display (enlarging vs. no-enlarging)

From the results shown in Figure 3.13, it can be seen that a higher number of votes goes for EI as far as the general layout of the action regions is concerned. Most of the participants preferred the centre of the interface to be as free as possible to provide clearer view of the scene. However, they were happy with the positioning of the "forward" action region in CI, as it looked more intuitive to them than it looked in EI.

Regarding displaying the captions on the action regions, a higher number of participants preferred static captions than dynamic ones. The mentioned reason was that the dynamic caption causes distractions, as it was predicted prior to the study (Section 3.5.1). Therefore, they preferred the captions to be displayed statically and continuously. However, a number of the participants mentioned that with more practice, they might discard all the captions due to the intuitive positioning of the action regions which makes them easy to recall.

Similarly a higher number of votes went to EI regarding the display from the downward camera. Most participants did not believe that further control over this display is necessary since it is used as a secondary view. More interestingly, two of the participants believed that a second camera is not necessary. They believed that the main camera can be aligned to provide the same view as the second camera if necessary.

On the other hand, most of the votes goes to CI when it comes to the automatic realignment of the main camera. Despite the additional control that EI provided to lock the camera at a different angle than the robot's alignment, most participants preferred automatic realignment. However, few participants mentioned that they would like this option if an alignment indicator exists on the interface.

In addition to the questions addressing these four key differences between the prototypes, some other points were addressed in the questionnaire. For example, the issue of displaying the action regions explicitly versus implicitly. The action regions are displayed explicitly because human intentions are expressed in a vague and unclear way [78]. Most participants (7 out of 10) preferred explicit action regions rather than implicit. The rest of the participants (3 out of 10) believed that the positioning of the action regions are intuitive enough to be remembered without being explicitly displayed on the interface.

The issue of displaying the POG as a superimposed crosshair on the interface was addressed in another question. Most participants (7 out of 10) preferred the POG to be displayed on the interface because it works as a confirmation message of where exactly the POG is. This finding contradicts the believe that superimposing the pointer causes visual distraction [87]. The rest of the participants however, preferred the current action region that contains the POG to be highlighted. One of the participants mentioned that he would like a focus point in the centre of the action region to help him gaze at the centre of the region. This however, contradicts with one of the objectives of TeleGaze that is enabling controlling while monitoring and not blocking one functionality for the sake of the other.

In general, from the results of the questionnaire it can be concluded that the participants preferred a rather simpler interface than a complicated one. Although some extra options might help at some times, it affects the usability at some other times by adding to the complexity of the system. The conclusions and observations from this study are implemented in designing a refined version of the interface as a ready prototype to perform navigational tasks.

3.6 Refined Interface Design

In the light of the results covered above, a refined interface (RI) as a third prototype has been developed for TeleGaze. RI meets all the user preferences observed in the study, in terms of layout and functionality. In RI the centre of the interface is mostly free, there is no control over the downward display, the cameras are automatically realigned and the captions are displayed statically. With these features RI meets the preferences of most of the users participated in the study. Figure 3.14 shows a snapshot of the refined interface (RI).

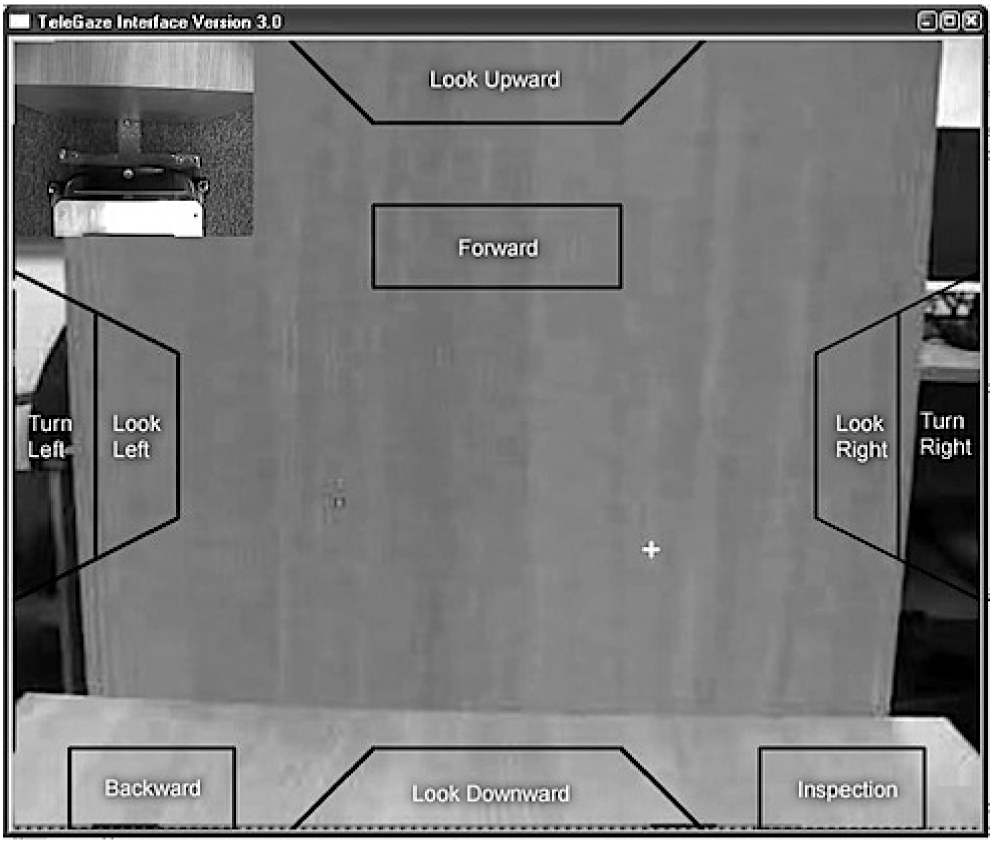


Figure 3.14: A snapshot of the refined interface (RI).

An interesting observation during this study is that almost none of the participants used the “backward” action region. The participants were free to explore the interface and try any of the actions during the allocated time. All the different action regions were tried more than once by a number of the participants. The “backward” action region however, was used only twice by two different participants. Therefore, the size of the “backward” action region was reduced in RI and moved out of the centre of the interface.

It can be argued that RI is the best interface design for TeleGaze. However, interfaces are tools designed and developed to perform certain tasks. The intended tasks determine the fitness of the design. Therefore, to maintain a reasonable balance between design and evaluation, it was decided to move forward at this stage to undertake a task oriented evaluation for the refined interface of TeleGaze.

3.7 TeleGaze Usability Testing

The aim of TeleGaze is to enable a human operator to teleoperate a mobile robot from a remote location. Therefore, the main element of any usability testing for TeleGaze should include, some sort of, teleoperation tasks. Since TeleGaze is not the only means for mobile robot teleoperation, the experience that it creates should be compared with other available means. Therefore, TeleGaze has been compared with a conventional means of teleoperation, such as joystick, for usability testing. Joysticks have been selected because most robotic platforms come with joysticks as the main interaction tool, particularly commercially available platforms [31].

Joysticks have been selected as a target for TeleGaze to meet. If TeleGaze can provide the interaction experience and the usability of a conventional joystick, then the advantage of having both hands of the human operator free puts TeleGaze forward. This is true even if TeleGaze does not beat the competitor joystick since the bonus of hands-free teleoperation still exists. Therefore, the aim of the usability testing is to measure the performance and the interaction experience of TeleGaze in completing a navigational task compared to a conventional joystick.

3.7.1 Experiment Design

A common approach to usability testing is to set a series of goals or tasks and to measure the time and effort necessary for a subject to accomplish that task [88]. Therefore, in order to be able to compare the task completion experience of TeleGaze with the one of the joystick, a navigational task had to be designed. The mobile robot had to be teleoperated to perform the navigational task using both TeleGaze and the joystick in order to measure different aspects of both means in question.

Navigational tasks are highly application dependent and might differ in many elements. Examples of these elements are speed, accuracy, and complexity of the navigational route. However, any navigational task can be divided into a number of subtasks, and furthermore, a number of actions. Moving along a straight line, turning right and left, and finally stopping at a designated point are essential subtasks in any navigational task. Eye tracking experiments on the other hand, are normally short to avoid the effects of fatigue, eye squinting, and the accumulated errors of eye tracking equipments [85], [89]. To include all the essential subtasks while keep to a practical limit, the navigational task illustrated in Figure 3.15 was designed for the usability testing of this stage of the research.

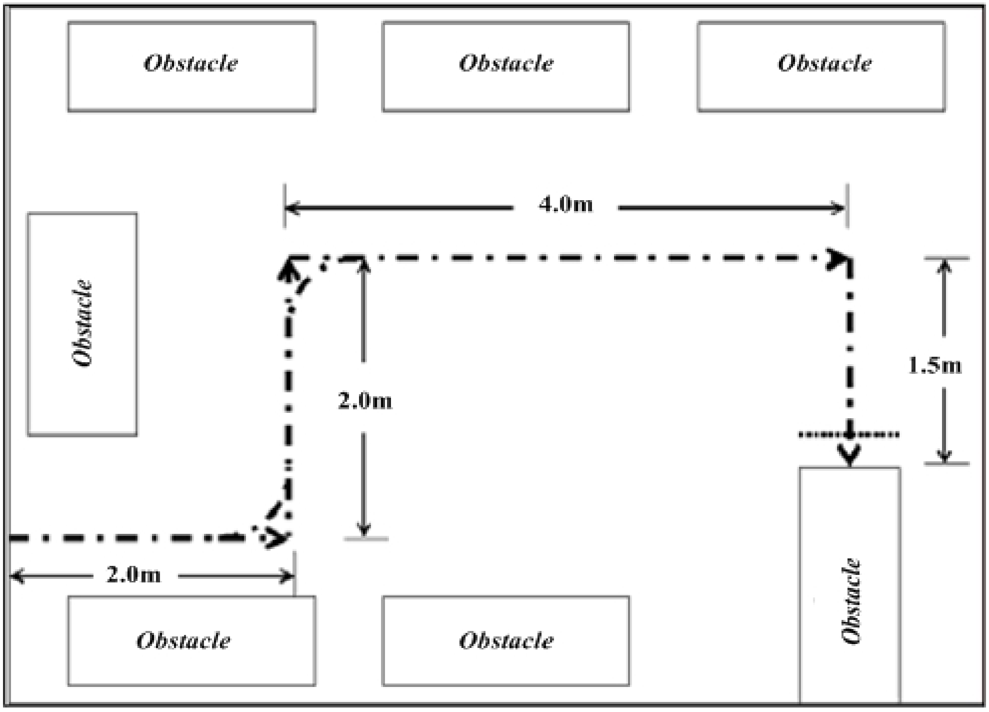


Figure 3.15: The navigational task used in the usability experiment of the native TeleGaze.

In addition to both TeleGaze and the joystick, a third mode of interaction was also included in the usability testing. This mode used the TeleGaze interface driven by a computer mouse instead of inputs from the eye tracker. Hence, it is called mouse-driven-interface (MDI). Perhaps the computer mouse is the most important change from a HCI perspective in the last 30 years [19]. Therefore, the MDI was included as a combination of a conventional input device and a novel interface design. Although a conventional joystick was used, not all the participants had prior knowledge with using joysticks. However, all the participants had prior experience with using a conventional computer mouse. This adds to the importance of this interaction mode in terms of user familiarity and interface novelty. Also by comparing eye tracking as an input device with a standard baseline device, such as the mouse, it is easier to determine how good the eye tracking system is [87], [22].

One major characteristic of teleoperation is that the control is conducted from a remote location. This makes the task far more difficult to perform, since the user is not interacting with the robot in real three dimensional spaces. Instead, the interaction is in two dimensional spaces. In this case, the user interacts only with feedback and different forms of data rather than the actual robot itself. Therefore, in all three modes of interactions, the participants were allowed to monitor the robot only through the interface to control the level of feedback information.

Participants of the experiment were asked to drive the robot along the track shown in Figure 3.15 using all the different modes of interactions including the TeleGaze interface. A brief explanation of each interaction mode was given to each participant with only one minute exercise prior to commencing the actual task. This was to get the participants familiar with each interaction mode and how the robot responds to individual commands. Then the time and accuracy of task completion were recorded for each participant and for each interaction mode.

3.7.2 Evaluation Metrics

To the present day, due to the diversity of HRI applications there are no standard metrics to evaluate any newly developed interaction systems. However, a number of common metrics can be adopted to evaluate a developed system for a particular application domain [90]. ISO 9241-11 suggests that usability should include efficiency, effectiveness, and user satisfactions [91]. For this stage of the work therefore, the following set of evaluation metrics was used to measure the usability of TeleGaze:

a. Objective Metrics

In many HRI applications, when performing certain tasks are involved, the two widely adopted metrics are efficiency and effectiveness. These two common metrics are used in the usability testing of TeleGaze too. The meanings of these two metrics however, are highly application, and hence task, dependent. Therefore, for the usability testing of TeleGaze, efficiency was defined as the time to complete the navigational task. Effectiveness on the other hand, was defined as the accuracy to keep the robot on track. Since the aim of the usability testing is to compare TeleGaze with other modes of interactions, the absolute values of these metrics are not meaningful as much as their relative values.

b. Subjective Metrics

Subjective metrics are closely adopted in any usability testing involving humans. Since the human operator is an important element of the TeleGaze system, subjective metrics too were adopted. One form of measuring subjective metrics is using questionnaires. Therefore, a specifically designed questionnaire for this stage of the research, was filled by the participants of the usability testing experiment. The used questionnaire this time is totally different from the one used for the observational study reported earlier in section 3.5.4. All the participants filled the questionnaire after completing the task with all three modes of interactions. The participants rated their agreement with a set of statements in favour of TeleGaze on a Likert rating scale. In addition to a set of statements, open questions were included in the questionnaire to collect personal opinions and comments. A full version of the used questionnaire is included in Appendix A (Section A.2), while individual questions are referred to in the coming relevant sections.

3.7.3 Data Analysis

As mentioned above in section 3.7.2, two different sets of evaluation metrics were used in this usability testing experiment to evaluate TeleGaze. A set of objective metrics, which evaluates the system's performance by evaluating the efficiency and effectiveness of the system. Also a set of subjective metrics, which evaluates the level of user satisfaction through a set of specifically designed questionnaire. Following are the results of the measurements obtained from the usability testing experiment at this stage of the research:

a. System Performance

Efficiency and effectiveness are two measures used to evaluate system performance. Efficiency, as mentioned above, is defined as the time-to-complete the navigational task. For the purpose of data analysis, the average of time-to-complete task (statistical mean) for all ten participants is calculated and plotted in Figure 3.16[20].

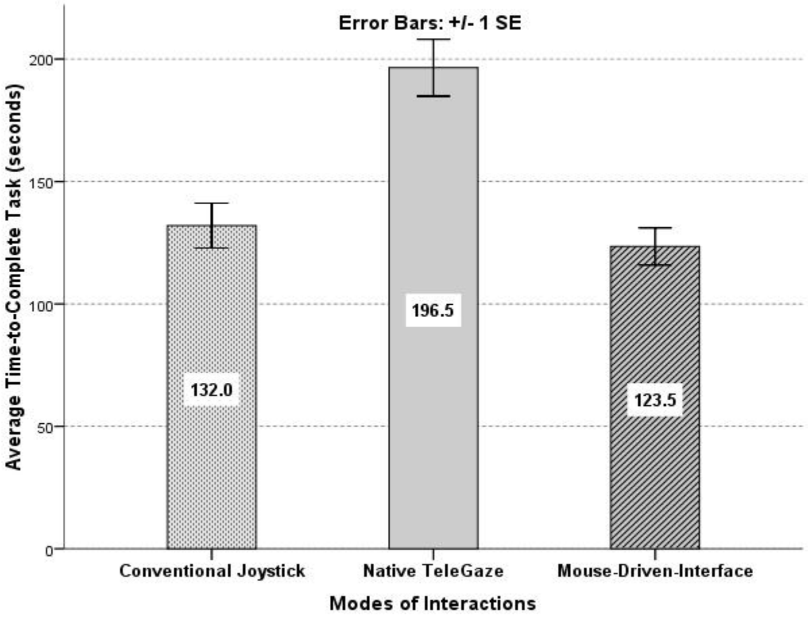


Figure 3.16: Average time-to-complete the task in all three modes of interactions.

Since the joystick is the target for TeleGaze to meet and efficiency is based on time-to-complete task, it can be inferred from Figure 3.16 that the native TeleGaze is not as efficient as its target. It can also be inferred that the native TeleGaze is also less efficient than the mouse-driven-interface (MDI). However, these conclusions are only based on differences in the absolute values of time-to-complete task. Further statistical analysis are required to determine the significance level of the differences in the calculated values of efficiency.

To analyze the efficiency of all three modes of interactions statistically, the following testing hypothesis is constructed[21]:

H_1 *: Time to complete task is different between the different interaction modes.*

---

[20] Corrections for error-bars are performed following the recommended procedure by Field in (p317, [86]).

[21] All constructed hypotheses in this section are expressed as “proposed outcomes” and not “proposed causes”. That is hypothesizing the values of the dependent variables and not the independent variables (p7, [86]).

The null hypothesis H_0 then, is that the average time-to-complete task is not different among the interaction modes. A one-way repeated measures ANOVA[22] is used to test the hypothesis. The results of the test show that efficiency as time-to-complete task is significantly affected by the interaction mode $F(2,8)=11.514$, $p<0.05$. This means that H_0 is rejected since it takes at least one of the interaction modes significantly more, or less, time to complete the task than one, or both, of the other two. Therefore, to find the source of this difference, more testing hypotheses are constructed as follows:

H_2 *: Time to complete task is different between joystick and native TeleGaze.*

H_3 *: Time to complete task is different between native TeleGaze and MDI.*

H_4 *: Time to complete task is different between joystick and MDI.*

This time, a two-tailed t-test is used to test the above hypotheses. The results of the test for H_2 show that on average it takes native TeleGaze ($M=196.50$, $SE=14.28$)[23] significantly more time to complete the task than the joystick ($M=132.00$, $SE=11.53$), $t(9)=3.308$, $p<0.05$[24], $r=0.74$. The results for H_3 show that it takes native TeleGaze significantly more time to complete the task than the MDI ($M=123.50$, $SE=8.59$), $t(9)=4.206$, $p<0.05$, $r=0.81$. However, the results for H_4 show that it does not take the joystick significantly more time to complete the task than it takes the MDI $t(9)=0.697$, $p>0.05$, $r=0.22$.

On average, the joystick is 32.82% faster than the native TeleGaze which is statistically significant. The MDI on the other hand, is 37.15% faster than the native TeleGaze which is also statistically significant. However, although the MDI is 6.43% faster than the joystick, it is not statistically significant. Therefore, based on mean ranking for the three modes of interactions, the MDI comes first, then comes the joystick and finally the native TeleGaze as far as speed is concerned. Considering statistical significance both the MDI and the joystick come before the native TeleGaze for speed of task completion.

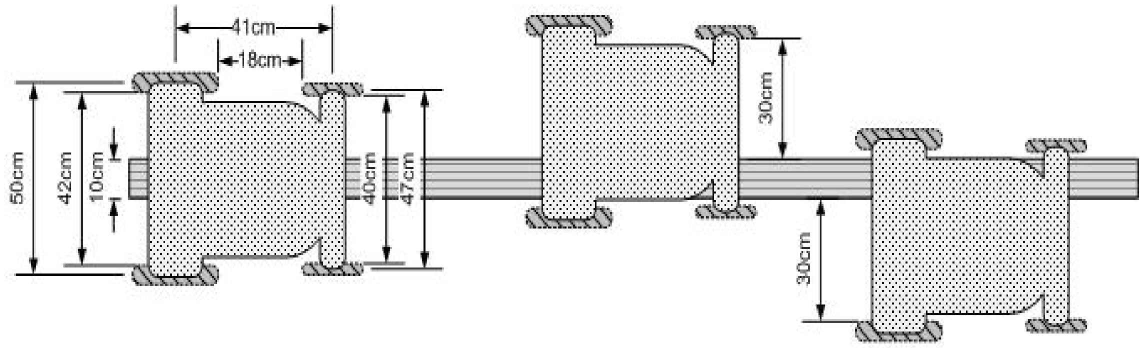
---

[22] Using Kolmogorov-Smirnov normality test, the results of testing for normality show that time-to-complete task for the joystick, the native TeleGaze, and the MDI is not significantly different from being normally distributed D(10)=0.139, 0.157, 0.136 , p>0.05, respectively (p145, [86]). Using Mauchly's test, the results of the test show that the assumption of sphericity is not violated $\chi^2(2)=2.202$, p>0.05 (p474, [86]).

[23] M= Mean or Median, SE= Standard Error.

[24] It is common to use 5% as the significance level in scientific research (p40, [92]). Therefore, 5% is used throughout this research to test for significancy.

To evaluate the effectiveness of TeleGaze, the accuracy was defined as keeping the robot on track. In this context, this is defined as keeping the track, which is marked by a tape on the floor, between the wheels of the robot. Considering the dimensions of the robot base, no more than 30cm diverge from the centre line is allowed at any point on the track. Any more divergence results in one of the wheels crossing the track, which is calculated against the accuracy. This means that despite moving out of a straight line, if none of the wheels cross the track, the accuracy is considered as 100%. The dimensions of the robot base and the maximum divergence allowed to score full accuracy is illustrated in Figure 3.17.



- a -
The robot on the centre of the track

- b -
Maximum divergence allowed to the left

- c -
Maximum divergence allowed to the right

Figure 3.17: The accuracy range for the robot in relation to the track.

Overall, the task was repeated thirty times. Ten times for each interaction mode. Based on the above defined accuracy criterion, eight out of those ten times in each interaction mode the task was completed with full accuracy. Regardless of the amount of divergence from the centre line, the accuracy for the remaining two attempts was scored as zero. Therefore, no difference was recorded in the accuracy for any of the interaction modes. However, full accuracy was not achieved with any of the interaction modes. The concluding results then is that TeleGaze meets its joystick target as far as effectiveness is concerned[25].

b. User Satisfaction

Using subjective metrics, user satisfaction is calculated from the results of the questionnaire. The participants rated their agreement with the statements in the questionnaire on a 5 point Likert rating scale, where 1 is *strongly disagree* and 5 is

---

[25] A total of three cases were recorded that required starting allover again due to failures not related to the interface or the system design. Examples are network failure, battery down and computer crash.

strongly agree. To visualize the results, the average of the ratings of the participants for each statement in the questionnaire is plotted in Figure 3.18.

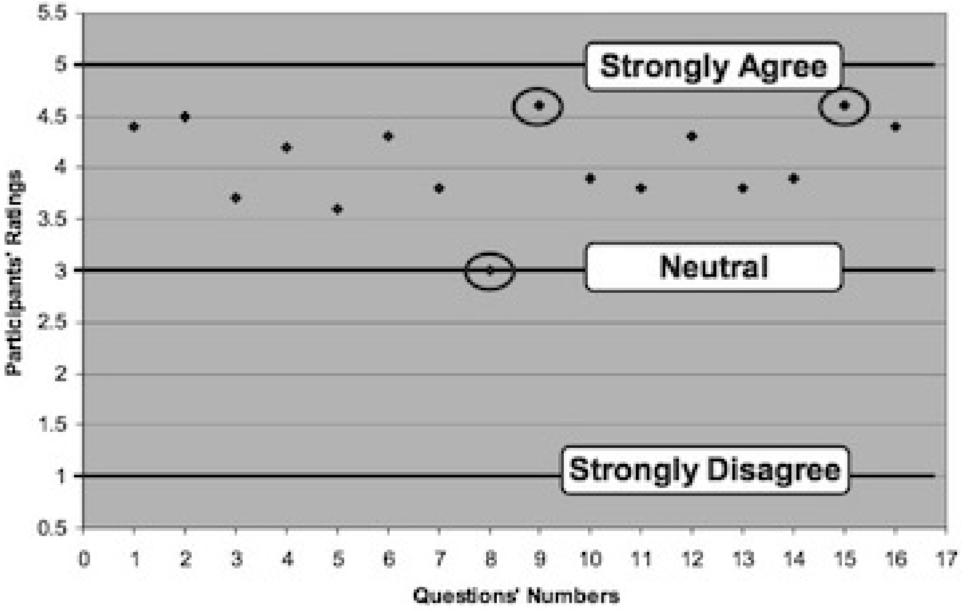


Figure 3.18: Average of participants' ratings for the statements in the questionnaire.

From the plotted results above, it can be seen that all the ratings (except for question 8) lie in the region between Neutral and Strongly Agree. Since the statements in the questionnaire are all constructed in favour of TeleGaze, all results above neutral are considered positive. Therefore, it can be inferred that the participants, on average, are satisfied with TeleGaze as a novel means for HRI and mobile robot teleoperation. However, to get better insight into the results of the questionnaire, the granulated ratings of the participants are plotted in Figure 3.19.

The minimum agreement obtained is for question 8, which stated that the user can perform more complex tasks with the current system. The majority of the participants on average did not agree with this statement, but did not disagree either. On the other hand, the maximum average ratings obtained is for both questions 9 and 15. Question 9 stated that the user can perform better with more training and practice and question 15 stated that the user would like the system to be developed further. These findings emphasize the need for further development despite the fact that the participants are, on average, satisfied with the system.

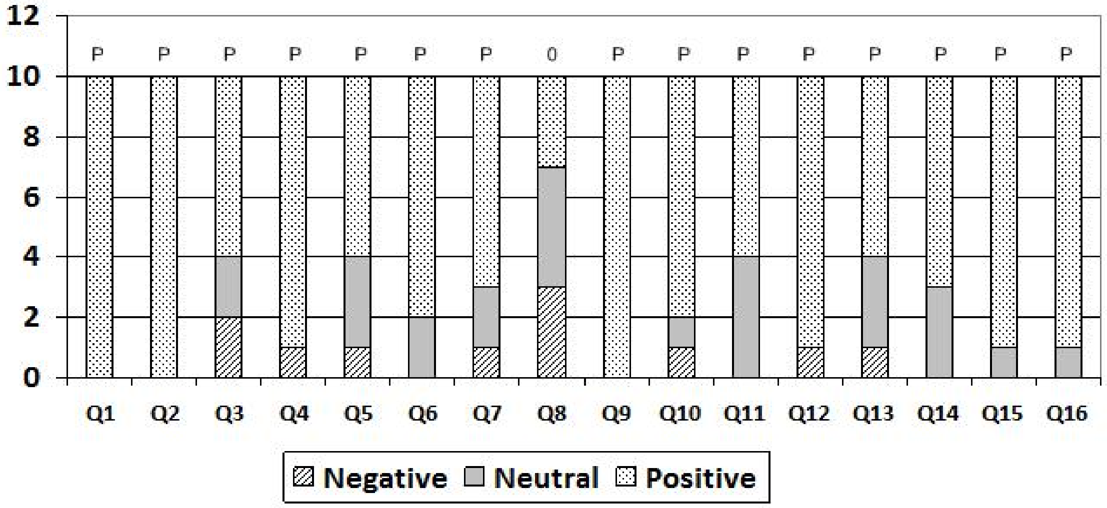


Figure 3.19: Granulated ratings of the statements in the questionnaire. P- Positive, 0- Neutral, N- Negative.

A second part of the questionnaire composed of only three questions, exploring the limitations of TeleGaze, if any, from the participant's point of view. The first statement suggested that the design of the TeleGaze interface is the most likely source of any limitations in the system. The second statement suggested that the way the system works, such as using dwell-time to confirm an action, is the most likely source. The third question stated that the eye tracking equipment is the most likely source. The same rating mechanism mentioned above used to obtain the participants agreements with each one of the three statements. The granulated ratings of the participants agreements are plotted in Figure 3.20.

It can be seen from the figure above that most participants (6 out of 10) do not agree that the design of the interface is the most likely source of any limitations. A similar number of participants (not the same participants), do not agree that the way the system works is the most likely source of any limitations. However, most participants (7 out of 10) agreed or strongly agreed that the eye tracking equipment is the most likely source of any limitations in TeleGaze. This statement suggested that the state of the art of the eye tracking technology and the used eye tracking equipment are the sources of the limitations and not the fact that inputs from eyes are used for controlling.

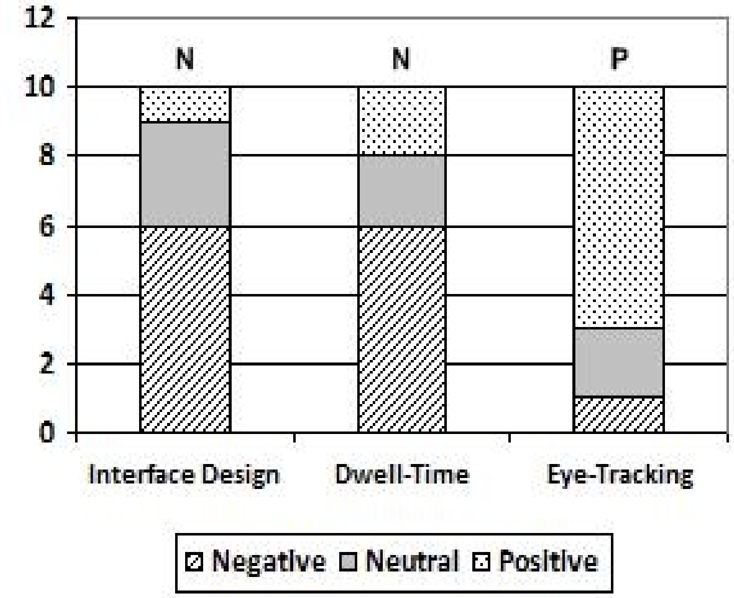


Figure 3.20: Granulated ratings for the three most likely sources of limitations (second part of the questionnaire)

3.7.4 Discussion

It can be concluded from the statistical analysis that the native TeleGaze does not meet its joystick target as far as efficiency is concerned. However, MDI, which adopts the TeleGaze interface driven by a computer mouse, outperforms the target. This is mainly due to the fact that all the participants have prior experience in using computer mice. This is true in most cases where novel interfaces compete with mouse driven interfaces [24]. However, the fact that the TeleGaze interface provides a list of necessary actions with the conventional click-of-mouse has played a significant role in forming the above results. The interface provides an intuitive interaction means while supported by the extensive experience people have in using computer mouse [22].

The significant difference in the efficiency of the native TeleGaze was observed to be due to another important reason, which was caused by the requirements of the task. The main requirement of the task was to follow the track marked on the floor. This required the operator to monitor the track from the view of the downward camera displayed on the upper-left corner (Figure 3.8).

In the case of using the joystick, the operator was able to look at the view while issuing commands with the joystick. In contrast, using the native TeleGaze, it is not possible to monitor that view and issue commands at the same time. Consequently, the overall time that took the operator to complete the task was not consumed in issuing commands. This phenomena was clearly observed during the experiment. However, no actual measurements were taken of the time that monitoring the track took from the overall time to complete the task, because it was not predicted prior to the experiment.

The above situation is less likely to happen in a different navigational task or interaction scenario. For example, when moving from one point to another does not require dominantly monitoring the close surroundings, or the alignment, of the robot. In this case, the operator would focus on the main view and would be able to issue commands that does not cost the task, almost, any extra time. This is one of the key issues that has been addressed in the coming stages of the research.

As far as accuracy is concerned, the results show that the same level of accuracy can be achieved regardless of the interaction mode. This can be due to one or two reasons. The first is that all the interaction modes provide the same level of situational awareness and command generation that is necessary to perform the task. The second reason is that the task is not challenging enough for these two features (situational awareness and command generation) of the interaction modes to show any differences.

User satisfaction on the other hand, is promising. The results of the questionnaire show that the participants are, on average, satisfied with TeleGaze as a means for mobile robot teleoperation. However, the questionnaire used at this stage of the research focused on the native TeleGaze only and did not address any elements of the other two interaction modes. Similar to the approach used in measuring the system's performance, comparative subjective metrics are likely to reveal more interesting results.

The results from the second part of the questionnaire show that the most likely source of limitations is the eye tracking equipment and technology. However, comments from individual participants suggest that the design of the interface and the way the system works have potentials to cause limitations for the system. Suggestions for larger action regions, more space between the looking and the turning action regions on the sides, and better eye tracking equipments were among those comments.

3.8 Conclusions

This stage of the research aimed at proofing the concept of TeleGaze as a novel means for mobile robot teleoperation. From the work conducted throughout this stage and the evaluation results obtained, the following key points can be concluded:

● TeleGaze has the potential to beat conventional means of mobile robot teleoperation with the significant advantage of total hands-free control. This is true if the level of compromise that TeleGaze requires pays for the advantages that it delivers. Therefore, the usability of TeleGaze needs to be investigated more in context rather than being generalized. To achieve this, a more thoughtful design of the usability experiment, which pushes TeleGaze to the limits, is required. This can be in the context of a particular application domain that is most likely to fit TeleGaze.

● To evaluate the usability of TeleGaze in any application context, a more extensive set of evaluation metrics is necessary. This set should evaluate TeleGaze from a multidisciplinary point of view. Also each metric should be defined specifically for the context in question. Specifically defined metrics are more likely to reveal more insights into the usability of the system and its limitations.

● To produce more meaningful results for fellow researchers in the field, a more standardized robotic platform is necessary to be integrated to the TeleGaze system. The mobile robot platform used in this stage of the research meets the hardware requirements for the robotic sub-system. However, some limitations of the TeleGaze system, such as the need for a secondary camera, are believed to be due to the platform. A more standardized robotic platform is likely to eliminate these limitations. It also aids in the definitions of the evaluation metrics and therefore, in the interpretations of the results.

● A redesigned usability experiment, an extensive set of evaluation metrics, and a more standardized robotic platform require a redesigned interface. Two of the objectives of the research are a platform-independent and an application-independent design for the TeleGaze interface. Changing the robotic platform and putting the application into context are good opportunities to test the design against these objectives.

- Finally, a multimodal approach is worth considering to better diagnose any limitations that might have been caused by the use of dwell-time in native TeleGaze. Depending purely on inputs from the human eyes, the native TeleGaze achieved performance close to the joystick target. The purpose of adopting a multimodal approach is increasing the performance of the system to meet its joystick target, while meeting the aim of TeleGaze of hands-free teleoperation.

Based on these key conclusions, the next stages of the research have been decided. In the next chapter, the details of an extensive set of evaluation metrics specifically designed for TeleGaze is covered. Also, details of the robotic platform upgrade, application domains and the redesigned usability testing experiment are covered. The multimodal approach and the redesigned interface are covered in the following chapter (Chapter 5).

Chapter FOUR

Evaluation Metrics and Experiment Design

4.1 Introduction

Designing interactive systems is an iterative process which cycles between the design of the system and the evaluation of it [19], [88]. Therefore, evaluation comprises an important part of the design process. Without this part, the design cannot be improved further. Evaluating the designed prototypes at each stage determines directions of further improvements in order to get closer to the target of the system [32].

The nature of the system determines the evaluation forms which are likely to produce the results that direct further improvements on the design. Also different forms of evaluations might be carried out for different stages of the design. Due to the limitless list of evaluation forms that have been used in evaluating interactive systems, selecting the most suitable set of evaluation metrics is not a trivial task [28]. In fact, the art of evaluation is quite difficult to manage for most interactive systems. Therefore, it is left for people specialized and experienced in evaluation. This common behaviour in designing interactive systems is mostly referred to as *expert evaluation* [36], [93].

In addition to the fact that evaluating any interactive system is not a trivial task, evaluating novel systems tend to be even more difficult. In this case, the design itself involves designing a new set of evaluation metrics. Considering the novelty of TeleGaze as an interactive system, it was realized that a new set of evaluation metrics is needed to be designed and used for evaluation. Therefore, the aim of this phase of the research is to develop a set of evaluation metrics which will be used throughout the rest of the research. More specifically, the evaluation metrics should meet the following criteria:

1. Evaluates TeleGaze against the design principles (Chapter 3).
2. Applies, either partially or fully, to more than one phase of the research. This enables comparison of the evaluation results of different phases of the research in order to determine the design trend and overall progress.
3. Selected from a range of highly recommended and experimented set of metrics by fellow researchers in the field.
4. Statistically analysable in order to better generalize and quantify the findings.
5. Evaluates the design from different points of view of the different disciplines related to the TeleGaze system.
6. Guides the design of the interface towards further improvements [1].
7. Does not require vast effort or expense to measure, record and analyse [91]

It is very essential for the designer to have a clear idea about the questions that the evaluation needs to answer. Therefore, the evaluation of TeleGaze should answer each of the following questions [94]:

1. Does TeleGaze produce the desired outcome or not?
2. Is TeleGaze then better than other modes of interactions or not?

3. If the answer to the previous question is yes, then why is it better and how much better it is?

4. Are there any components of the system that can be removed without affecting the outcomes?

In order to design a set of evaluation metrics that meet the objectives and answer the questions mentioned above, this chapter covers the details of designing a set of evaluation metrics. The set of metrics is going to be used in the evaluation of the coming phases of the research. This chapter covers the details of the evaluation metrics in relation to the limitations of TeleGaze highlighted in chapter 3. Prior to this, some other considerations such as hardware upgrade and experiment design, are covered.

4.2 Hardware Upgrade

One of the key features of the native TeleGaze interface experimented in the previous phase was the view from the downward camera (Figure 3.12). This feature provided necessary situational awareness to the operator as far as the close surroundings of the robot body is concerned. Due to the height of the robot neck, which is also the height of the main camera, manoeuvrability was difficult without this view. Research shows that both vision and proprioception are combined in a very efficient way to plan movement if, for example, the hand in a robotic arm is visible prior to movement [18]. This is also true when the camera display shows part of the robot's body while controlled from a remote location [4].

The visibility of the robot body, or more specifically the robot nose, increases the operator's situational awareness and hence the efficiency in planning necessary movements. However, it was found in the usability experiment that monitoring this view while performing the navigational task using TeleGaze adds to the overall time-to-complete-task. This led TeleGaze to score less efficiency when compared to the joystick. To solve this problem, similar level of situational awareness needs to be provided through the display of the main view only which saves the time needed to monitor the secondary view.

Another limitation of the hardware used in the previous phase of the study was the lack of granulated steering values, i.e. values other than absolute forward and absolute turn left/right. The previous robotic platform provided four discrete actions of *forward, backward, turn right,* and *turn left*. The heavy weight of the robot also added significantly to the response time. These limitations appeared to have affected the performance of the whole system. Although some of these limitations affected the joystick as well as the TeleGaze, the granulated steering problem was less significant using the joystick. To achieve granulated steering, the users were pressing both the forward and, for example, the turn right buttons at the same time to produce a combination of linear and angular velocities. However, this functionality was not possible with TeleGaze due to the fact that the POG can be in one action region at a time. It can be in either forward or, for example, turn right.

Another limitation of the previous platform was in the vision system. The non-smooth response of the pan/tilt behaviours of the camera appeared to have affected the overall performance. Due to physical characteristics of the camera not every command produced the same pan/tilt result. This was mistakenly understood by some users and created some level of confusion among them. The users explained the difference in the pan/tilt results as their non-consistent use of the interface, which was not true. This misunderstanding is believed to highly affect the interaction experience and therefore, is necessary to be eliminated.

In addition to the reasons mentioned above, it is believed that a more standardized robotic platform is necessary in order to produce more informative results. Therefore, the robotic platform has been upgraded at this stage to a Pioneer P3-DX research platform from Mobile Robots Inc. [31]. This platform is a Wi-Fi enabled mobile robot with differential steering. The platform is also equipped with a video camera mounted on a pan/tilt unit. In order to keep the same naming conventions throughout the research, the new robotic platform has been called GazeBot. Different views of GazeBot and its dimensions are shown in Figure 4.1.

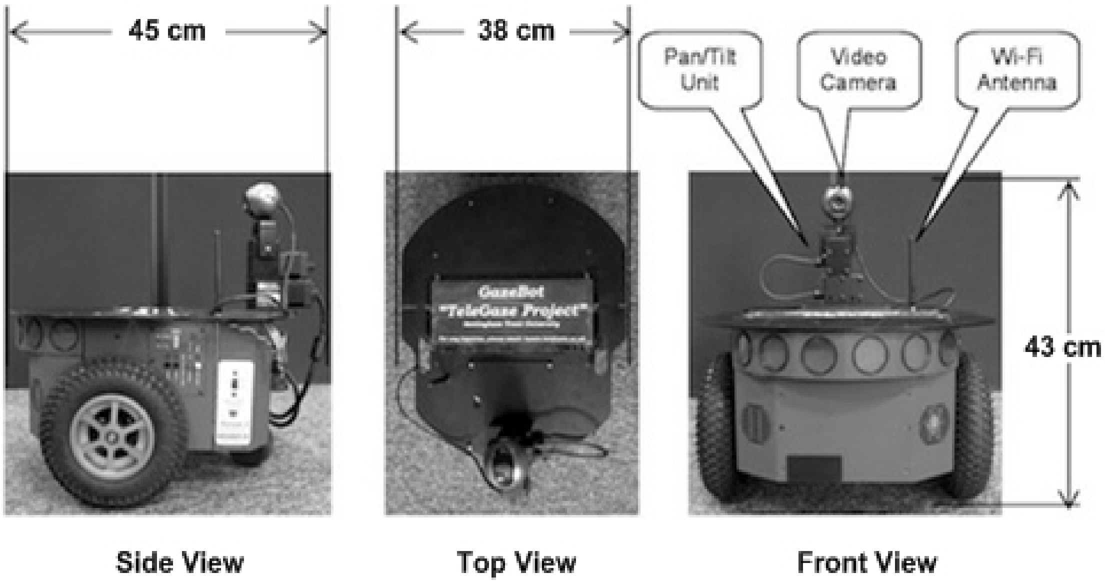


Figure 4.1: The new robotic platform: GazeBot.

Since the hardware architecture of GazeBot is similar to the previous platform, no change in the conceptual design of TeleGaze is necessary. This rather unplanned change in hardware tests TeleGaze for one of its main objectives, which is a platform-independent interface. The fact that TeleGaze can be easily integrated to GazeBot demonstrates that TeleGaze implementation is not tied to a particular robotic platform. Instead, it can be easily implemented on any mobile robotic platform as long as the platform uses the same hardware architecture set in the conceptual design (Chapter 3).

In addition to the fact that Pioneer P3-DX is a more familiar research platform in the robotics field, GazeBot has some main interesting features in comparison with the previous robotic platform. For example, tilting the camera downward very slightly makes the nose of the robot visible in the view of the camera. This feature aids in gaining more affective driving, as previously mentioned. More importantly, it eliminates the need for a secondary camera to provide visibility of the close surroundings of the robot body. Also with two differential wheels and a third scooter wheel, it is possible to rotate GazeBot around itself. This helps in getting through, or out of, more difficult and narrow pathways, the thing that was more difficult with the previous platform.

Furthermore, the feedback system that the PTU provides regarding the exact pan/tilt angle at any time enables better situational awareness as far as the camera alignment is concerned. All these extra capabilities of the platform are used and exploited in order to address some of the limitations of TeleGaze discovered in the previous phase of the research. More details of these capabilities and their role in improving the performance of the system are covered in the coming relevant sections.

4.3 Application Domain

One objective of TeleGaze is to investigate the possibility of developing an application-independent interface that can be used for a range of navigational tasks in teleoperation contexts. This was the reason behind using a rather generic navigational task in the usability experiment of the previous phase of the research. However, evaluating interactive systems gets more complicated without specific contexts and specified application requirements [26]. Brooke argues that "*it is impossible to specify the usability of a system without first defining who are the intended users of the system, the tasks those users will perform with it, and the characteristics of the physical, organisational and social environment in which it will be used*" [91].

On the same subject, Ravden and Johnson argue that "*evaluating an interface requires evaluators to carry out realistic tasks using the system as part of the evaluation*" (p17, [77]). Therefore, it is necessary to understand the user's needs and skills in order to develop evaluation metrics with this information in mind [95]. Although TeleGaze is applicable to a wide range of mobile robot teleoperation applications, it was decided at this stage of the research to narrow down the application domain. This is in order to enable better design of navigational tasks and clearer understanding of the results of the usability experiments. Figure 4.2 illustrates the three most likely applications for TeleGaze and their requirements in general.

Extra hardware components, such as robotic arms or grippers, are required for both library and supermarket robots. Therefore, it was decided to withdraw them from the list of the most likely applications for TeleGaze. In addition to hardware requirements few other constrains exist, such as the need for close positioning in order to pick up objects for example. The need for close positioning adds to the difficulty of the task. Although "*teleoperation is fatiguing and stressful even without the requirement for close positioning*" [81], minimizing the need for close positioning reduces the workload that the task creates. Furthermore, fine movements are required in these two cases in order to move between large numbers of crowds, as it is likely to be the case in supermarkets and libraries. Therefore, the closest and most realistic application of TeleGaze is a museum or gallery robot.

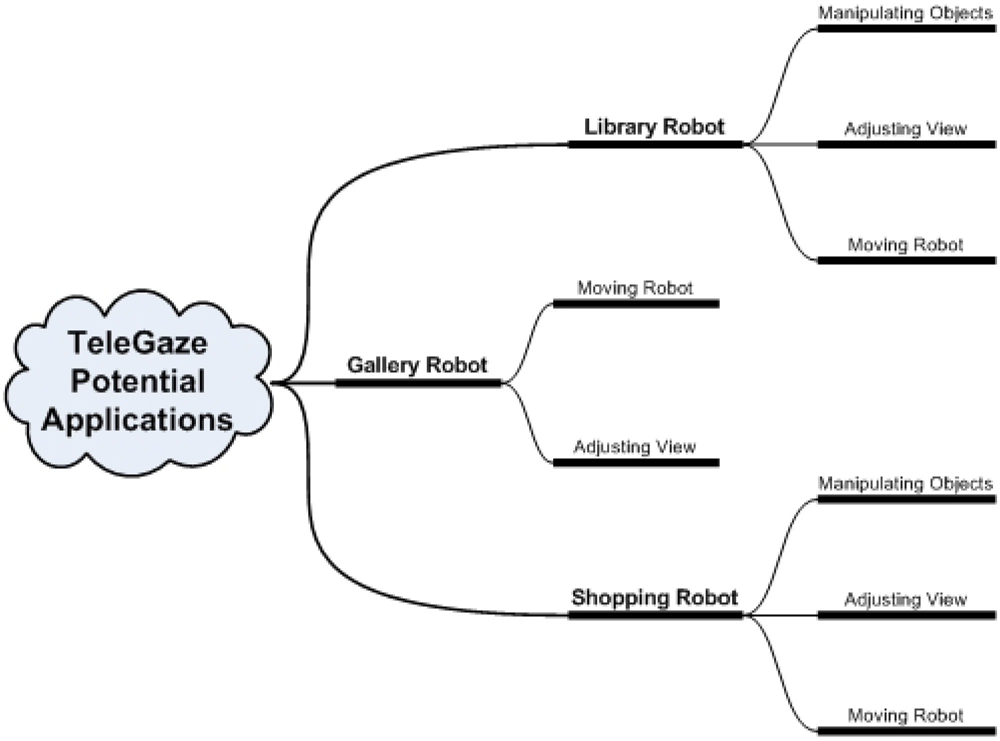


Figure 4.2: TeleGaze most likely applications and their functionality requirements.

Driving a robot around a gallery has less constrains in comparison with the other two applications and it is used to experiment new teleoperation interfaces [34]. For gallery robots, less fine movements are required as galleries are less likely to be as crowded as supermarkets or libraries. Also once the painting or the artefact is visible to the operator, then no further manoeuvring is required. In most cases no object manipulation is required too. On the other hand, more natural and quicker scanning is required over the object of interest or the artefact. Another very interesting characteristic of gallery applications is that poor adjustments of the robot can be compensated by adjustments of the camera and vice-versa.

A very interesting example of a teleoperation application, where TeleGaze might come very handy, is the GestureMan. The GestureMan is used by a remote operator to give explanations about some of the exhibits in a science museum [96]. The orientation of the GestureMan's head, which consists of three cameras, is controlled by a joystick to project the orientation of the operator's gaze. Using TeleGaze in this case, saves the need for a joystick and gives the operator more opportunities to concentrate on the explanations and not controlling the head of the robot.

Narrowing the application domain of TeleGaze to gallery robots does not contradict with the objective of designing an application-independent interface. It only helps constructing more realistic experiments to test the usability of the interface. TeleGaze still enjoys a wide range of likely applications where it can substantially benefit the operator by allowing hands-free teleoperation.

4.4 Experiment Design

Eye movements are known to be task-dependent [26]. Therefore, extra care must be taken in designing a task-oriented evaluation usability experiment. Task planning also means scenario planning as scenarios can be represented as sequences of tasks in specific orders [97]. Considering the application domain and a mixture of likely life scenarios, a navigational task for TeleGaze was designed at this stage of the research. The scenario behind the task states that a human operator is driving a mobile robot in a gallery like environment to inspect a number of paintings. The task this time is more demanding than the task used in the previous phase of the research because it is meant to push TeleGaze to its limits.

Tasks used to evaluate any system should be as close as possible to the work that is to be carried out using the system. Also they should test as much of the system as possible [77]. Therefore, the navigational task requires more than just basic navigation sub-tasks because TeleGaze is meant to handle more than basic navigation tasks. Also, the task has a number of specific requirements to test certain features of TeleGaze. For example, in addition to driving the robot around the gallery like environment, there were paintings hung around the environment which the operator needed to inspect and report results back. The navigational task and the galley like environment are illustrated in Figure 4.3.

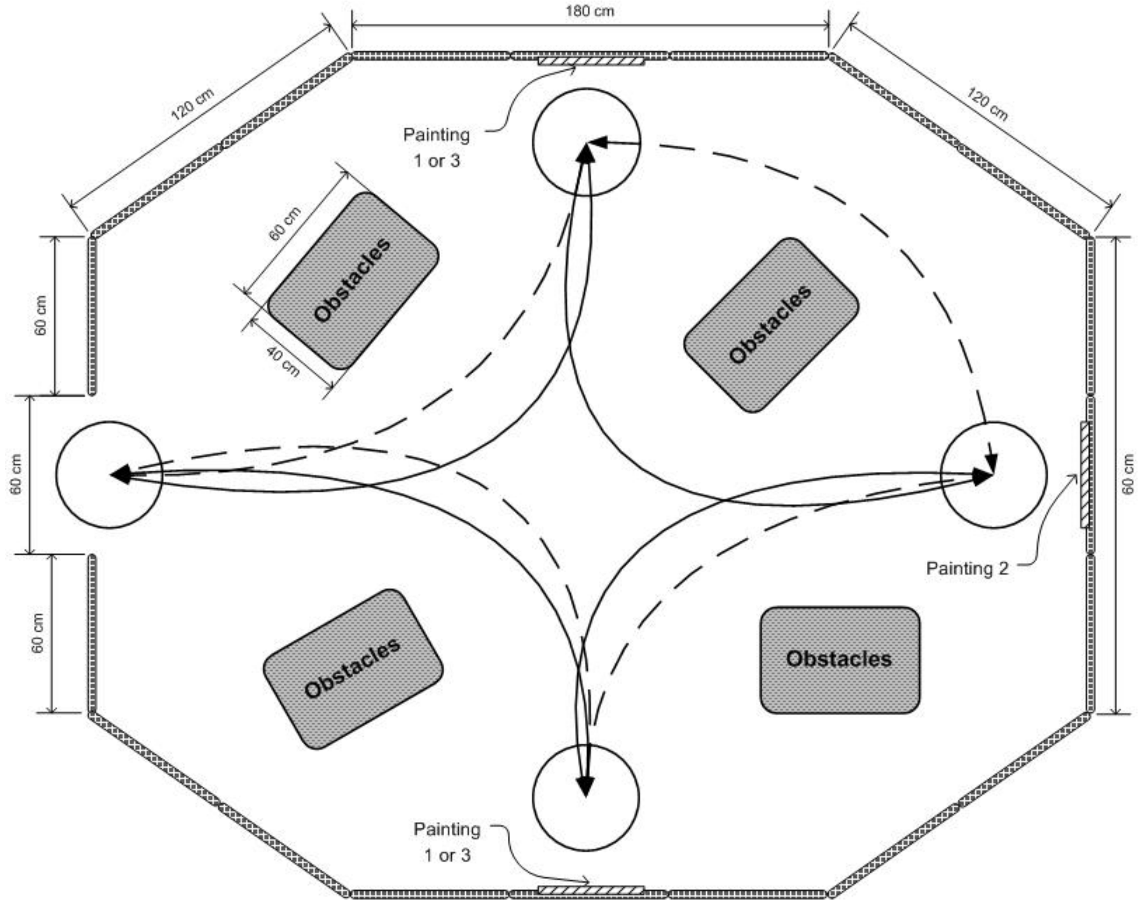


Figure 4.3: The layout and the likely routes of the navigational task.

To gain better understanding of the navigational task, following are the requirements of the task and some fine details of the sub-tasks because different types of subtasks require different amount of workload [83].

4.4.1 Paintings' Contents

Three different paintings with different contents are hung in the environment. The contents of the paintings are designed to measure how the concentration level of the operator gets affected in different modes of interactions. For this purpose, the contents of the paintings are different sets of numbers in different colours. In addition to driving the robot around, the operator is required to write down the summation of a particular set of numbers from each painting. This ensures that the operator actually is getting close enough to the paintings to be able to read the contents. Also it is used in measuring the operator's concentration level through the results of the mathematical operation required when getting to any of the paintings [1]. Even when getting to the paintings the operator is required to scan across the whole painting to get the necessary readings for the mathematical operation.

A further interesting purpose of asking for the results of the mathematical operation to be written down is to test the advantage of hands-free driving. In the case of using the joystick for example, extra time is required to put down the joystick, get the pen, write down the results, and hold the joystick back. Whereas in TeleGaze, both hands are free any way, which means the time required to change between the joystick and the pen is saved. Details of the contents of the paintings are illustrated in Figure 4.4.

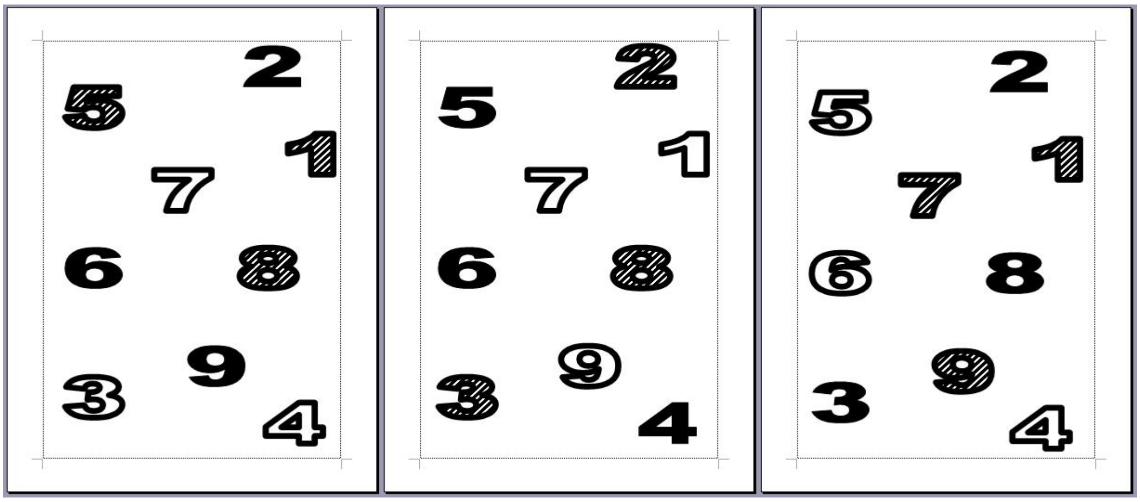


Figure 4.4: The contents of the paintings to be inspected as part of the task.

4.4.2 Paintings' Positions

In the previous navigational task most of the participants did not use any of the camera controlling functionalities as it was not necessary to complete the task. Therefore, no evaluation results were obtained as far as this part of the interface is concerned. Because the camera controlling functionalities make a significant part of the TeleGaze interface, it is necessary to be tested and evaluated in the navigational task. In this task, the paintings are hung at different heights in accordance to the robot height. This forces the operators to use the camera controlling functionality of the interface to inspect different paintings otherwise they would not be able to obtain the readings.

In terms of heights, one of the paintings is hung higher than the robot's line of sight, one of them lower, and one of them at the same level of the robot's line of sight.

Camera height adjustments depends on the skills of the operator because it is correlated to the distance between the robot and the painting. Figure 4.5 shows the differences in the height of the paintings in accordance to the height of the robot.

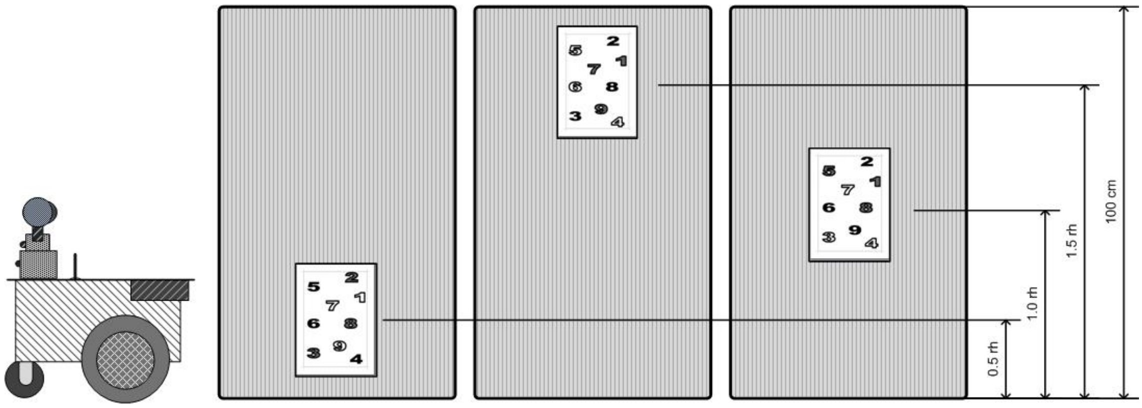


Figure 4.5: Paintings heights in relation to the height of the GazeBot line of sight. (rh = robot height)

4.4.3 Positioning Obstacles

To simulate a real life navigation scenario in a gallery like environment a number of obstacles are positioned in the environment (Figure 4.3). The positioning of the obstacles are designed so that there are more than one obvious route to drive from one point to another. This is unlike other works where in order to save the operator from getting lost, the navigational task has only one possible route [4]. It is also ensured that reading the next painting is not possible from the position of the current painting due to the heights of the obstacles in addition to their positioning. This forces the operator to actually navigate the robot to get to a different point where inspecting the next painting is possible. Hence, it is required to navigate the robot among the obstacles in order to get to the different paintings.

4.5 Anatomy of TeleGaze for Evaluation Metrics

The novelty of TeleGaze requires a specific set of evaluation metrics that evaluates the system and the interface from a multidisciplinary point of view. In order to clarify and determine the involved disciplines, it is necessary to look into what constitutes TeleGaze as an interactive system. TeleGaze can be defined as ***a human operator sitting in front of a computer screen interacting with a mobile robotic platform via an intelligent user interface using inputs from her eyes.*** This thorough definition can be used to pick out all the different disciplines that constitute TeleGaze. The following anatomy therefore, highlights those disciplines:

- **Human-Computer Interaction (HCI):** Quoting from the definition it can be seen that ***"a human operator sitting in front of a computer screen"*** implies that there is an interaction going on between the human operator and a computer. This shows that TeleGaze can be looked at from a HCI point of view since the actual interaction is taking place between the human operator and a computer.

- **Human-Robot Interaction (HRI):** Quoting again from the definition using ***"a human operator …... interacting with a robotic platform"*** is rather self explanatory in the sense that there is interaction between a human operator and a robotic platform. HRI therefore, is one of the disciplines that constitute TeleGaze since the aim of the interaction is interacting with a robotic agent and not a computer.

- **Intelligent User Interface (IUI):** Another quote from the definition is ***"via an intelligent user interface"*** which shows that the interaction is achieved using an intelligent user interface. Thus a third discipline is IUI. Interacting with IUI is not necessarily limited to computer or robotic applications. Interacting with mobile devices through IUI is an example of applications relatively outside the world of desktop computers and robots.

- **Eye Tracking:** The last part of the definition which is ***"using inputs from her eyes"*** explains the input channel of the data which is the operator's eyes. This shows that eye tracking is another discipline that constitutes TeleGaze. Although eye tracking might be the least mature disciplines in comparison with the previous three, it is mature enough to have a set of specific evaluation metrics and scientific methodologies.

It is difficult to decide which one of these disciplines is the dominant discipline in TeleGaze. For example, the aim of TeleGaze is to control a mobile robot platform. However, the actual interaction is not taking place between the operator and the physical robot. Therefore, it cannot be considered as a pure HRI due to the fact that the human operator is not interacting with the physical robot. Also because of the aim mentioned above it cannot be considered HCI although the human operator is actually interacting with a desktop computer. Eye tracking on the other hand cannot be considered as the dominant discipline of TeleGaze as it only consists part of the system both conceptually and physically.

The common evaluation metrics used in any of the mentioned disciplines are not completely strange to the ones used in any of the other disciplines. HCI being the most mature discipline, most of the evaluation metrics used in HRI, IUI, and eye tracking are derivatives of evaluation metrics commonly used in HCI [26]. However, HRI and IUI differs from HCI in a number of different dimensions [98]. Therefore, it was decided to derive a set of evaluation metrics from the most common evaluation metrics that mostly coexist in all the involved disciplines instead of simply following textbook knowledge [39]. This set of evaluation metrics should meet the requirements mentioned in the introduction section. Most importantly, it should evaluate TeleGaze from a multidisciplinary point of view.

4.6 Scope of Evaluation

To systematically select a set of evaluation metrics from the vast number of evaluation metrics used and recommended in all the disciplines mentioned above, a filtering mechanism is necessary. For this purpose, a comprehensive evaluation metrics tree is built at this stage of the research (Figure 4.6). Heading down from the top of the evaluation metrics tree, the set of evaluation metrics that are likely to suit TeleGaze gets closer to the scope of the research. Following is the route that has been taken to narrow down the domain of evaluation:

- **One-human/One-robot:** HRI problems are not limited to only one-human/one-robot, but this is certainly one important type [36]. Therefore, the scope of TeleGaze is limited to one-human/one-robot interaction for the purpose of this research. That eliminates the set of evaluation metrics common in the other three forms of interactions which are one-human/multiple-robots [99], multiple-humans/multiple-robots, and multiple-humans/one-robot [100].

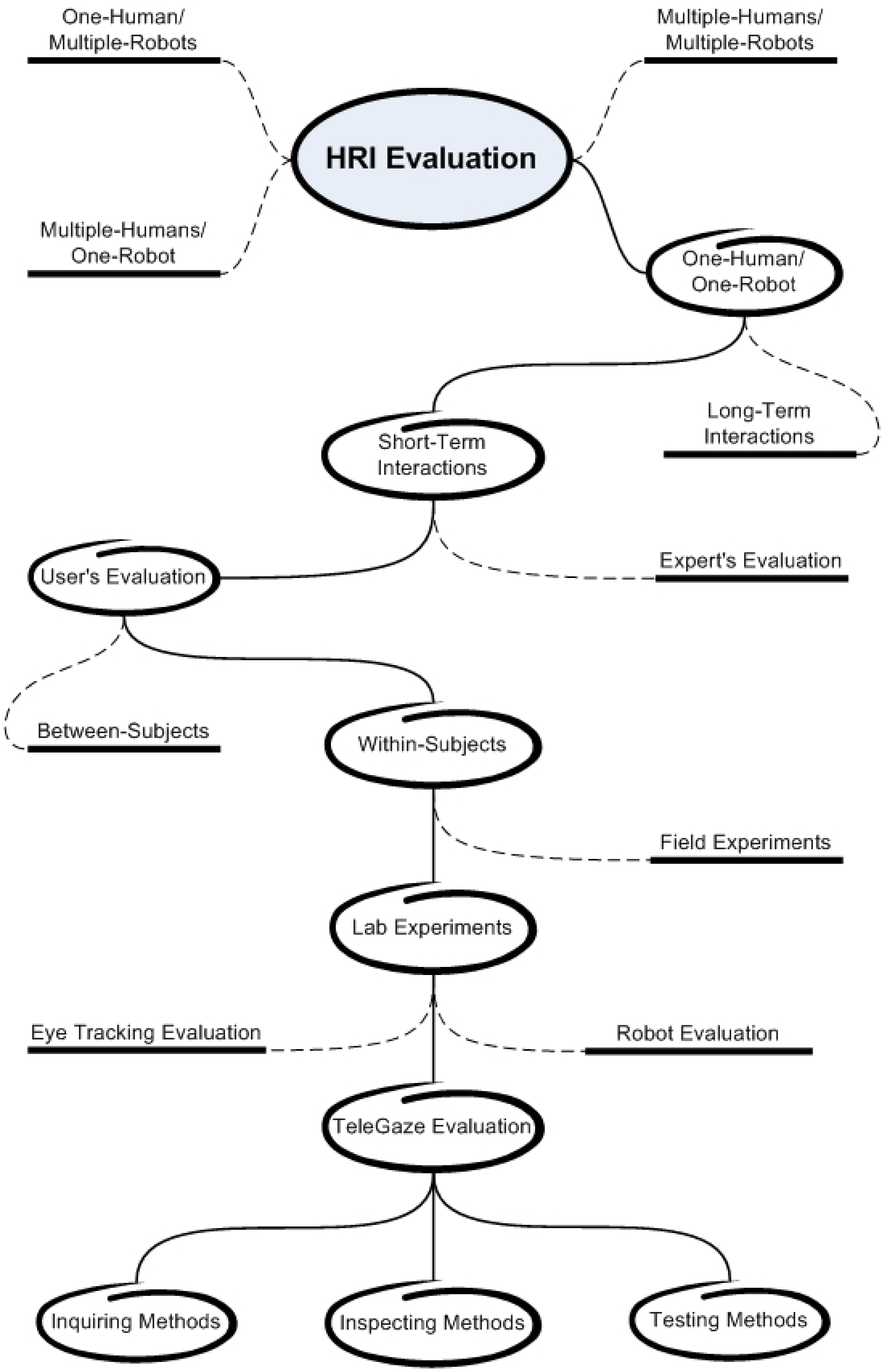


Figure 4.6: HRI evaluation tree. Continuous lines are used for selected routes.

- **Short-Term Interactions:** Based on the duration of interactions, short-term interactions and long-term interactions require different approaches of evaluation. Long-term interaction studies are demanding in terms of labour, time and equipments [39]. Also due to the limitations of current eye tracking equipments, such as the need for recalibration over longer periods [26], mostly short term interactions are considered in gaze driven systems [85]. In some cases, if the interaction extends beyond few minutes, researchers intervene to provide help and end the interaction as soon as possible [74]. Following the same approach, only short-term interactions is considered in the evaluation of the TeleGaze.

- **User's Evaluation:** Due to the novelty of TeleGaze, it was decided not to consider expert's evaluation as part of the set of evaluation metrics. Expert's evaluation are mostly used for systems with previous histories where experts have knowledge in terms of general expectations and standards. This is less likely to be the case in TeleGaze. Therefore, only user's evaluation is considered.

- **Within Subjects:** As far as the group of participants is concerned, either a within-subjects or a between-subjects design is used in the evaluation of interactive systems. A within-subjects design uses one group of participants for different conditions, while a between-subjects design uses different groups for different conditions. For a between-subjects design, more participants are required and statistical analysis may be complicated if there is too much variance among the participants groups [26]. Therefore, a within-subjects design is followed for TeleGaze.

- **Lab Experiments:** Special eye tracking equipments are necessary for conducting field experiments which require flexible placements of camera and light sources [61]. Therefore, the limited mobility of the used eye tracking equipment in TeleGaze is a key factor in considering lab experiments rather than field experiments. Also in general, lab experiments can be better controlled than field experiments [32]. Therefore, only lab experiments are going to be used in the evaluation of TeleGaze.

- **Interaction Experience:** As it is set in the research boundaries (Chapter 1), this work does not involve neither design and evaluation of any robotic platforms, nor any eye tracking platforms. The focus is the interaction experience that TeleGaze creates for the operator and the hardware components of the system. Therefore, the evaluation is limited to the interaction experience of TeleGaze additional focus on the TeleGaze interface as the interaction medium. However, observations regarding the hardware components of the system are taken into considerations too.

4.7 Selection of Evaluation Metrics

Different methods of evaluation exist and have been followed by practitioners in any of the disciplines that constitute TeleGaze. However, each method has its limitations when it comes to analysing the results and generalizing the findings. Therefore, using more than one method is recommended in the evaluation of any particular system. This approach is likely to overcome the limitations of a particular method by the use of another method [19]. For this reason, more than one evaluation method and more than one set of evaluation metrics are used in the evaluation of TeleGaze. The followings are the details of the selected methods and the reasons behind selecting particular metrics:

4.7.1 Testing Methods using Quantitative Metrics

Since a robot is a dynamic system of which the primary job is to accomplish tasks through execution of motions [97], the main requirement in the usability experiment is to accomplish a navigational task. Therefore, testing methods using quantitative metrics are highly applicable and necessary. The performance of the system, including the operator, can be measured with quantitative metrics in terms of efficiency and effectiveness. Hence, two different forms of quantitative metrics are included in the evaluation metrics. Followings are those quantitative metrics with specific definitions and details of their use in the context of TeleGaze.

a. Efficiency

Efficiency is one of the most common metrics used in the evaluation of interactive systems [26]. The definition of efficiency is contextual and likely to differ from one interactive system to another. However, the most common definition of efficiency is the time-to-complete a particular task using the system under evaluation [90]. The same definition of efficiency applies in the context of TeleGaze. Therefore, efficiency is measured as the time-to-complete the navigational task in the usability experiment.

b. Effectiveness

Effectiveness is even more contextual than efficiency and in most cases it can be defined in a number of different ways. It can be defined as the number of errors or damages occurred during the performance [1], [26]. Alternatively, it can be defined as the amount of tasks completed accurately [90]. For the purpose of TeleGaze effectiveness is defined as the overall goal achievement. In addition to driving the robot around the environment, this includes avoiding the obstacles and reporting correct readings from the paintings. To better quantify this, the overall percentage of effectiveness is calculated based on the following expression:

$$e=[(0.2\times c_{rds})+0.4-(0.1\times h_{obs})]\times 100 \qquad 4.1$$

Where c_{rds} is the number of correct readings and h_{obs} is the number of hitting obstacles.

The award is calculated as 10% of the overall effectiveness for avoiding any obstacle and 20% of effectiveness for reporting any correct reading. Reporting the readings is awarded higher level of achievement due to the mental and concentration demand it requires in addition to all the manoeuvring efforts. The maximum effectiveness therefore, is where all four obstacles are avoided and readings from all three paintings are reported correctly (i.e. [(0.2 x 3) + 0.4 – (0.1 x 0)] x 100 =100%).

4.7.2 Inquiring Methods using Subjective Metrics

The interaction in TeleGaze takes place between a human operator and the TeleGaze system. Therefore, the user is one major part of the interaction and her attitude is very significant in determining the future directions of TeleGaze. Subjective metrics have been widely used in the evaluation of HCI systems [32]. In fact they are recommended for evaluating most interactive systems in general [26] including HRI systems as measures of the quality of the effort [90]. Although subjective metrics are more exposed to individual interpretations in comparison with objective metrics, they are highly valuable when cross analysed with the results from the objective metrics. More interestingly, cross analyses is possible within the subjective metrics themselves. This is highly recommended to ensure the integrity of the obtained results from the subjective metrics.

Subjective metrics are highly flexible and can be constructed entirely for evaluating the system in question. However, some subjective metrics are more tested than others for a wide range of applications. Therefore, in order to minimize the effects of individual interpretations of the subjective metrics' results, two different sets of subjective metrics are used in the evaluation of TeleGaze. A set of specifically designed questionnaire is used in order to get results regarding some specific aspects of TeleGaze. Also NASA-TLX, as another set of subjective metrics, is used to get a more general insight into the TeleGaze system. Followings are the details of the subjective metrics used in the evaluation of TeleGaze.

a. Specifically Designed Questionnaire

As a subjective metric for measuring user satisfaction level, a set of very carefully designed questionnaire is included in the evaluation metrics of TeleGaze. Through a rating scheme, mostly Likert, the participants rates their agreements with a number of pre-designed statements. The statements address the questions that are believed necessary to be answered through the evaluation. It is often assumed that a Likert scale is based on forced-choice questions, where a statement is made and the respondent then indicates the degree of agreement or disagreement with the statement [91]. Although this can be seen as an advantage of getting answers to specific questions and narrowing down the interpretation possibilities, it can be seen as limitations to statement based questionnaires. To overcome this limitation therefore, some open questions are included in the questionnaire in addition to the statements. Also some other considerations are taken into account in designing the questionnaire. Followings are some of the main considerations.

The questionnaire is divided into two sets of questions. The first set of questions addresses the interaction experience for each mode of interaction. Therefore, the same set of questions, eight in total, is filled after the navigational task is completed with each mode of interaction. Access to the answers of the previous mode is not allowed while answering the questions for the current mode. This is in order to avoid any influence to the answers based on the previous mode of interaction. This method also insures better analysing possibilities when it comes to result comparison of the different interaction modes. The first set of questions are presented in Table 4.1.

Table 4.1: The first set of questions. Repeated for all modes of interactions [26].

No.	Statements	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	I found it easy to learn how to use the system	1	2	3	4	5
2	The system has all the capabilities I expect it to have	1	2	3	4	5
3	I felt confident using the system	1	2	3	4	5
4	It was difficult to get familiar with the system and how it works	1	2	3	4	5
5	I don't think I can perform better than how I did	1	2	3	4	5
6	I don't see the system appropriate for this kind of applications	1	2	3	4	5
7	I would like more chances to try more difficult tasks	1	2	3	4	5
8	Overall, I am satisfied with the system	1	2	3	4	5

Within the set of the questions above, some of the key questions are repeated but in a paraphrased way with opposite direction. This ensures that the participant knows what is the question about and answers with attention to the question. If the answer to question one for example (*I found it easy to learn how to use the system*) is positive, then the answer to question four (*It was difficult to get familiar with the system and how it works*) should be negative. This is because these two questions imply the same meaning but with different directions. If the answers do not match, then either the participant has not understood one, or both, of the questions or she has answered the question without paying attention.

To gain a better understanding of the participant's attitude, the answer to the final question in the first group (*Overall I am satisfied with the system*) should match the average answers from the other questions in the same group. This question although seemingly very simple and plain, is a common approach in designing questionnaires [101]. The answers to this question is compared to the overall answers to the rest of the questions in the same set. This ensures a tight cross-analysing within the answers of each group of questions for each particular mode of interaction for each particular participant.

---

[26] The questions used in the questionnaire are mostly obtained from [19] and [26], but paraphrased and tailored to fit the application context of TeleGaze.

The second set of questions in the questionnaire addressed the design elements of the TeleGaze interface. A different type of scale is introduced in this section for two of the questions. This is in order to change the style of answers required and is meant to increase participant's attention when filling the questionnaire. Also two open questions are added in order to allow more freedom if participants wanted to comment or raise any concerns. A whole version of the questionnaire, including the first set of questions, is presented in Appendix A (Section A.3).

b. NASA-TLX

The aim of TeleGaze is to enable mobile robot teleoperation through eye gaze in order to minimize body engagement. To achieve this aim, compromising certain elements of the user interaction experience might be necessary. Along this compromise, freeing the hands of the operator while adding constrains to the eyes is likely to be accepted, only up to a certain limit. To gain better insight into the task workload, this limit - the limit of compromising - needs to be measured however. Comparing the task workload measurements for each mode of interaction gives better understanding of the trade-off that is necessary to achieve the aim of TeleGaze. Therefore, it is necessary to measure the task workload for each mode of interaction. This includes the overall task workload and particular components of the workload. The individual components and their definitions are included in Appendix A (Section A.4).

NASA-TLX is used, as part of the subjective metrics, to measure the task workload for each mode of interaction for comparison purposes. NASA-TLX is a popular evaluation tool in interactive systems and it has been proved to produce reliable results with applications to HRI [102]. It also is more known to the community than less familiar task workload measuring tools. Hence the measurement results produce clearer and better understandings of the system in question. Although the definitions of individual components of the tool can be altered to fit a particular application, the default definitions used in TeleGaze because they suit the context.

4.7.3 Inspecting Methods using Monitoring and Recording Metrics

Due to unfamiliarity of people with the practice of controlling through eye gaze, TeleGaze might raise a number of psychological and physiological concerns. Self reports and interviews might not reveal what people actually do because people not necessarily report, or even know, what they do [74]. Using inspecting methods is likely to reveal some understandings of the reactions of the user while interacting with the system. Mostly qualitative metrics are going to be used as inspecting methods which might be more difficult to quantify. However, qualitative metrics reveal interesting information regarding uncounted behaviours. Therefore, the following inspecting methods are included in the set of evaluation metrics.

a. Eye Tracking Data

In some cases, traditional usability metrics might reveal a range of usability problems while they may be enhanced by additional measures such as eye movement recordings [88]. Eye tracking data can provide both quantitative and qualitative information about the two common stages of visual search which are perceptual and cognitive. There are some known problems in analysing and using eye tracking data. An example of such problems is the difficulty to determine whether the subject is thinking about the task or something else when measuring cognitive activities through eye tracking [84]. Another example is the difficulty of recognizing similarities in eye movement patterns between individuals due to individual differences [103]. However, eye tracking data visualization and analysis is believed to gain insights into the subject's attentive behaviour and is frequently viewed as a window into the internal cognitive processes [26], [74].

One form of extracting diagnostic information from eye tracking data is based on fixations. The average duration of fixations usually reveal the amount of cognitive load required to understand the scene or reveal the attention allocation. In these cases duration of fixations negatively correlate to the efficiency of task execution [78] and longer fixations indicate higher cognitive demands [74], [103]. This is the most common and probably the best use of fixations for diagnostic purposes. However, this approach is not likely to be so effective when used for interactive interfaces. In interfaces that react to fixations, which is the case in TeleGaze, the fixations can be far from being correlated to the cognitive demand. Therefore, not all the fixations can be correlated to attention or cognitive demands. Hence, they cannot be correlated to efficiency [78] because subjects have to fixate whether by, or against, their will.

Another form of extracting diagnostic information from eye tracking is the variations in pupil dilation during the course of interactions. Not all experiments have shown the relationship between task difficulty and pupil dilation. However, some show that pupil dilation is a reliable and valid measure of mental workload [78]. Therefore, it might be very appealing to use this indicator to measure the operator's mental workload throughout the navigational task in the usability experiments.

Pupil dilation is very sensitive to environment variations and extra care must be taken as far as environment illumination is concerned. Controlled illumination is more achievable with static and controlled interface backgrounds such as when subjects interact with a document on the screen. With dynamic backgrounds, such as real-time images from the video camera in TeleGaze, it is almost impossible to control the illumination in the background. Thus pupil dilation cannot be trusted as a measure for mental workload. Also, if successful, pupil dilation measures differences in the workload throughout the execution of a task. The aim of the usability experiment is to measure the difference in workload for different modes of interactions and not during task execution for individual modes. Using NASA-TLX should reveal an overall index of workload including mental workload in a more useful form for the purpose of the usability experiments in TeleGaze.

For the reasons mentioned above, the use of eye tracking data in the traditional way is not applicable to TeleGaze. Fixation durations do not reveal cognitive demands as they are not affected by cognitive demands, but rather by action demands in TeleGaze. Also pupil dilations do not reveal differences in mental workload between different interaction modes, but rather during one interaction mode. To overcome the limitations of using eye tracking information in the traditional way, it was decided to aid the information with other forms of inspecting methods. As Jacob and Karn argue "*eye tracking alone is not a complete usability engineering approach, but it can make a significant contribution to the assessment of usability*" [40]. Therefore, video recordings of the scene, which is the same through the eyes of the robot and the operator, is combined with eye tracking data to fit the purpose of the usability experiment of the TeleGaze system.

b. Video Recording

It is both feasible and useful to log and process interaction events [84]. Video recording is reported to be the richest source of information for usability experiments [95]. The recorded data can be granulated to time-stamped, task-stamped, or action-stamped data for comparison purposes which then reveals stamped specific information. In general, video recording of the subject's behaviour during the interaction sometimes reveal interesting and useful information. Because the eye tracking data and the video recordings complement each other, the combination of both are recorded during the experiments. The result is the actual interface that the subject interacts with during the experiment containing the real images from the on-board camera and the gazing data projected on top of it. Going through the video recordings of each subject for each interaction mode reveals the flow of the task and difficulties the subject has, if any. The data includes the actions that the subject issues at any particular time and/or stage of the teleoperation process.

Although eye tracking data is not needed for the interaction using the joystick, the data is recorded and logged for the evaluation purpose. The interface presented to the user when using the joystick is free from any action regions. However, the eye tracking data recorded while using the joystick is projected on an interface that has all the action regions. This is to compare the distribution of the fixations when using the joystick with the positions of the action regions in the TeleGaze interface. This reveals a clear comparison of the distribution of the fixations between the different modes in accordance to the positions of the action regions on the TeleGaze interface.

4.8 Other Evaluation Metrics

In addition to the evaluation metrics that are used in the evaluation of TeleGaze, other evaluation metrics are used in the evaluation of systems with similar purposes and requirements. *Human error*, for example, is well known to affect the performance of the system and the overall completion of a task. However, it is not included in the evaluation metrics because one of the causes of human errors in teleoperation is, for instance, lack of feedback information. In all the modes of interactions in the usability experiment of TeleGaze the same level of situational awareness[27] is provided. Therefore, the possibilities of human error is the same for all modes of interactions [3], which means no comparable results are produced.

Another example of other evaluation metrics is *think-aloud* [26]. Although think-aloud is a valuable engineering usability method, it may influence the way the user attends to a certain task and change the patterns of the gaze. Furthermore, it might affect the performance measurements as performing a task is likely to take longer when verbalized [26]. Also people cannot always verbalize what they do [74]. Eye tracking data can be used as an alternative to think-aloud with the advantage of no cognitive imposition on the operator during task performance [104]. Therefore, this evaluation metric too is not included in the set of the evaluation metrics. Similarly, the lack of direct applicability of a number of other evolution metrics is the reason behind not considering them for the evaluation of TeleGaze. As Benyon and colleagues say "*just because something can be measured, it does not mean that it should be*" [19].

4.9 Participants

Researchers have different opinions as far as the number of participants necessary for similar usability experiment studies. Suggestions vary from as few as six to as many as twenty participants [84]. Although even fewer than six is used in some cases [62], [81], recommendations state ten as an in-between number for usability experiments of interactive systems [19], [32], [84]. Furthermore, some researchers use different numbers for different stages of usability experiments of the same system [4]. In the usability experiment of the previous phase of the research (Chapter 3), a group of ten users, two females and eight males, aged between 22 and 45 years old participated. Based on the recommendations above, the criteria for the participants stays the same for the coming usability experiments.

Also in the previous usability experiment, there were participants with high familiarity of using computers and high familiarity of using joysticks. Non of the participants however, had any prior familiarity of using eye tracking systems or

---

[27] Level one of situational awareness, which is enough information to proceed [98], is provided in TeleGaze.

robotics. Therefore, in order to create the same level of familiarity among the participants, four participants from the previous usability experiment are included in the coming usability experiments. This is because they have gained some experience in using eye tracking systems and robotics. With this approach, the level of familiarity of using computers, joysticks, and eye tracking is distributed evenly among the participants. It is common to use pre-test questionnaires to test participants familiarity and background knowledge [102]. However, this approach is not followed in order not to overwhelm the participants with answering a long list of questions.

4.10 Conclusions

The set of evaluation metrics designed at this stage evaluates TeleGaze from a multidisciplinary point of view using a variety of methods. It can be argued whether this is the best set of evaluation metrics that can be used to evaluate an interactive system or not. However, in order to keep a practical balance between the design of TeleGaze and its evaluation, it is necessary not to include evaluation metrics that are not highly related and applicable. Furthermore, adding extra evaluation metrics, even with potential contribution to the usability experiment, increases time, cost, and personnel demands, most likely beyond the access of the research.

This stage of the research revamped the evaluation metrics used in the previous phase of the research. Although it is realized that the basic principles behind the usability experiment stayed the same, re-definition and re-design of some of the metrics are found to be necessary. This is in order to obtain more informative usability testing results, such as in the case of re-defining effectiveness and re-designing the questionnaire. Extra evaluation metrics were added to the set of previous evaluation metrics such as NASA-TLX, eye movement data and video recordings.

In addition to the set of evaluation metrics, more sophisticated experiment design and more specific application domain are believed to aid in the usability experiment. Taking all this into account, it is believed that the results of the usability experiments produce significant amount of useful information that help determining the usability of TeleGaze in comparison to conventional means of HRI.

Chapter FIVE

Multimodal TeleGaze

5.1 Introduction

Natural human-human interaction (HHI) is the ultimate aim for human-computer interactions (HCI) and human-robot interactions (HRI) [48]. Since HHI naturally does not depend on single modes of interactions, experimenting multimodal approaches is a must in HCI [28] and HRI systems [10]. When one single mode of interaction fails to totally fulfil the requirements of a particular system, "*multimodality provides the crucial key*" [66]. Although multimodality has advantages on one hand, it has certain disadvantages on the other hand [38]. It is not certain that multimodality is always the best solution. Experimenting the approach is necessary for each particular system and application context. Therefore, experimenting the approach was necessary to see if any improvements in the system's performance and user satisfaction can be achieved.

It is likely that additional requirements, both functional and non-functional, emerge as the design process of any interactive system goes on [19]. Therefore, in addition to experimenting with a multimodal approach, this stage of the research includes some major refinements in the TeleGaze interface. This is mainly driven by the findings in the usability testing experiment from the previous stage of the research

(Chapter 3). It is also driven by the extra capabilities that the updated robotic platform (GazeBot) provides, such as feedback on the current pan/tilt angles of the camera.

Findings from the previous stage of the research show that more than one source of limitations exist in the native TeleGaze. Using dwell-time to trigger an action, issues in the design of the interface, and the eye tracking equipment used are the main three sources of these limitations. This chapter therefore, starts with the details of the multimodal approach as the obtained solution to address the dwell-time problem. It then, moves to the details of the refinements of the TeleGaze interface to address the design issues. The limitations of the eye tracking technology and the equipments are also covered. Using the redesigned set of evaluation metrics and the redesigned navigational task (Chapter 4), the details of the experiment of the multimodal TeleGaze are presented. Finally, the findings from the experiment and key conclusions from this stage of the research are discussed.

5.2 Action Confirmation

5.2.1 Dwell-Time, the Problem

One of the common problems in using inputs from human eyes for gaze contingent interfaces is the *midas-touch* problem [24]. This is the problem of distinguishing between fixations that are necessary to obtain information on a point and fixations that are required for confirming an action. This happens because eyes are always engaged and every point on the interface is likely to become eye activated. One of the common solutions to this problem is *dwell-time* [44]. Dwell-time is the time a fixation, or more than one, needs to take in order to be registered as a confirmation for an action. However, depending purely on inputs from the eyes for both scanning and selecting, or as a direct controlling device, raises a number of concerns.

Barcelos [24] argues that "*the anatomical properties of our eyes give us indication that completely eliminating the manual operations can overload the eyes with a manipulation task that they are not prepared for*". On the same subject, Zhai [38] argues that "*it is unnatural to overload a perceptual channel such as vision with motor control task*". The constrains and unnaturalness of overloading human eyes with controlling tasks affects the performance of the system, and most importantly, the interaction experience of the human operator. In addition to the constrains, dwell-time can only substitute for one click [38]. This too plays a significant role in preventing the dwell-time from being the perfect solution for the midas-touch problem.

The native TeleGaze used dwell-time as a solution for the midas-touch problem (Chapter 3). However, the findings from the usability testing experiment highlighted the dwell-time as one of the main sources of the system's limitations. It was observed that the participants had difficulties in knowing the duration of the dwell-time exactly. In some cases, fixations for longer time than required caused extra commands to be issued than what the subject originally intended. When steering for example, extra commands led the robot to face a different direction than the intended one, which then required reverse steering. In some other cases fixations for shorter than required caused the intended actions not to be issued with one attempt. Both situations caused some frustrations for the subject. Many of the participants therefore, did not find dwell-time a convenient solution for action confirmation. Instead, a more deterministic form of confirmation was believed to be more convenient.

5.2.2 Multimodal, the Solution

Extra modalities have been added to gaze-driven interfaces for various reasons, but mostly to overcome one main problem. As mentioned above, depending purely on inputs from human eyes for such interfaces creates the midas-touch problem (Section 5.2.1). Using dwell-time to solve this problem, although is common, poses a number of other challenges. Therefore, other approaches instead of dwell-time, such as using additional natural and artificial modes of interactions, are followed to tackle the midas-touch problem.

Although selection by dwell-time is considered more natural than blinking [105], blinking is one of the natural forms that is used for action confirmation [43]. Speech is another example of the natural forms that is used for the same purpose [106]. On the other hand, the computer mouse [38], the spacebar [87], and certain keys of the keyboard [79] are among the less natural modes of interactions that have been used too. In some cases, both natural and non-natural modes of interactions are integrated into the same system [10]. Also, novel interfaces have been experimented as additional modalities for gaze-driven interfaces such as brain-computer interfaces (BCI) [107].

Although common and seemingly promising, multimodal approaches do not always produce better results at all levels. Results vary among reported works. Where some elements of performance or user satisfaction have been improved by the multimodal approach, other elements have been compromised [38]. In [107] for example, the multimodal approach produced better accuracy than the dwell-time. However, it resulted in slower performance. On the other hand, where higher speed has been achieved, the accuracy has been affected by the multimodal approach [87].

One of the main reasons for these problems is the required coordination between the different devices that produce the final instruction for the system. In [87] for example, where the spacebar is used instead of dwell-time for action confirmation, participants "*either pressed the spacebar before fixating or after*". Similar phenomena where participants leave the focus zone before pressing the button has been reported elsewhere. This phenomena has been referred to as "*leave-before-click*" [79]. The fact that certain problems are likely to appear in multimodal approaches indicates that some elements are necessary to be compromised. To determine the nature and the significance of this compromise, empirical results are necessary to be collected for any particular application context.

5.2.3 TeleGaze, the Multimodal

Designing an interface that is driven purely by inputs from human eyes for mobile robot teleoperation is achievable, as the native TeleGaze proved so. However, it turned out to be less efficient and less satisfactory when compared with conventional modes of interactions, such as a joystick. The disadvantage of being stressfully careful not to issue a command unintentionally seemed to overshadow the advantage of hands-free teleoperation. Adding inputs from an additional device to the inputs from the human eyes while keeping minimum human body engagement, seems to aid in better achieving the aim of TeleGaze. Therefore a multimodal approach is experimented at this stage of the research, as an additional control to show operational context [21].

Non of the extra modalities mentioned above (Section 5.2.2) is believed to be the best for the TeleGaze system. Using the computer mouse [38] or the keyboard [79] and [87] as a confirmation mechanism contradicts the TeleGaze aim of hands-free teleoperation. On the other hand, adding extra constrains to the eyes, such as specific blinking pattern [43], affects the naturalness of the interface. Therefore, a different form of confirmation mechanism is necessary to create a multimodal TeleGaze interface. Considering the fact that many people have driving experiences, and considering mobile robot teleoperation mainly as a driving experience, an accelerator pedal is believed to best suit the multimodal TeleGaze.

An accelerator pedal is used as a contextual triggering mechanism. The aim of the pedal is to eliminate the likelihood of unintentional actions. Regardless of the fixation points and their durations, and regardless of the status of the pedal, both inputs should match in order to produce a command. The final instruction to the system results from the combination of the fixation point and the status of the pedal at any time. For example, if fixations happen in the forward action region and the pedal is pressed, then forward commands are issued. If fixations happen in the turn right action region and the pedal is pressed, then turning right commands are issued. Therefore, for an action to be confirmed fixations, naturally, still need to occur in the action regions.

TeleGaze still mainly depends on inputs from human eyes. The functionality of the pedal only substitutes the need for the dwell-time. The intentions for the actions and the directions of the movements are determined by the direction of the gaze. The actual act however, is triggered with pressing the pedal. This natural combination of the two modes is likely to cause less physical constrains and fatigue, while producing better performance when compared with the dwell-time. It also ensures that TeleGaze still creates a natural HRI experience with the advantage of hands-free teleoperation. Therefore, this multimodal approach does not contradict with the aim of TeleGaze.

5.3 Interface Design Issues

The refined-interface shown in Figure 5.1 was built on the findings of the observational study discussed earlier in section 3.5 that was carried out for two different prototypes (Figure 3.11). Some design parameters, such as the size of the action regions, were selected as experimental values to be modified at the later stages of the research based on empirical findings. As expected therefore, some design issues were observed during the usability testing experiment of the native TeleGaze (Section 3.7).

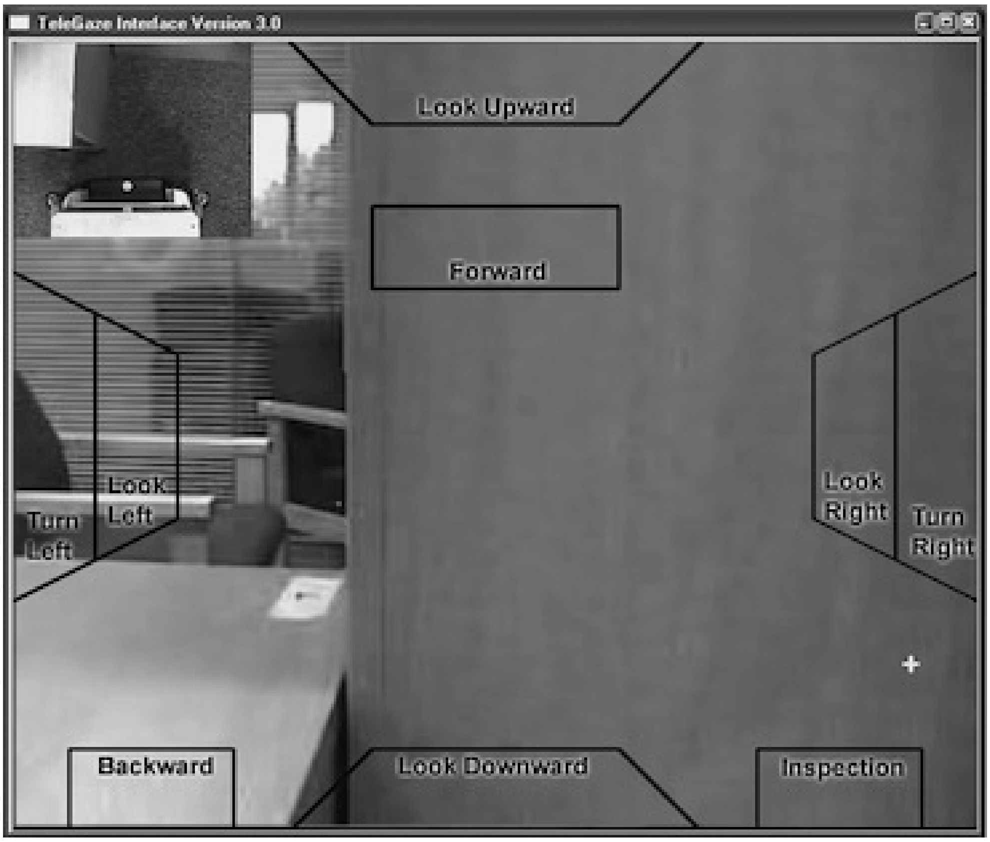


Figure 5.1: Another snapshot of the refined-interface (RI). Displayed here for ease of comparison.

One example of the design issues is the display from the downward looking camera in the upper-left corner of the interface. This display, as mentioned previously, was provided for better situational awareness when it comes to monitoring the close surroundings of the robot. To comply with the accuracy requirements of the navigational task, the participants spent a significant amount of time monitoring this display. Monitoring this display added to the overall time to complete the task and affected the efficiency of the system since efficiency is calculated based on time-to-complete task.

This issue is no longer a problem since GazeBot, the robotic platform that is used from this stage onwards, provides the required level of situational awareness with one single camera, that is the main view. Therefore, a secondary view from a downward looking camera is not necessary anymore. Also the accuracy requirements in the coming experiments are different from the requirements of the previous experiment in that monitoring marked tracks is not necessary. This saves the subject the time that was spent in the native TeleGaze usability testing experiment on monitoring the track.

Another example of the design issues is the adjusted positioning of camera and robot turning actions. Despite clear captioning for both regions, it was observed during the usability experiment that the subjects experienced some level of confusion. Prior to the experiment, it was believed that issuing one command instead of the other, turning right instead of looking right for example, is consequent free. However, the participants experienced additional stress due to this issue while under the stress of performing the task. In addition to personal observations, two of the participants mentioned this issue explicitly in their comments.

Another issue that was affected by this is the distance between the turning action regions and the forward action region. When consequent issuing of turning and moving forward commands was required, the participants found the distance between these regions affecting their achievements. This was also observed to be due to the lack of incremental steering action regions. Prior to the usability experiment, it was believed that incremental steering, such as moving on a curvature, can be achieved with the right combination of moving forward and turning commands. However, it was observed that incremental steering would have been more convenient than forcing the subject to use a combination of moving forward and turning commands.

To address these design issues, some major refinements in the design of the interface are believed to be necessary. These refinements are also driven by the capabilities that GazeBot, the new robotic platform, provides in comparison with the platform used in the previous stage of the research. The refinements include the layout and the design of the action regions, some added functionalities to the interface, and rethinking some of the design parameters. Figure 5.2 shows the layout of the interface designed at this stage of the research and Figure 5.3 shows an actual snapshot of it.

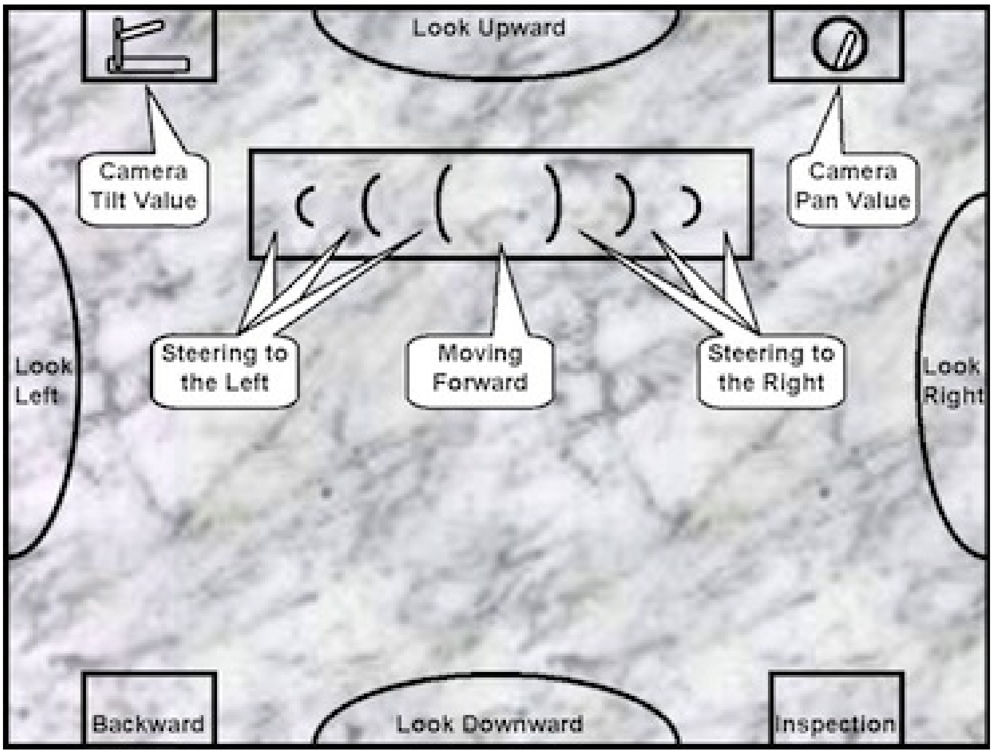


Figure 5.2: The layout of the multimodal TeleGaze interface.

Followings, are the details of the major refinements carried out at this stage of the design of the TeleGaze interface:

5.3.1 Steering Action Regions

One of the main objectives of TeleGaze is to achieve intuitive interpretations of the gazing behaviours into teleoperation commands. This is mainly targeted through intuitive positioning of the action regions on the interface. The forward action region for example, is positioned where people look naturally if they want to move forward. Unlike "*turning the eye from the centre position to extreme up position and coming back to the centre without delaying*" to issue a forward command [50], the positioning of the action regions in TeleGaze matches their functions. However, it was observed during the usability testing experiment of the native TeleGaze that the same does not apply to all the action regions. Turning right/left for example, did not seem intuitive enough for the participants to avoid confusion with looking right/left. Therefore, a totally different approach is used to position the turning right/left action regions on the interface at this stage of the research.

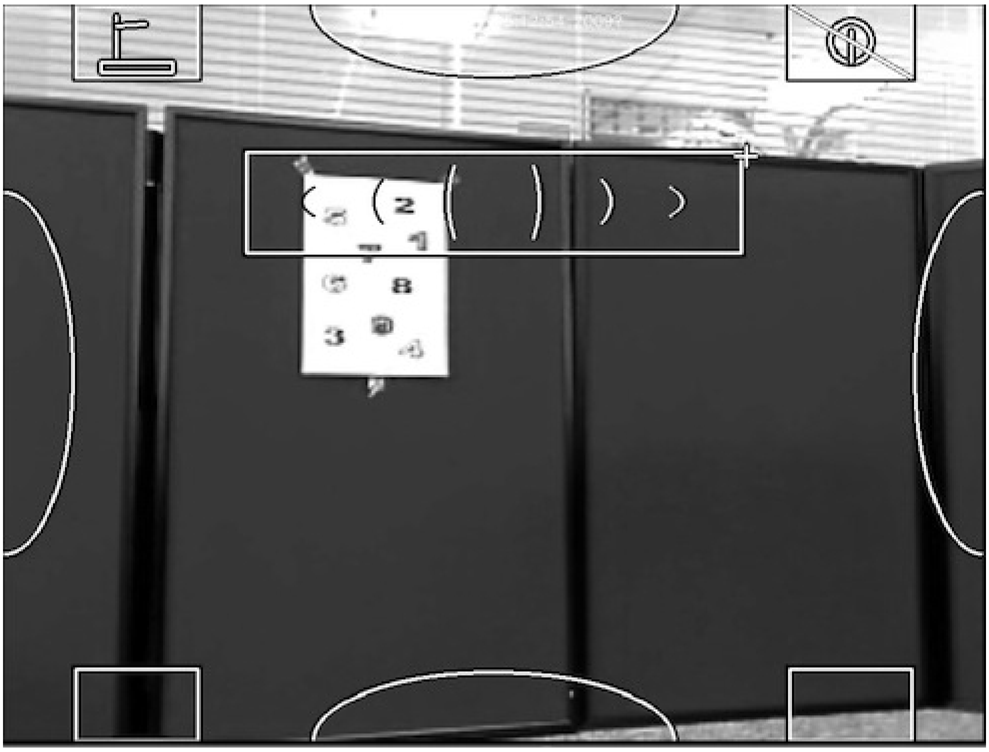


Figure 5.3: An actual snapshot of the multimodal TeleGaze interface.

According to Gestalt's law of continuity "*we tend to perceive smooth, continuous patterns rather than disjoint, interrupted ones*" (p114, [19]). Using a fuzzy representation of the forward and turning right/left action regions, one major modification in the interface is the continuous *forward/turning* action region. This representation brings all robot controlling action regions together, except backward action region. It extends the forward action region to include the turning right action regions on the right, and the turning left action regions on the left. A shorter distance between the forward and the turning action regions than that of the refined-interface (RI) is achieved, which results in less physical demand on the eyes of the subject [88].

In terms of functionality, this design provides a granulated steering control over the robot through a fuzzy combination of linear/angular speeds. This functionality was not possible with the previous platform due to limitations in the hardware components of the platform. Therefore, this is one of the refinements that is driven also by the capabilities of GazeBot. The forward/turning action region is magnified and illustrated in Figure 5.4.

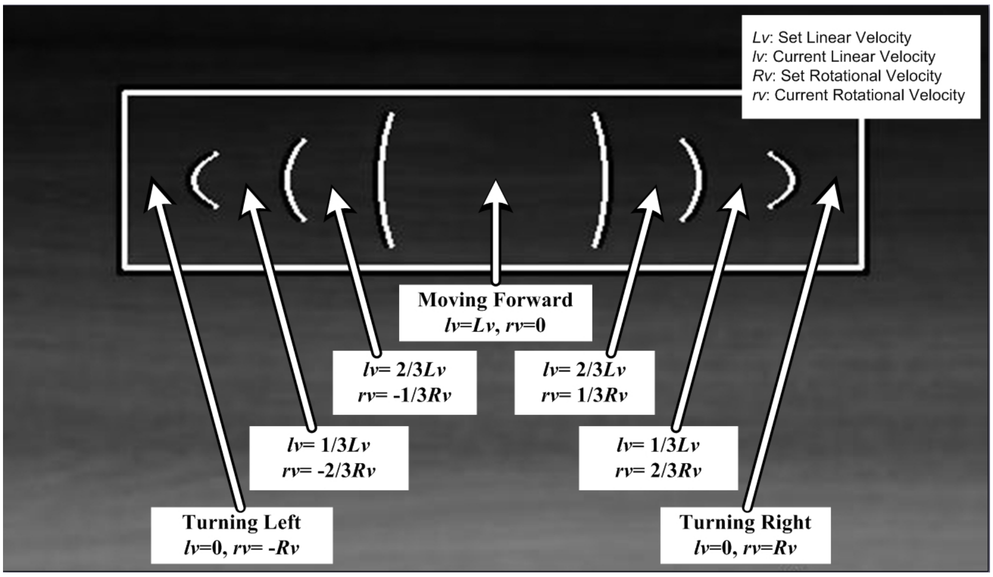


Figure 5.4: Details of the forward/steering action region in the multimodal interface.

The introduced design of the forward/turning action region is believed to provide a more desired interaction experience than what was experienced in the previous stage of the research. Also the granulated steering functionality is believed to deliver a smoother control when it comes to directing the robot. Moving the robot related action regions out of the edges of the interface, clears the way for the camera related action regions. Hence, a more intuitive positioning of the camera control action regions are achieved in this design which adds to the overall intuitiveness of the interface.

5.3.2 Different Geometric Shapes

The action regions on the interface can be mainly divided into *robot related action regions* and *camera related action regions*. Moving forward, backward, turning right, and turning left compose the robot related action regions, while looking upward, downward, right, and left compose the camera related action regions. Moving turning right/left action regions out of the way of looking right/left action regions resulted in a significant difference in the layout of the interface. With no robot related action regions along the sides of the interface, only camera related action regions occupy these areas (Figure 5.2). This resulted in positioning the camera related action regions more intuitively than that achieved previously (Figure 5.1).

The *looking up* action region for example, is located in the centre of the upper edge of the interface. This is where people naturally focus if they intend to look upward and move the camera upward. The same principle applies to looking down, right, and left action regions. With this design, the position of each camera related action region is believed to be highly correlated to the natural gazing behaviour of that action. In addition to this achievement in the design of the interface, further refinements are believed to aid in the intuitiveness of the interface.

In RI (Figure 5.1), similar geometric shapes are used for both robot related action regions and camera related action regions. To better distinguish between the two categories of action regions, different geometric shapes are used in the redesigned interface for this stage of the research (Figure 5.2). Increasing visual cues to distinguish between the action regions is believed to increase learnability of the interface. Recognition seems to be easier than recall (p105, [19]). With different geometric shapes, better visual recognition can be achieved and hence, less confusion while under the stress of performing any tasks. This approach has nothing special to do with the functionalities of the interface. It is instead, the "*look and feel*", which matters significantly in the user experience [19]. This is one of the new features of the redesigned interface that is included in the evaluation of the interface.

5.3.3 Camera Alignment Information

In addition to the status of the robot as a whole, the operator needs the current status of any robotic sensors [98]. The upgraded robotic platform, GazeBot, is equipped with some feedback mechanisms that provide the current status of the pan-tilt-unit (PTU). It provides the current values of both the pan and the tilt angles, individually, at any time. Therefore, one of the major refinements in the redesigned interface is driven by the capabilities of the upgraded robotic platform which also meets one of the needs of the operator.

In this version of the interface, the operator is provided with the current status of the PTU in real-time. The pan/tilt values of the PTU are represented through two individual action regions at the upper edge of the interface. The displayed pan/tilt values on the interface are synchronized with the actual pan/tilt values of the PTU, which determines the pan/tilt values of the camera alignment.

In the previous designs of the interface explained in chapter 3, the automated camera home configuration aligns the camera with the robot alignment whenever a robotic action is issued. Depending on the operator's desire, this capability was enabled or disabled through the interface. One limitations of this functionality was that both the horizontal and the vertical automated alignments are either enabled or disabled together. Therefore, the operator did not have the capabilities to set one of the values (pan or tilt) to automated alignment and the other to controlled alignment. In this design however, this capability is provided for further control over the alignments of the camera.

In realistic application scenarios, the operator does not necessarily require having the camera aligned both vertically and horizontally with the robot alignment. Horizontal alignment (pan angle) is more crucial when the robot is moving around as it affects the orientation of the operator. However, vertical alignment might be more useful if controlled and set at different values than being aligned with the robot alignment. Therefore, the extra level of control over the pan/tilt values of the camera individually is believed to add significantly to the feeling of being in control when using the interface. In order to follow the same design language, the automated realignment for pan/tilt is activated through the same action region that displays the pan/tilt status. When the pan/tilt is set to automated realignment, a diagonal line is drawn across the pan/tilt action region.

5.3.4 Action Regions Size

As it was mentioned in chapter 3, very limited information is reported in the literature regarding the design criteria for the layout of gaze-driven interfaces. Specifically, the size of looking zones (action regions in TeleGaze) is rarely reported with sufficient details on the criteria for selecting particular sizes. Most design parameters are selected empirically for each application context. Therefore, an experimental value for the size of the action regions was selected in the previous stage of the research. The experimental value was selected so that it provides a reasonable balance between action regions and no-action regions on the interface. Since most action regions on the interface were in rectangular shapes, this value constrained the smaller dimension of the action regions.

The approach above required the participants to adapt their natural gazing behaviours to the design of the interface. Achieving a required balance between action and no-action regions on the interface is a system-centred approach. Being human-centred on the other hand, is not easy to achieve. It involves "*observing people, talking to them, and trying ideas with them, which is expensive in terms of time*" [19]. A human-centred approach however, seemed tempting to be explored after the initial stage of the research. "*Both the structure and the functionality of the human visual system components place constrains on the design parameters of a visual communication system*" (p38, [26]). Therefore, a different approach for selecting this design criterion, the smaller size of the action regions, is experimented with in this stage of the research.

The maximum acuity of the human eye is one of the natural characteristics that is different from one person to another. However, it is believed to be around 1º visual angle at the centre of the eye [26]. Therefore, one of the human centred approaches is to maintain, at least, 1º visual angle for the smallest dimension of looking zones [24]. Since this value is considered as a minimum value, a visual angle of 1.5º is selected as the determining criterion for the smaller dimension of the action regions on the TeleGaze interface.

For the eye tracking equipment to perform at its best, the optimum distance between the subject's head and the interaction screen is 65cm. Although it is unlikely that this distance can be secured for every participant, it is used as the basis for deciding on the size of the action regions. A visual angle of 1.5º at this distance is equivalent to 3cm on the screen. Therefore, 3cm is considered the minimum value for the smallest dimension of the action regions on the interface. Using this approach, the overall area occupied by action regions is nearly ¼ of the total area of the interface. Coincidently, when compared with RI in Figure 5.1, this is the same area as that was occupied there. However, due to change in the overall layout of the action regions and their shapes, this ¼ ratio resulted in larger action regions than that of RI.

5.3.5 Fixation Calculations

In order to mainly meet some frequency matching requirements, a duration of 330ms was used as the dwell-time value to register an action in the native TeleGaze interface. GazeBot is a more flexible platform than the previous robotic platform, hence more flexibility can be obtained in terms of the necessary frequency matching. When possible "*designers need to put people rather than technology at the centre of their design process*" (p3, [19]). The duration to register a fixation this time therefore, is mainly driven by the natural characteristics of human eyes, instead of the characteristics of the used hardware.

As it was covered in chapter 3, findings and recommendations as far as the fixation duration is concerned vary significantly despite the fact that even minor changes in the parameters might affect the results dramatically. Ranging from 50ms to 750ms [74], [87], it is difficult to decide on the optimum fixation duration unless extensive experiments are conducted and empirical values are obtained for particular systems. Therefore, a duration of only 200ms is used in this design of the interface to register fixations since this value is believed to "*provide an adequate balance between speed of interaction and accuracy*" [25]. With this fixation value, a maximum of 5 commands/second can be obtained if fixations continuously happen in one action region while the pedal is pressed. This is more (2commands/second) than what was achieved in the previous stage of the research, which is believed to increase the efficiency of the TeleGaze system.

5.4 MoSCoW Rules

Additional requirements emerge as the design and evaluation of any interactive system goes on [19]. These requirements follow the *MoSCoW* rules[28], where:

- **M** stands for *must have* requirements that make the system work.
- **S** stands for *should have* requirements that are essential if resources permit.
- **C** stands for *could have* requirements that can easily be left for later stages.
- **W** stands for *won't have* requirements that are not necessary for this stage.

---

[28] The details of this section is mainly obtained from Benyon and colleagues (p212, [19]).

The interface design issues mentioned above (Section 5.3) lie in the first three categories of the MoSCoW rules. Steering action regions and camera alignment information for example are believed to be *must have* requirements. Different geometric shapes and redesigned action region sizes are believed to be *should have* requirements. Fixation calculations are believed to be *could have* requirements. The accelerator pedal, as a deterministic form of action confirmation (Section 5.2), is believed to be a *must have* requirement too. The usability testing experiment however, determines the correctness of associating these requirements to their categories in the MoSCoW rules.

5.5 Eye Tracking Equipment

The state of the art of eye tracking technology puts some limitations on the performance of available eye tracking equipments and their applications. Benyon and colleagues argue that "*very often the technology gets in the way of people and the goals they want to achieve*" (p58, [19]). Therefore, one of the main challenges facing gaze driven interfaces is the eye tracking technology itself [38]. In some cases, a totally different approach than eye tracking, such as face tracking, is used to obtain the direction of the gaze because "*it is easier to measure accurately*" [85]. In TeleGaze, the findings from the previous usability experiment draws the same conclusions[29].

Drifting, which is defined as the difference between the actual point-of-gaze (POG) and the measured POG [7], is a rather technical problem that is mainly caused by the eye tracking equipment itself. It might be caused by other reasons too, such as insufficient focus or head movements during calibration [8]. One of the big issues in drifting is that it builds up as time goes on. Whether it starts with the beginning of the session or during the session, it gets to a point where recalibration is required and the session has to be stopped. This is one of the main reasons that eye tracking experiments are generally kept short.

Calibration in itself, is another well known problem in using eye tracking equipments. Calibration requires the operator to gaze as steadily as possible on a single point on the screen when that point is registered. This is a very difficult task to fulfil due

---

[29] The eye tracking equipment has been ranked by the majority of the participants (7 out of 10) as the most likely cause of any limitations in the system (Chapter 3).

to the jittery nature of human eyes. Any shift from the point of calibration during the registration of the point causes drift in the later use of the system. This was one of the clearly observed points during the usability experiment in the previous stage of the research. In some cases it was necessary to repeat the calibration before embarking on the actual experiment because "*calibration is the only way to ensure the highest accuracy of recorded data*" (p178, [26]).

One example of a problem specific to the used platform is the optical head tracking that the system uses. The eye tracking camera comes with an optical head tracking functionality which tracks slight movements in the head of the operator to keep the ball of the eye in the centre of the camera. While this feature works well most of the time, it causes problems at some other times [44]. For example, when the camera looses the reflections from the eye of the operator, it grabs any similar sort of reflections in the scene. These are not necessarily meaningful reflections and do not relate to the eye of the operator. This situation caused the camera to track something else rather than the eye of the operator which in turn caused panic for the operator. This problem causes increase in the overall time to complete the task, as well as extra workload and stress on the human operator.

In addition to these technical problems, other types of problems are likely to be faced when using eye tracking equipments. Noise and data loss for example, due to eye moisture and blinks are amongst eye tracking specific problems that are caused for physiological reasons [25]. Some other problems exist on the other hand that are caused for biological reasons. Disassociation between the gazing behaviour and the visual attention of the human operator is an example of such problems [79]. This problem is difficult both to measure and to analyse because eye trackers track the movements of the eye and not the movements of the visual attention [26].

Whether these problems are specific to the used eye tracking equipment or they are rather general, they have their effects on the overall performance of the TeleGaze system. Two out of the three sources of limitations, which are the interface layout and the action confirmation mechanism, are addressed. However, the third cause, which is the eye tracking equipment and the technology, is beyond the scope of this work.

5.6 Multimodal TeleGaze Usability Testing Experiment

The multimodal TeleGaze enables hands-free mobile robot teleoperation. This is a significant advantage if no tradeoff in the usability of the system is necessary. Therefore, the aim of the experiment at this stage of the research is to measure the usability of the multimodal TeleGaze against the usability of the conventional joystick. The usability of the conventional joystick, as it has been mentioned, is the target for the multimodal TeleGaze to meet.

Comparing the native TeleGaze to the multimodal TeleGaze reveals how the multimodal approach affects the usability of the system [24]. Therefore, although the main competitors in this experiment are the multimodal TeleGaze and the joystick, TeleGaze is experimented in its native form as well. This is due to the refinements in the design of the interface which also includes some changes in functionalities. Also it is due to using the upgraded robotic platform, which is different from the platform that the native TeleGaze was experimented on.

The native TeleGaze used in this stage of the research uses a different dwell-time value for action confirmation. As it was covered earlier in section 5.3 , a duration of 200ms is used to register a fixation. The same value is used for the dwell-time, which is needed to issue an action using the native TeleGaze. This is shorter than the dwell-time used in the previous stage of the research (330ms) because GazeBot is more responsive than the robotic platform used then. Theoretically, shorter dwell-time results in more actions per unit of time and hence, less overall time to complete the task. However, this is likely to add to the overall load of the task on the subject. The evaluation metrics identified in chapter 4 are designed to measure these aspects of each mode of interactions in the usability testing experiment.

5.6.1 Before Carrying Out the Task

The participants were informed of the requirements of the task in detail prior to starting the session. They were also informed of the nature and the aim of the experiment and the details of the measurements taken during and after the experiment. To minimize the difficulty of sensing the environment through feedback information only, they were shown around the environment to get an idea of the navigational task. More importantly, they were informed that this is a comparative experiment that tests each interaction mode in comparison with the other two. They were also informed that the different interaction modes are being tested, and not them.

The participants had practising sessions for each modes of interaction before commencing the actual experiment. The same amount of practising time was allowed for all three interaction modes [32]. No full scale trial sessions was allowed in order to minimize the effects of fatigue. Any concerns that the participants had regarding the experiment were answered between the practicing session and the actual experiment. The participants were informed that no assistance or information is provided once the experiment starts. This was to test the learnability of the interfaces and the systems for all three modes of interactions.

5.6.2 While Carrying Out the Task

A within-subject method is used where each one of the participants performed the same navigational task with all the three different modes of interactions [32]. To counter balance for any learning or boredom effects, the modes of interactions were shuffled in a systematic way. Familiarity of the context is one of the factors that influences attention selection behaviour [80]. Therefore, in addition to the sequence of interaction modes, the sequence of the paintings were changed from one mode to another. This minimizes familiarity with the route and the content of the paintings. Although specific orders for inspecting the paintings were forced, the possible routes were left for the subject to decide and try.

5.6.3 After Carrying Out the Task

After completing the task with each interaction mode, the participants were left to have some rest in order to minimize any boredom or fatigue effects. During the rest, they were asked to fill out the questionnaire specific to that particular mode of interaction. The details of the time-to-complete the task and accuracy were not given to the participant until all three modes of interactions were completed. This is to avoid any effects of the objective metrics on the subjective metrics.

Necessary assistance was given during filling out the questionnaire, especially with providing the NASA-TLX ratings. While providing the answers to the current mode of interaction, access to the answers to the previous mode of interactions were not allowed. This is again to avoid trying being comparative or consistent in answering the questions. The aim is to obtain answers as natural and as intuitive as possible without being affected by any other factors rather than opinions from the interaction experience.

5.7 Results and Findings

As it was covered in chapter 3, different evaluation metrics are used to evaluate the usability of TeleGaze in comparison to a joystick target. Data analysis of these evaluation metrics holds the answer to the research question. Therefore, in this section the results and findings of these measurements are presented.

5.7.1 Testing Methods

a. Efficiency

Efficiency is one of the important objective metrics that is used when task completion with user novel interfaces is in concern. Following the definition that is used throughout the research, efficiency is calculated as the overall time-to-complete the task. Statistical analysis for the efficiency of all ten participants for all three modes of interactions are conducted. Averages (statistical mean) of time-to-complete the task for all ten participants in all three modes of interactions are shown in Figure 5.5.

Based on time-to-complete the task, it can be inferred from the figure that it takes the joystick (*M*=358.4, *SE*=26.2)[30] 5.2% less time than the native TeleGaze (*M*=377.9, *SE*=41.7) to complete the task. It takes the native TeleGaze 8.2% less time than the multimodal TeleGaze (*M*=412.0, *SE*=43.2) to complete the task. Subsequently, joystick takes 13.0% less time than the multimodal TeleGaze to complete the task. This means that the joystick is the most efficient among the three modes of interactions. The native TeleGaze is the second most efficient, while the multimodal TeleGaze comes last. However, to statistically analyse these differences in efficiency and find out their significancy, the main testing hypothesis is constructed as follows[31]:

---

[30] M= mean or median, SE= standard error.

[31] This is the same testing hypothesis used in the previous stage of the research since the aim of the research is the same for all the different stages of the research.

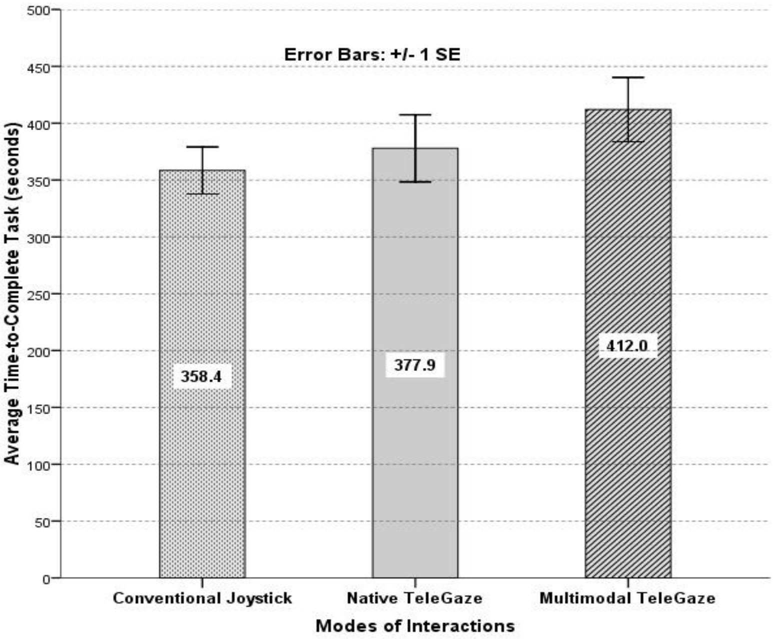


Figure 5.5: Average of time-to-complete the task for all three modes of interactions.

H_1 *: Time to complete task is different between the different interaction modes.*

The null hypothesis H_0 then, is that average time-to-complete task is not different among the interaction modes. A one-way repeated measures ANOVA[32] is used to test the hypothesis. The results of the test show that time-to-complete task is not significantly affected by the interaction mode $F(2,8) = 0.701, p > 0.05$.

The results of the statistical test show that despite the joystick being more efficient than the TeleGaze, the difference is not statistically significant. This means that *both modes of TeleGaze meet the joystick target as far as efficiency is concerned*. Although it was believed that the multimodal TeleGaze will score higher efficiency, the results show otherwise. Again, *despite the native TeleGaze being more efficient than the multimodal TeleGaze, the difference is not statistically significant*.

Since the calculated efficiency is highly related to time-to-complete the task, some observations need addressing. Being introduced to the TeleGaze as a novel interface, a high level of excitement in the participants was observed while using the system.

---

[32] Using Kolmogorov-Smirnov normality test, the results of testing for normality show that time-to-complete task for the joystick, the native TeleGaze, and the multimodal TeleGaze is not significantly different from being normally distributed D(10)=0.197, 0.225, 0.215, p>0.05, respectively (p145, [86]). Using Mauchly's test, the results of the test show that the assumption of sphericity is not violated $\chi^2(2) = 1.367$, p>0.05 (p474, [86]).

Interested in challenging the system, some of the participants tried harder navigation techniques than what was necessary to complete the task. For example, one of the participants parked the robot in parallel to the wall and used the camera to turn 90 degrees towards the painting. This technique, although accepted and interesting, required more time and effort than what is basically required to complete the task.

The extra time consumed due to trying challenging techniques, affects the average calculated efficiency of the TeleGaze. What technique to use, or what technique is best was totally left to the participant to decide. Overall, the task in this stage is more complicated and more demanding than the navigational task of the previous stage of the research. Despite the extra demands, both modes of TeleGaze meets the joystick target as far as efficiency is concerned.

b. Effectiveness

Effectiveness, as defined in expression 4.1, is one of the objective metrics used to compare the three modes of interactions. Figure 5.6 shows the results of the average of effectiveness of all the participants for each mode of interaction.

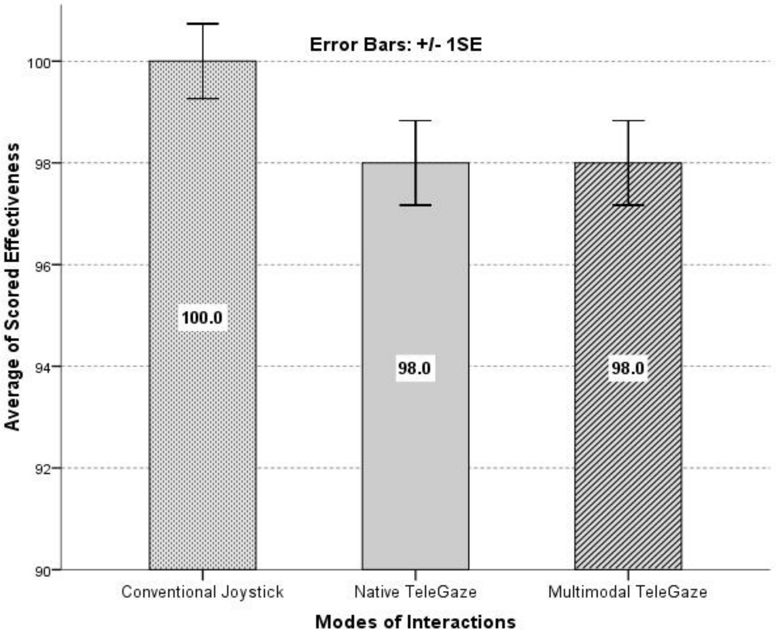


Figure 5.6: Average of scored effectiveness for all ten participants in all three modes of interactions.

It can be seen from this figure that a 100% effectiveness is scored using the joystick interaction mode. While the average effectiveness achieved with either mode of

TeleGaze is 98% of the overall goal achievement. This means that on average, the joystick is only 2% more effective than both modes of TeleGaze. Using the Friedman's ANOVA[33], the results of the test show that effectiveness is not significantly affected by the interaction mode $\chi^2(2)=2.667$, $p>0.05$. Therefore, *the statistical analysis show that as far as effectiveness is concerned, both modes of TeleGaze meet the joystick rival.*

5.7.2 Inquiring Methods

a. Questionnaire

One of the subjective metrics among the inquiring methods is a specifically designed questionnaire. The questionnaire is composed of two main sets of questions[34]. The first set of questions addresses the interaction experience for each mode of interaction. This set of questions is asked for each interaction mode in order to obtain comparable results. Some of the questions in this set are constructed in positive forms and some are constructed in negative forms. The average (statistical median) of the participants' answers to each one of the questions for each mode of interaction is presented in Figure 5.7.

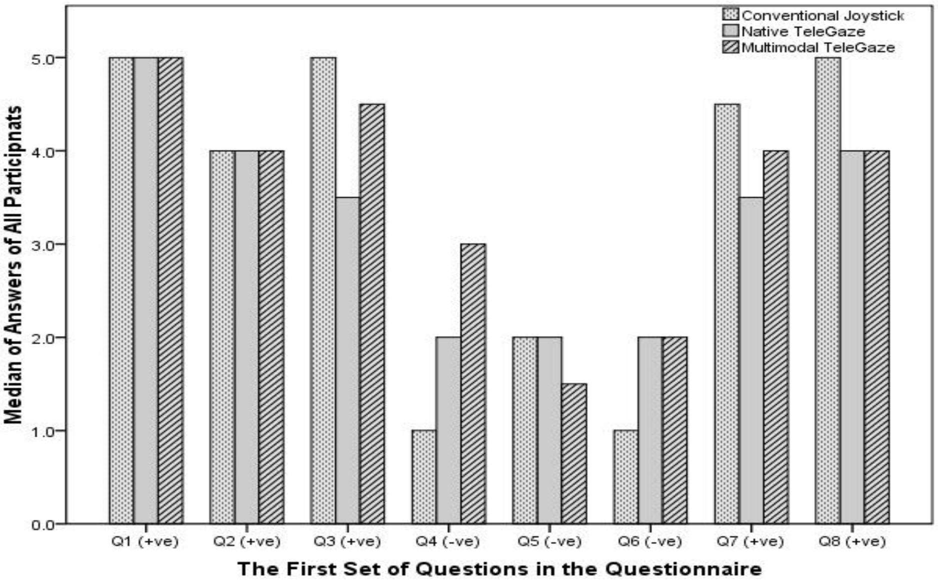


Figure 5.7: Average of the answers to the first set of questions in the questionnaire. The scale of the y-axis is (1- Strongly Disagree, 2- Agree, 3- Neutral, 4- Agree, 5- Strongly Agree).

---

[33] Using Kolmogorov-Smirnov normality test, the results of testing for normality show that effectiveness of the native TeleGaze and the multimodal TeleGaze is significantly different from being normally distributed D(10)= 0.482, 0.482, p<0.05, respectively (p145, [86]). Therefore, non-parametric tests are adopted. The scored effectiveness with the joystick is consistent at 100% for all participants.

[34] The complete questionnaire is included in Appendix B. The details of designing the questionnaire is covered in Chapter 4 (Section 4.7.2).

From Figure 5.7, the following can be inferred:

● For two of the positive questions (Q1 and Q2), the same ratings are obtained for all three modes of interactions. This means that on average, no comparative preferences can be seen. However, the ratings for all three modes of interactions are positive (strongly agree for Q1 and agree for Q2).

● For the rest of the positive questions (Q3, Q7, and Q8) higher ratings for the joystick than the TeleGaze are obtained. However, the ratings for the TeleGaze are positive too. The maximum difference between the ratings is only 1.5 agreement factor (Joystick is strongly agree while native TeleGaze is between neutral and agree in Q3).

● For three of the positive questions (Q1, Q2, and Q8) the same ratings are obtained for both modes of TeleGaze (strongly agree for Q1 and agree for Q2 and Q8). For the rest of the positive questions (Q3 and Q7), higher ratings for the multimodal TeleGaze than the native TeleGaze are obtained.

● The results obtained for Q4, theoretically, should match the results obtained for Q1, but in an opposite direction. The same applies to Q5 against Q7. The final results for questions 1 and 4 after calculations of directions put the joystick (4.5) first, the native TeleGaze (4.0) second, and the multimodal TeleGaze (3.5) last. The results for questions 5 and 7 put the joystick (4.25) first, the multimodal TeleGaze (3.5) second, and the native TeleGaze (3.25) last. Therefore, as far as TeleGaze is concerned, the results of these questions (Q1, Q4, Q5, and Q7) show no preferences of the modes (native versus multimodal). However, as far as the joystick is concerned, it is rated higher than both modes of TeleGaze.

● The final question (Q8), which is inquiring about the overall satisfaction with the system, puts the joystick higher than the TeleGaze. However, the same results are obtained for both modes of TeleGaze.

● The ranking based on the overall median of the ratings for all the questions[35]

---

[35] Complement of rating for negative questions are calculated (ex. 2 → 3 since *disagree* with a negative statement is equivalent to *agree* with a positive statement).

puts the joystick at the first place (M=4.25), the multimodal TeleGaze at the second place (M=4.0), and the native TeleGaze at the last place (M=3.25) as far as user preferences are concerned. This shows that although both modes of TeleGaze do not meet the joystick target, the multimodal TeleGaze is more preferred than the native TeleGaze.

To gain better insight into the significancy in the differences of the rankings, the following testing hypothesis is constructed:

H_1 *: User satisfaction level is different among the three modes of interactions.*

The null hypothesis H_0 then, is that user satisfaction does not differ from one mode of interaction to another. To test the hypothesis, a Friedman'a ANOVA[36] test is conducted. The results of the test show that user satisfaction is significantly affected by the mode of interaction $\chi^2(2)=6.381$, p<0.05. Since the differences in user satisfaction are significantly different, further follow-up tests are required. Therefore, the following testing hypotheses are constructed:

H_2 *: User satisfaction is different between joystick and native TeleGaze.*

H_3 *: User satisfaction is different between joystick and multimodal TeleGaze.*

H_4 *: User satisfaction is different between native and multimodal TeleGaze.*

Using the Wilcoxon signed ranks test[37], the results of the test show that user satisfaction is significantly different between the joystick and the native TeleGaze z=-2.121, p<0.05, r=0.47. The results of the test for the joystick and the multimodal TeleGaze show that user satisfaction is not significantly affected by the interaction mode z=-1.807, p>0.05, r=0.40. The results also show that between the modes of TeleGaze, user satisfaction is not significantly different by the mode of interaction z=-0.667, p>0.05, r=0.14. Therefore, it can be inferred that the native TeleGaze does not score user satisfaction as high as the joystick. *The multimodal TeleGaze however, meets the joystick target as far as user satisfaction is concerned*.

---

[36] Friedman'a ANOVA is used because answers to questionnaire are, arguably, considered ordinal values and parametric tests are not recommended (p8, [86]).

[37] The test of difference for two dependent variables. The non-parametric equivalence to the student t-test (p552,[86]).

The second set of questions in the questionnaire, addresses the design and the layout of the TeleGaze interface. This set of questions is believed to reveal some insights into the design of the interface from the participants' points of view, regardless of the mode of TeleGaze (native or multimodal). Whether the interface is used in the native or in the multimodal mode, there is likely to be some design issues that is applicable to both modes. Unlike the first set of questions which is answered after each interaction mode, this set of questions is answered only once at the end of the whole experiment. Therefore, the answers to this set of questions are influenced, implicitly, by the interaction experience of both modes of TeleGaze.

The TeleGaze interface is designed to provide both monitoring and controlling simultaneously. This is one of the main objectives of TeleGaze as highlighted in chapter 1. Therefore, the first question in this set inquires the proportion of these two elements in the design of the interface. It inquires which of monitoring and controlling the interface provides *most*, and which one the interface provides *best*. The answer to this question is restricted to the proportion of one capability (monitoring) in accordance to the proportion of the other capability (controlling). Therefore, this question is construed as follows[38]:

Which one of the capabilities do you think the interface provided most?

Monitoring	*	*	*	*	*	Controlling

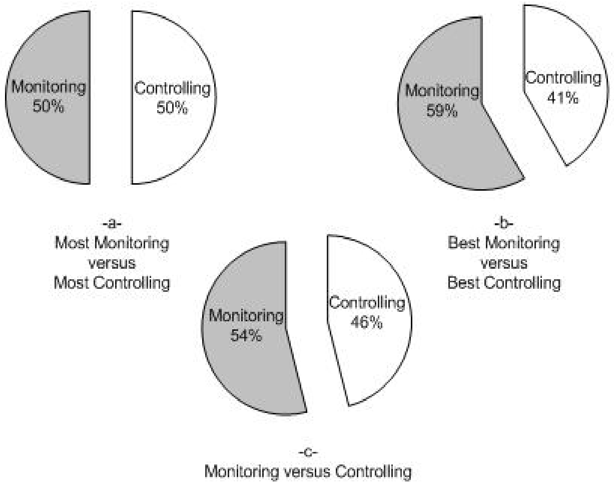
And which one of the capabilities do you think the interface provided best?

Monitoring	*	*	*	*	*	Controlling

Answering this question requires drawing a circle towards the side of the scale that the subject believes the interface provides most and best. The averages (statistical median) of the answers to both parts of this question are visualized in Figure 5.8.

---

[38] The whole questionnaire is included in Appendix B.



-a-
Most Monitoring
versus
Most Controlling

-b-
Best Monitoring
versus
Best Controlling

-c-
Monitoring versus Controlling

Figure 5.8: Average of ratings for monitoring versus controlling of the TeleGaze interface.

It can be inferred from the figure that the interface provides the same magnitude of both monitoring and controlling (Figure 5.8-a). However, it does not provide the same quality as shown in Figure 5.8-b. As a result, the participants are, on average, more satisfied with the monitoring capability than the controlling capability of the interface as shown in Figure 5.8-c. Ideally, the same satisfaction level should be obtained for both monitoring and controlling to meet the objective of TeleGaze.

To determine the significance of the difference in user satisfaction as far as monitoring and controlling is concerned, further statistical analysis is required. Using Friedman's ANOVA[39] to test the significance of the difference, the results of the test show that the difference in user satisfaction is not statistically significant $\chi^2(2)=4.667$, $p>0.05$. Therefore, despite the difference in the quality of monitoring and controlling that TeleGaze provides, it meets its objective of not compromising one of the capabilities for the other.

The rest of the questions in the second set of questions addresses other elements of the design of the interface. They do not ask for participant's agreements with pre-constructed statements as in the first set of questions. Instead, they as for participants' ratings for some features of the interface. The positioning of the action regions, their shapes, their sizes and an overall rating of the interface are all inquired in this set of

---

[39] The same reason explained above in footnote 28.

questions. The average (statistical median) of the participants' ratings for these features are visualized in Figure 5.9. The details of the features are included in Appendix A.

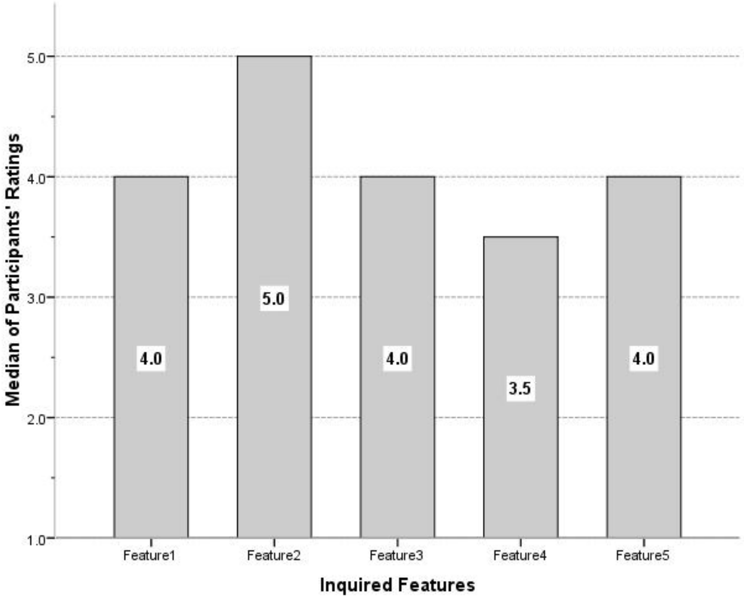


Figure 5.9: Average of the participants' ratings for the inquired features of the interface.

It can be seen from the figure that all the ratings are on the positive side of the scale. Maximum rating is obtained for the positioning of the camera related action regions (Feature 2). Minimum rating on the other hand, is obtained for the size of the action regions (Feature 4). Overall, the participants rate their interaction experience with the interface at 4 out of 5 (Feature 5). The fact that no feature is rated below 3 shows that the participants are, on average, satisfied with the interface. However, there is potential for some features to be further improved, such as the size of the action regions.

b. NASA-TLX

NASA-TLX is a questionnaire based form of inquiry that measures the amount of the experienced load in performing any task. It does not only measure the amount of the overall task load, but it also measures the amount of individual components of the load. However, it is a subjective metric that depends on inquiring methods and the obtained answers might not necessarily represent the actual measurements. That said, using NASA-TLX in combination with other subjective metrics, such as questionnaires, is likely to overcome the drawbacks that single subjective metrics generally have.

Since the aim of the usability testing experiment is to compare the usability of TeleGaze to the one of the joystick, NASA-TLX measurements are obtained for all three modes of interactions. Similar to the first set of questions in the questionnaire, the necessary answers are obtained from the participants after the task is completed with each mode of interaction. The final results of the metric therefore, are compared to each other for all the three modes of interactions. The average (statistical mean) of NASA-TLX for each mode of interaction is plotted in Figure 5.10.

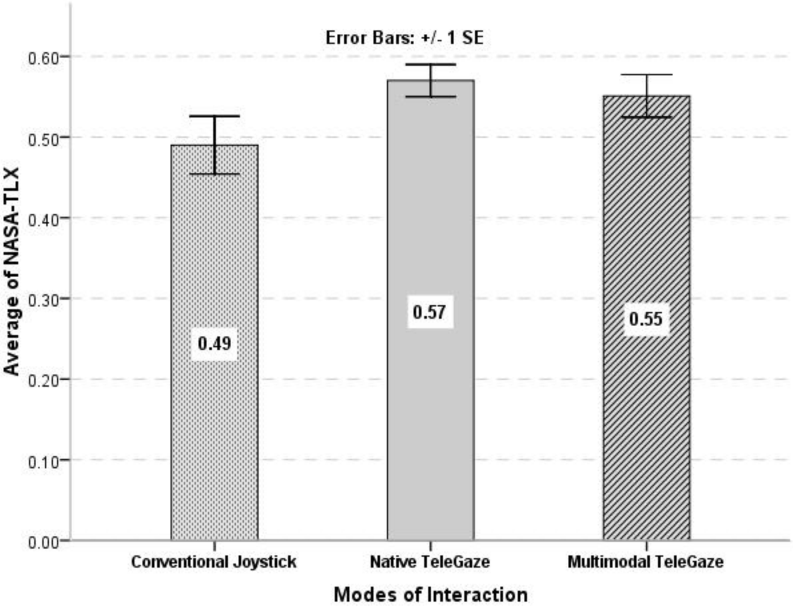


Figure 5.10: Average of NASA-TLX for all three modes of interactions.

From the results shown in the figure, it can be seen that the highest value of the index is obtained for the native TeleGaze (*M*= 0.57, *SE*= 0.050), while the lowest value is obtained for the joystick (*M*= 0.49, *SE*= 0.042). Consequently, the value of the index for the multimodal TeleGaze (*M*= 0.55, *SE*= 0.050) lies between the values of the other two modes of interactions. This shows that although the joystick created the least task load, the multimodal TeleGaze created less than the native TeleGaze. This exactly what was predicted prior to the experiment since the accelerator pedal is meant to minimize the task load on the subject. To statistically determine the significance of the difference in the task load, the following testing hypothesis is constructed:

H_1 : *NASA task load index is different among the different interaction modes.*

The null hypothesis H_0 then is that the task load is not different among the interaction modes. To test the hypothesis, a one-way repeated measures ANOVA[40] is used. The results of the test show that NASA-TLX is not affected by the interaction mode, $F(2,9)=1.468$, $p>0.05$. Therefore, although the joystick created less load on the subject than the two modes of TeleGaze, the difference in the load is not statistically significant. This means that, *as far as task load is concerned, TeleGaze meets its joystick target*.

NASA-TLX usually is analyzed based on the value of the overall index. However, looking at the individual components reveals some interesting insights into the participants' opinions of the system [88]. Therefore, although the difference in the overall index is not statistically significant, deeper look into the individual components of the index is worth while. The average (statistical mean) of the index for each individual component is plotted in Figure 5.11 for all three modes of interactions.

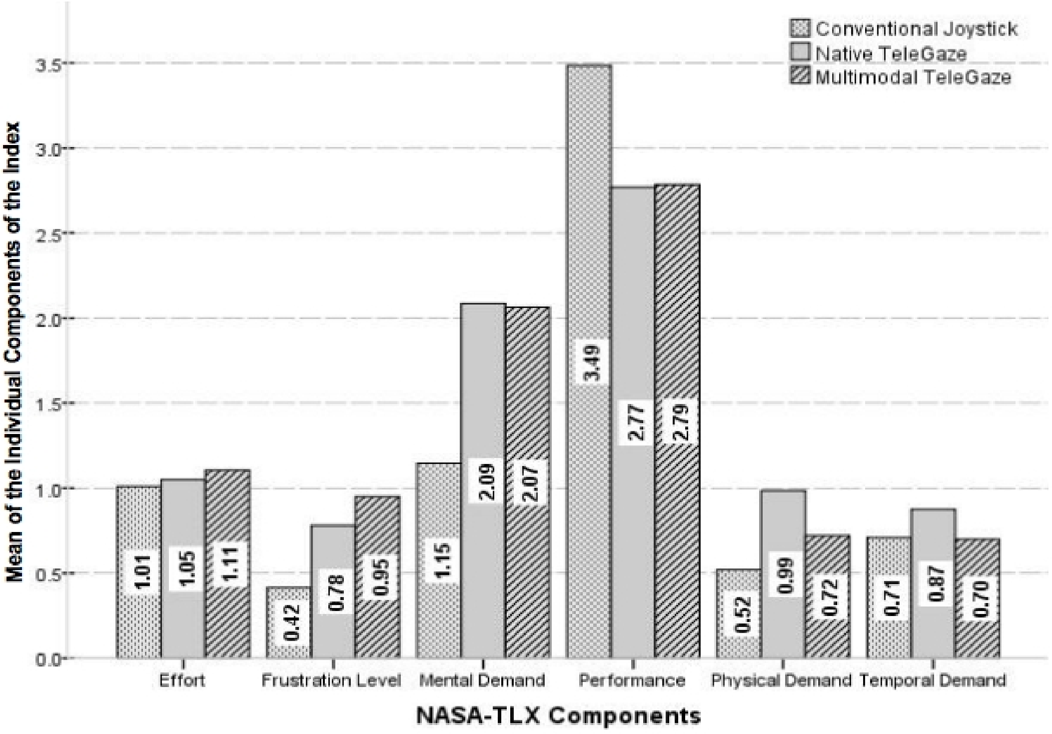


Figure 5.11: Average of each individual component of the NASA-TLX for all three modes of interaction.

---

[40] Although subjective metrics tend to be not normally distributed, the results of the Kolmogorov-Smirnov normality test show that NASA-TLX for the joystick, the native TeleGaze, and the multimodal TeleGaze is not significantly different from being normally distributed D(10)=0.200, 0.205, 0.183, p>0.05, respectively (p145, [86]). Therefore, the parametric test of ANOVA is used to test the hypothesis. Using Mauchly's test, the results of the test show that the assumption of sphericity is not violated $\chi^2(2)=4.417$, $p>0.05$ (p474, [86]).

The fact that NASA-TLX is a subjective metric makes it not as reliable as some of the objective metrics. As it was mentioned earlier, the meanings of the individual components and the ratings of the subjects might vary significantly from one person to another[41]. However, the mean of answers of a number of participants (ten in this experiment) increases the reliability of the metric. Therefore, from the results shown in Figure 5.11, the following findings can be highlighted:

● The minimum value of the index is obtained for *frustration* in the joystick mode, which means that the participants did not get frustrated by using the joystick to perform the task. The maximum value on the other hand, is obtained for *performance,* again in the joystick mode. This shows that the participants, on average, were more concerned about their performance when using the joystick and not when using the different modes of TeleGaze.

● Although no time limits were set for the participants, a noticeable proportion of the load is due to *temporal demand*. More interesting, the value of the component is higher for the native TeleGaze than both the joystick and the multimodal TeleGaze. This shows that the participants were more concerned with finishing the task in a specific time while using the native TeleGaze. However, this concern seems to have been less while using the other two modes of interactions.

● The values for both *performance* and the *mental demand* are very close for both modes of the TeleGaze. This shows that having a deterministic form of action confirmation, such as the accelerator pedal, does not affect the load that performance and/or mental demand put on a subject while performing a navigational task. On the other hand, the multimodal TeleGaze requires more *effort* and creates more *frustration* than the native TeleGaze.

● Another interesting finding lies in the results of the *physical demand*. The value for the physical demand for the native TeleGaze is higher than the value for both the joystick and the multimodal TeleGaze. This finding is interesting because both the joystick and the multimodal TeleGaze, apparently, involve physical activities more than

---

[41] The standard definitions of the components are included in Appendix A.

the native TeleGaze. However, this occurrence can be referred to the reason that while being tracked by the eye tracking equipment, a high level of tension is sensed in the shoulder muscles and the neck. Alternatively, and more interestingly, it can be due to a total misunderstanding of the component and its role in the overall task load.

Although some interesting findings can be obtained from the results, the fear that some of the components of the index might have been misunderstood, raises the issue of the reliability of the index. This issue was raised explicitly by some of the participants while filling the questionnaire. Another concern is that due to the fact that the ratings are collected after the experiment, they might not represent the actual cognitive load that was experienced during the experiment [93]. On the other hand, the fact that the overall findings are inline with the findings of the other metrics shows the reliability of this index. Therefore, the findings from the index are considered in the final decisions of the usability of TeleGaze.

5.7.3 Inspecting Methods

Inspecting methods reveal some insights into the subjects' experiences and reactions during the execution of the task. Tracking what the users have seen and how they have reacted to the scene is an important source of information that can be used for evaluation purposes [108]. By seeing what they have seen, inferences can be made about specific strategies that they have used to deal with demands of different situations [93]. For this reason, data is collected and combined from two different sources to create one form of inspecting method that is used as a qualitative evaluation metric. As it was covered earlier in chapter 4, eye tracking data is not suitable as an evaluation metric for TeleGaze in its traditional form. Therefore, the data is combined with video recordings of the robot's eye, which is also the subject's eye, for the whole duration of the operation. This is one of the most important forms of inspecting methods in evaluating gaze-driven interfaces as "*eye movements collected over a prototype of an interface may guide the designer on the interface's layout*" [26].

One of the most interesting findings revealed by this metric is related to the positioning of the forward action region. Looking at the results of this metric in the joystick mode, the participants mostly are gazing below the forward action region when trying to move forward. This is influenced by the proportion of the height of each individual and the height of the interaction screen. Regardless of the subject's height, the height of the interaction screen was kept constant for all the participants. Also the overall distribution of the gazing points on the screen turned out to be different between the participants regardless of their height. Therefore, it is realized that the height of the forward action region does not suit all the participants. This issue was raised clearly by two subjects when they mentioned that gazing at a higher level than their comfort line of sight is tiring and causing fatigue.

This significant finding is only revealed by studying the results in the joystick mode, because in this mode the gazing behaviour is not constrained by the interface. In the TeleGaze modes however, the gazing behaviour is constrained by the positioning of the action regions and the intended actions. Also since no specific questions regarding the positioning of the action regions individually is included in the questionnaire, this issue is not revealed by the results of the questionnaire. This finding shows the importance of the combined recordings of the scene and the gazing points on the interaction screen for all the interaction modes, particularly the joystick.

One of the other interesting findings revealed by studying the results of this evaluation metric is the default setting of the camera auto-home configuration. Despite the option of enabling/disabling this capability, almost no participant ever used this functionality. In contrast, the auto-home configuration, set to *enabled* by default, caused panic and frustration in some situations. Following is an example of a frequently observed situation where this capability caused extra time and task load:

- The subject stops moving the robot towards the painting because she believes that the robot is close enough to be able to inspect the contents, if the camera alignment is properly set.

- The subject performs a sequence of camera adjustment operations. However, when the camera is facing the painting, the subject realizes that the distance between the robot and the painting is more than what she needs to be able to inspect the contents.

● The subject then wants to move the robot few small steps closer to the painting while keeping the camera alignment the same. This is because she believes that the current camera alignment is the best to be able to inspect the painting.

● Once the first robot motion command is issued, all camera alignments are canceled and the camera is realigned with the robot. Consequently, all necessary camera alignments have to be repeated allover again. This happens because the camera auto-home configuration is *enabled* by default.

● The whole situation can be easily avoided be setting the auto-home configuration to *disabled* once necessary camera alignments are done. In this case, the camera alignments stays the same even when robot action commands are issued.

Similar situations occurred when participants wanted to avoid one of the obstacles by aligning the camera and moving the robot. These kinds of situations added to the overall time-to-complete the task and the overall task load of the experiment. This significant finding too is only revealed by analysing the results of this evaluation metric. Non of the other evaluation metrics highlighted this issue or the effect of the camera auto-home configuration settings.

5.8 Discussion

The obtained results are not consistent for all the evaluation metrics used to evaluate the TeleGaze against the joystick target. Efficiency results for example, show that the native TeleGaze is more efficient than the multimodal TeleGaze. In contrast, user satisfaction questionnaire and NASA-TLX results show that the multimodal TeleGaze is more desirable than the native TeleGaze. On the other hand, effectiveness results show that both modes of TeleGaze are the same in terms of overall goal achievement. This shows that different modes of TeleGaze have different advantages. It also shows the importance of using different types of evaluation metrics [74].

In order to get an overall idea of how the results of the evaluation place the interaction modes against each other, comparative analyses are needed. Based on average rankings (mean or median) of individual evaluation metrics, both modes of TeleGaze are relatively ranked against the joystick target. The overall ranking then is calculated based on the average of the rankings for the individual evaluation metrics, which is presented in Table 5.1:

Table 5.1 : Average (Mean or Median) rankings of the interaction modes relative to the joystick target.

	Conventional Joystick	**Native TeleGaze**	**Multimodal TeleGaze**
Efficiency (Time-to-Complete Task)[42]	358.4 (*MR*= 1.00)	377.9 (*MR*=0.95)	412.0 (*MR*=0.87)
Effectiveness (Overall Goal Achievement)	100.0 (*MR*= 1.00)	98.0 (*MR*= 0.98)	98.0 (*MR*= 0.98)
Satisfaction Questionnaire (1-5 Likert Scale)	4.20 (*MR*= 1.00)	3.25 (*MR*= 0.77)	4.00 (*MR*= 0.95)
Overall Task Load (NASA-TLX)	0.49 (*MR*= 1.00)	0.57 (*MR*= 0.86)	0.55 (*MR*= 0.89)
Relative Average Rankings	1	0.89	0.92
Final Rankings Against the Joystick	**First Place**	**Third Place**	**Second Place**

From the findings presented in the table, it can be seen that the joystick comes at the first place. The multimodal TeleGaze comes second with only 8% (1.0-0.92=0.08) less usability than the joystick, and the native TeleGaze comes last with 11% (1.0-0.89=0.11) less usability than the joystick. In this calculation, all four evaluation metrics are given the same weight in the shaping of the final usability index. Also no statistical significance is taken into consideration for the differences between the values of the individual metrics. Therefore, another comparative analysis is conducted based on statistical tests for the differences in individual measurements and their level of significance.

In this analysis, the significance of the difference between individual evaluation metrics is considered for the rankings. Also whether the results of the significancy test is obtained using parametric or non-parametric statistical tests is considered. In addition to these main factors, whether the metric is subjective or objective and whether it is more standardized metric, such as NASA-TLX, or less, such as the questionnaire, are also considered. Therefore, the usability index, *U*, is calculated according to the following formula:

---

[42] The ranking for two of the evaluation metrics (efficiency and NASA-TLX) is calculated differently than the other two of the evaluation metrics (effectiveness and questionnaire). This is because more time to complete task means less efficient and more NASA-TLX value means less usable. Therefore, the inverse of the actual readings are taken into account when calculating the ranks.

$$U = T_u - [0.3 * S_{efficiency} + 0.2 * S_{effectiveness} + 0.2 * S_{questionnaire} + 0.3 * S_{nasaTLX}] \qquad 5.1$$

Where T_u is the usability of the joystick target, which is set to 1 in this context. $S_{efficiency}$ is 1 if statistical analysis shows that efficiency is significantly different from efficiency of the joystick, and is 0 if not. $S_{effectiveness}$ is 1 if statistical analysis shows that effectiveness is significantly different from effectiveness of the joystick, and is 0 if not. $S_{questionnaire}$ is 1 if statistical analysis shows that questionnaire results are significantly different from questionnaire results of the joystick, and is 0 if not. $S_{nasaTLX}$ is 1 if statistical analysis shows that NASA-TLX results are significantly different from NASA-TLX results of the joystick, and is 0 if not.

Due to violations of normality distribution requirements, both effectiveness and the questionnaire are analysed using non-parametric tests. Whereas efficiency and NASA-TLX are analysed using parametric tests. In general, some people believe that parametric tests are more reliable than non-parametric tests in detecting significant differences between means [86]. Therefore, in the equation above, 30% of the weight is given to each of efficiency and NASA-TLX, while only 20% of the weight is given to each of effectiveness and the questionnaire. The used weighting technique ensures a desired balance between results obtained using parametric and non-parametric tests. Also it ensures a desired balance between objective and subjective metrics when the two objective metrics combined weigh 50% and the two subjective metrics combined weigh the other 50%. The significancy index S of any element in the equation is given a value of 1 if the difference between the TeleGaze mode and the joystick is statistically significant for that element, and zero otherwise.

Recalling some results from section 5.7 , the results of the one-way repeated measures ANOVA show that efficiency is not significantly affected by the interaction mode $F(2,8) = 0.701$, $p > 0.05$ ($S_{efficiency} = 0$ for both modes of TeleGaze). Using the Friedman's ANOVA, the results of the test show that effectiveness is not significantly affected by the interaction mode either $\chi^2(2) = 2.667$, $p > 0.05$ ($S_{effectiveness} = 0$ for both modes of TeleGaze). The results of the Friedman'a ANOVA conducted for the questionnaire show that user satisfaction is significantly affected by the mode of interaction $\chi^2(2)=6.381$, $p<0.05$. Further analysis using the Wilcoxon signed ranks test, the results of the test show that user satisfaction is significantly different between the joystick and the native TeleGaze $z=-2.121$, $p<0.05$, $r=0.47$ ($S_{questionnaire}=1$ for native TeleGaze). However, user satisfaction is not significantly affected by the interaction mode $z=-1.807$, $p>0.05$, $r=0.40$ when the test is conducted for the joystick and the multimodal TeleGaze ($S_{questionnaire}=0$ for multimodal TeleGaze). Finally, using Friedman's ANOVA the results of the test show that the difference in task load based on NASA-TLX is not statistically significant $\chi^2(2)=4.667$, $p>0.05$ ($S_{nasaTLX}=0$ for both modes of TeleGaze).

Substituting the values of S based on the results from the statistical tests, the following usability index values are obtained for TeleGaze:

$$U_{Native}=1.0-[0.3*0+0.2*0+0.2*1+0.3*0]=0.8 \tag{5.2}$$

$$U_{Multimodal}=1.0-[0.3*0+0.2*0+0.2*0+0.3*0]=1.0 \tag{5.3}$$

From the results of the equations above, it can be seen that the multimodal TeleGaze scores total usability in comparison with the joystick target, while the native TeleGaze does not. These results are consistent with the results obtained based on mean rankings of TeleGaze in comparison with the joystick. Taking mean rankings and statistical significancy tests into account, it can be concluded that *the multimodal TeleGaze meets the joystick target* from all considered usability points of views. On the other hand, *the native TeleGaze does not meet the target* since one element of the usability index (user satisfaction level) is significantly different from the joystick target.

The relative improvements in TeleGaze against the joystick target is a significant achievement when compared with the obtained results in the previous stage of the research. Looking at the evaluation metrics individually, better results are achieved as far as efficiency is concerned. Both modes of TeleGaze in this stage of the research are not significantly less efficient than the joystick target. Also, with the complexity of the navigational task of this stage of the research, achieving the same effectiveness as the joystick target is highly promising. There is no significant improvements in terms of effectiveness however, when compared with the previous stage of the research.

Efficiency and effectiveness evaluation metrics have not shown any advantages of the multimodal approach over the native approach. Instead, the efficiency of the native TeleGaze is higher than the multimodal TeleGaze. However, the subjective metrics show significant advantages of the multimodal approach. Higher user satisfaction and lower task loads are obtained for the multimodal TeleGaze than the native TeleGaze. Also the same user satisfaction level is achieved with the multimodal TeleGaze, but not the native TeleGaze. These findings are not comparable with the findings of the previous stage of the research since similar measurements were not taken then. However, the obtaining the same usability index as the joystick for the multimodal TeleGaze is a significant step forward.

Although the findings from the testing and inquiring methods show that TeleGaze has met its target, the findings from the inspecting methods show some room for some interesting improvements. The positioning of the forward action region for example is an easy, as well as interesting, improvement that is likely to increase the chances of TeleGaze to beat the joystick target instead of only meeting it. Following an adaptable user interface approach that allows the user to control it [28], a relocatable forward action region might be the solution.

The observations from the inspecting methods reveal that the height of the forward action region does not suit all the participants. Instead, some participants preferred the action region to be lowered while some others did not have any concerns with its height. Therefore, as one of the further improvements on the design of the TeleGaze interface, a relocatable forward action region is necessary. The relocatable forward action region can be adjusted prior to commencing any interaction session to best fit the height of individual subjects.

In addition to a relocatable forward action regions, the camera auto-home configuration default settings can be altered. As it was covered earlier, setting this capability to enabled by default created some panic and costed extra time in certain situations. Therefore, an alternative approach is necessary to be experimented. In addition to these two main improvements, the size of the action regions can be improved also. Most of the participants commented that one of the main reasons for preferring the multimodal TeleGaze is the fact that they can have bigger action regions. Therefore, the size of the action regions is another element of improvement that can be easily addressed.

5.9 Conclusions

Throughout this stage of the research TeleGaze has seen significant improvements in terms of design, functionality, and usability. The redesigned interface and the upgraded robotic platform played an important role in achieving these improvements. Also the extensive set of evaluation metrics and the redesigned usability testing experiment played their role in proofing these improvements.

Based on statistical analysis of the obtained measurements for all the evaluation metrics, *the multimodal TeleGaze meets its joystick target in terms of overall usability*. The native TeleGaze however, does not meet the target since it has not achieved the same user satisfaction level as the joystick. Therefore, the multimodal TeleGaze is the answer to the research question. With the multimodal TeleGaze, *mobile robot teleoperation through eye gaze is possible, with both hands totally free from the task*. Furthermore, *it is as possible as mobile robot teleoperation with conventional teleoperation means*.

With the research question being totally answered, the end of this stage of the work marks the end of the research. However, some interesting findings from the inspecting methods show potentials for further improvements. Improvements that might push TeleGaze to beat the joystick target and not just meeting it. Therefore, it was decided to carry on the research in order to conduct some modifications in the design of the interface. Also conducting the necessary usability testing experiment to evaluate these modifications. Based on observations from the inspecting methods, following are the key modifications that are going to be carried out.

● One of the main modifications in the design of the TeleGaze interface is introducing a relocatable forward action region. This is necessary to adapt the interface to the height of the subject. The height of the relocatable action region can be adjusted during the calibration process to best suit the subject's need.

● Another modification is changing the default settings of the camera auto-home configuration. Findings have shown that setting this capability to enabled by default resulted in some extra time and load in completing the task. Therefore, in the next version of the interface this capability is going to be set to disabled by default. The operator still will have the option to enable/disable the capability, similar to the current version of the interface.

● One other modification in the interface is using bigger action regions. With the multimodal TeleGaze, the fear of issuing a command unintentionally is minimized. Therefore, the action regions can be enlarged to ease the task of gazing and keeping the gaze in the region for the desired duration. As it has been mentioned earlier, this was explicitly requested by some of the participants.

● The last improvement is reducing the granularity of the steering action regions. It has been observed that the granulated steering action regions has not been used frequently. Therefore, less granulated steering action regions, which adds to the clarity of the action regions, can be used.

To carry out these modifications, a refined version of the TeleGaze interface is designed. Also to evaluate the effects of these modifications on the usability of TeleGaze, the usability testing experiment is repeated for the refined interface. Therefore, the next chapter covers the details of these modifications and the results of the usability testing experiment.

Chapter SIX

Refined Multimodal TeleGaze

6.1 Introduction

The results of the last usability testing experiment showed that the multimodal TeleGaze performs well in comparison with the joystick target. This is based on the fact that within the bounds of statistical significance it met the target in a straightforward experimental comparison of usability [73]. In addition to the refinements in the design of the interface from the earlier stage, the accelerator pedal has played a significant role in achieving the recorded usability scores. The important role of the pedal lies in making the interface act on the inputs from the user's eye only when she wants it, which is a difficult aim to achieve in general [44]. Hence, the multimodal approach has minimized the level of ambiguity in the user's input and enriched the output.

Although there are several dimensions along which gaze-driven communication can be viewed, the usability is the most important of these dimensions [16]. TeleGaze has met its joystick target as far as usability is concerned. However, the data from the usability testing experiment showed some potentials for further improvements of the interface. In particular, the results from the video replay of the scene combined with the eye tracking data is the inspiration behind this stage of the research. Although the most relevant metrics related to eye tracking data vary from task to task [40], this sort of eye tracking metric has been used as tools for further improvements [109].

Therefore, this chapter covers some final refinements in the design of the TeleGaze interface. These refinements are mainly inspired by the results of the video recordings of the last usability testing experiment, but also supported by supervisory observations of the experiment. To evaluate the impact of these refinements, the usability testing experiment has been repeated for the refined multimodal interface (RmI). The results of the usability testing experiment are then statistically analysed and compared with the results of the previous usability testing experiment and final conclusions are drawn.

6.2 Interface Refinements

Analysing and visualizing multimodal data on user interaction is a difficult task because video recordings often lack specific details despite the richness of the data [94], [104]. However, some key design points have been highlighted by the video recordings obtained for the last usability testing experiment which are also inline with some personal observations. Based on the findings from both sources, followings are the key refinements that are believed to be necessary to further improve the interface:

6.2.1 Relocatable Action Region

One of the objectives of TeleGaze is to design an interactive system that interprets the natural gazing behaviours of a human operator into teleoperation commands through an intuitive interface. Therefore, the positioning of the action regions are meant to be intuitive and inline with the natural gazing behaviours of human operators. However, the information revealed by the video recordings show otherwise for one of the, supposedly most intuitive, action regions. The forward action region seemed to be positioned higher than the natural line of sight of most of the participants. This information was revealed when the density of the fixations on the interface was observed closely while using the joystick. The observations showed that the density of the fixations naturally lies below the position of the forward action region when the fixations are not driven by the action region, such as in the joystick mode.

During the interaction sessions, the height of the user's seat was adjusted so that the line of sight of the subject is level with the centre of the interaction screen, which is also the centre of the interface. Consequently the forward action region is located at a higher level than the line of sight of the subject. This is shown in Figure 6.1. This is the most likely reason that the density of fixations, in general, is located below the forward action region. However, since the observations vary for the participants, this might not be the definite reason. Therefore, lowering the forward action region to the centre of the interface, which will be level with the line of sight of the subject, does not solve the problem for all the participants. Instead, a rather dynamic solution is required, such as a *relocatable forward action region.* With the relocatable forward action region, the height of the action region is adjusted based on each individuals needs and comfort, which makes the interface an adaptable user interface[43] [28].

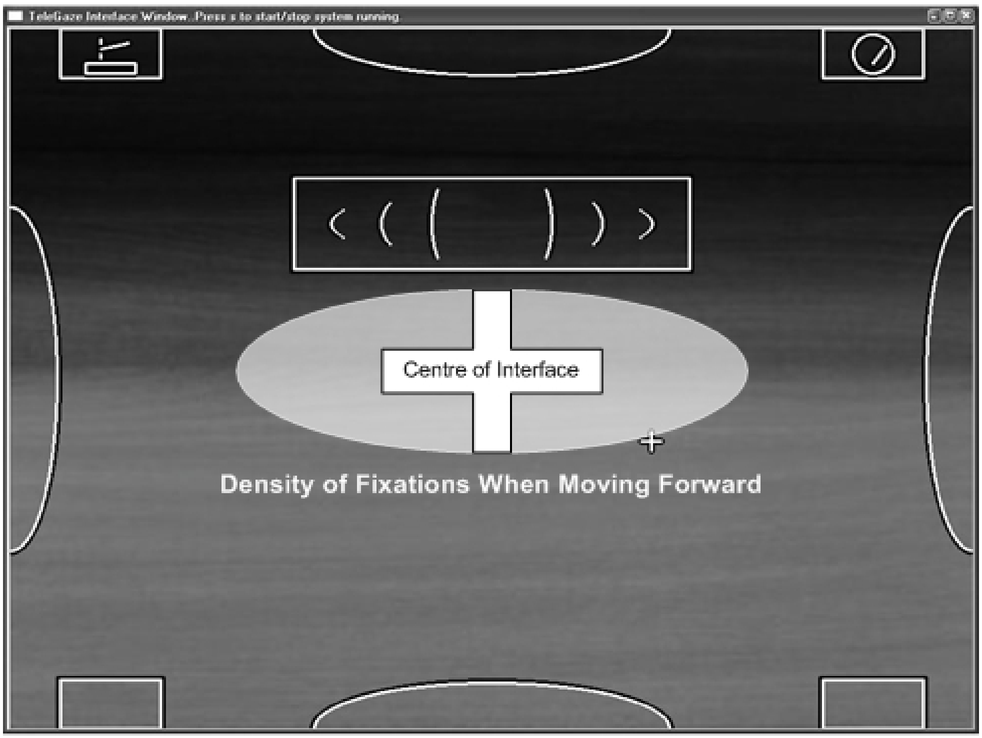


Figure 6.1: The position of the forward action region in accordance to the line of sight, or the centre of the interface.

In this design, the height of the forward action region can be adjusted to the best comfort of the subject, within the available space on the interface. Prior to commencing the actual interaction session, participants have the choice to alter the height of the forward action region. However, once the most comfortable height is selected, then this

---

[43] Adaptable user interfaces allow the user to control some of its features, while adaptive user interface adapts those features automatically to the user [28].

option is disabled and participants does not have the option of altering it during the actual interaction session. This is to avoid creating extra task load of adjusting the height of the forward action region while being engaged in the actual experiment. In this design, adjusting the height of the forward action region is done by the supervisor using the keyboard, and not the user herself. The span of the relocatable forward action region is illustrated in Figure 6.2.


TeleGaze Interface Window..Press s to start/stop system running.

Figure 6.2: The span of the relocatable forward action region on the refined multimodal interface.

6.2.2 Camera Auto-Home Functionality

The camera auto-home functionality ensures that the camera is aligned with the robot whenever robotic actions are executed. This is believed to be necessary in order to avoid moving the robot in one direction while looking at another direction. However, the operator has the capability of enabling/disabling this functionality for pan and tilt, separately. This means that this functionality can be disabled when it is not necessary, or more importantly, when the functionality does not fit the nature of the task. An example of such situations is that when the robot has to move along a wall but keep looking at the wall. In this case, enabling the functionality causes all necessary camera alignments to be repeated after any robotic movements. To match the level of control between the

TeleGaze and the joystick, this functionality is also available for the joystick mode. The camera alignment information is also presented to the user on the joystick interface with indications of whether the functionality is enabled or disabled for pan and tilt separately.

The video recordings of the scene, combined with the eye tracking data, revealed interesting information regarding the camera auto-home functionality. Despite introducing the enable/disable capability of this functionality to the participants, no participants disabled this functionality during the experiment. Interestingly enough, even with the joystick no participants ever disabled this functionality. Not disabling this functionality throughout the navigational task, caused the participants to face some unpredicted situations. In some cases, these situations caused some panic and added to the overall time-to-complete task and the overall task load. Therefore, in this design of the interface, this functionality has been modified to best suit the task and the participants natural gazing behaviours.

The underlying principle of the auto-home camera functionality stays the same. The camera is aligned with the robot whenever robotic actions are executed. However, instead of realigning the camera with the current headings of the robot, the robot is aligned with the current headings of the camera. This ensures that the camera and the robot are aligned when the robot is moving. It also ensures that the robot moves to the direction of the camera, which eliminates the need for camera realignment after the robot has moved. In addition to this change, the default configuration in this design is that the functionality is disabled, unlike the default configuration of the previous design. The participants are expected to enable the functionality if they need to. However, even with the functionality being disabled, the interaction experiment is expected to run smoother than the previous stage.

6.2.3 Less Granulated Steering

One of the objectives of the evaluation metrics mentioned in chapter 4 is to show the usage of each one of the components of the interface individually. Removing ineffective components can benefit both the design and the user [94]. It benefits the design in the sense that it makes it less cluttered, and it benefits the user in the sense that it creates less interaction objects to deal with. One of the major additions to the multimodal interface, designed at the previous stage of the research, was the granulated steering action regions. The granulated steering action regions allow linear control over the steering values of the robot by presenting the user with different proportions of both the linear and the angular velocities. This is achieved by adding two extra regions to each side of the forward action region before the turn right/left action regions. Therefore, the forward/steering action region is composed of seven action regions in total[44], as illustrated in Figure 5.4.

Studying the video recordings of the scene combined with the eye tracking data revealed that the in-between steering action regions are significantly less used than anticipated. The users mostly either moved forward or turned right/left without depending much on the granulated steering action regions. Therefore, to reduce the number of the action regions cramped in the middle of the interface, the number of the granulated steering action regions is halved. Instead of two in-between action regions on each side of the forward action region, only one granulated steering action region is used in the refined multimodal interface. The components of the modified forward/steering action region with relative values of each portion of the action region are illustrated in Figure 6.3.

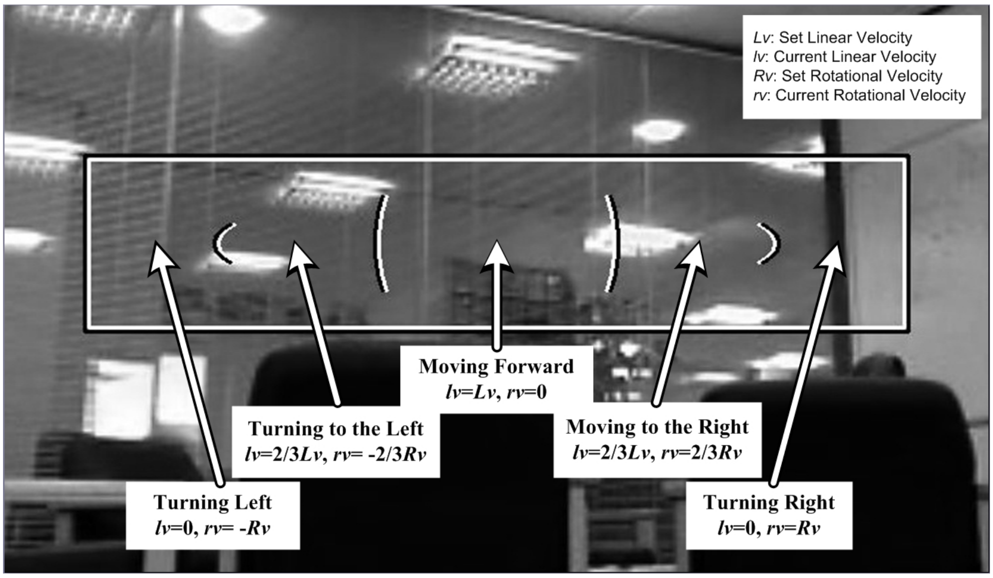


Figure 6.3: The design of the forward/steering action region in the refined multimodal interface.

---

[44] Although this action region behaves as a continuous function, in order to gain better understandings of the density of the fixations and the use of the action regions, a discrete function governs the action region.

Another significant advantage of this modification is that the available space now is used for less components, which means that the remaining components can use more space. The available space between the forward action region and turning right/left action regions is now used for only one action region instead of two. Therefore, the granulated action region has more space and hence has a larger size, which makes it more comfortable to use.

6.2.4 Size of the Action Regions

Gaze may act as a faster pointing device than a mouse if targets are sufficiently large [7]. Therefore, one of the main concerns in designing any gaze-driven interfaces is the size of the individual components of the interface. In particular, the components that replace buttons on conventional graphical user interfaces [73]. With the current status of eye tracking equipments, fine pointing on high resolution displays is not possible, which "*restricts the size of the displayed objects that can be selected*" [9].

Furthermore, the accuracy problem is not only due to the resolution of the eye tracking camera, but also due to the jittery nature of the eye movements [11]. As a result, the eye tracker is viewed as having a much coarser resolution than that of a typical input device, "*perhaps more like a touch screen*" [44]. Therefore, larger action regions on the TeleGaze interface is always desirable, if it does not contradict with the other needs of the interface, such as space for inspection or resting the eyes.

The existence of the accelerator pedal and the transparency of the action regions make it possible to enlarge them even more. Larger action regions make the interface easier to use. This issue was explicitly pointed out by few participants during the experiment. Therefore, in the refined multimodal interface, the size of the action regions are enlarged by a factor of 1.5, which makes the smaller dimension of the action regions equal to 2.25° degrees visual angle. This is equivalent to nearly 4.5cm on the interface when the subject is seated at a distance of 65cm from the interaction screen. Figure 6.4 shows a snapshot of the refined multimodal interface with larger action regions than the multimodal interface.

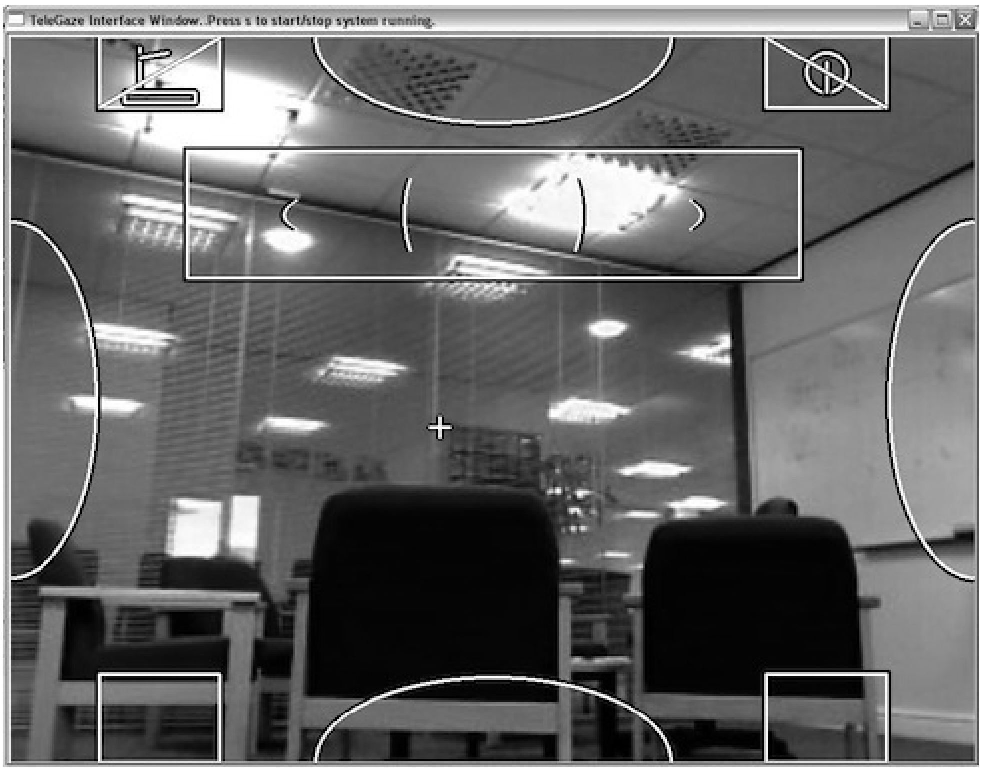


Figure 6.4: An actual snapshot of the refined multimodal interface.

6.3 Usability Testing Experiment

The elements of the usability testing experiment and the evaluation metrics used at this stage of the research are the same as the ones used at the previous stage of the research (Chapter 5). However, since the experiment is repeated only one time for the refined multimodal interface, some details are necessary to be covered. The details of the experiment and the used set of questionnaire are explained in the following sections.

6.3.1 Sequence of the Paintings

In the previous usability testing experiment two different orderings of the paintings were used. The participants either started with painting 1, 2 and then 3, or they started with painting 3, 2 and then 1. This was not because of any differences in the contents of the paintings, but because the operator's geographic knowledge affects her gazing behaviours [110]. Research shows that familiarity with the road affects the sequence and numbers of fixations [111], [112]. Therefore, different orderings introduced in order to ensure equal familiarity with the environment for the different modes of interactions.

Due to the fact that a total of three interaction modes were experimented by each individual, one of these orderings was performed twice while the other one was performed only once. Since the same participants participate in the usability testing experiment of this stage, the orderings can be counter balanced with this mode of interaction. With four modes of interactions, each one of the orderings can be performed twice. Therefore, in this usability testing experiment, the ordering that the participant perform is the one that they performed only once in the previous experiment.

6.3.2 Questionnaire

As it has been mentioned earlier in chapter, the questionnaire is composed of two sets of questions, where one is filled out after each interaction mode and the other is filled out at the end of the experiment, including all three interaction modes. In this experiment however, there is only one interaction mode. Therefore, the whole questionnaire is filled out at the end of the experiment.

The results of the first set of questions, which inquires the interaction experience, can then be compared with the results obtained for the other three modes of interactions. The results of the second set of questions, which inquires the design of the interface, can be compared with the results obtained at the end of the previous usability testing experiment. This allows obtaining comparative results for both the interaction experience and the design of the interface between the previous stage of the research and this stage. A copy of the whole questionnaire used in this experiment is included in Appendix A (Section A.5).

6.3.3 The Numbers on the Paintings

Three different sets of numbers exist on each painting for the participants to inspect. In the previous experiment, the participants were asked to report the summation of the numbers of one of the sets in each interaction mode. Hence, a total of three different sets of numbers are provided for three different modes of interactions. In this experiment however, a different approach is required since only one mode of interaction is experimented. To minimize any learning or boredom effects, the participants are asked to use each set of numbers only for one painting. A different set is then used for the second painting and the last set is used for the third painting. This is believed to be necessary in order to achieve comparable, as well as even, results with the results obtained in the previous experiment.

Except the changes mentioned above, no other changes are made in the parameters of the usability testing experiment. The participants are allowed the same amount of practising prior to the actual experiment as in the previous experiment. The fact that participants needed practising prior to the actual experiment shows that the effect of learning is not significant enough to influence the results of the experiment. The participants are informed of the changes in the details of the experiment. However, they are not informed of the changes in the design of the interface. This is in order to see whether the changes in the design are significant enough to be noticed by the participants or not.

6.4 Results and Findings

As it was mentioned earlier, in order to obtain comparable results, the same set of measurements are taken in this experiment as of the previous experiment. For ease of comparison, the same structure of presenting the results and the findings is used in this chapter as of the previous chapter. The results from this experiment are presented with the results from the previous experiment. However, the results for the joystick interaction mode and the refined multimodal TeleGaze is the main focus of discussion and analysis, similar to some related works in the field [93]. Following are the obtained results for each evaluation metric and the findings:

6.4.1 Testing Methods

a. Efficiency

In order to measure efficiency, the average of time-to-complete the task using the refined multimodal interface is calculated. The results are compared and statistically analysed against the average time-to-complete the task for the other three modes of interactions. To better visualize the findings, the calculated averages for all four modes of interactions are plotted in Figure 6.5.

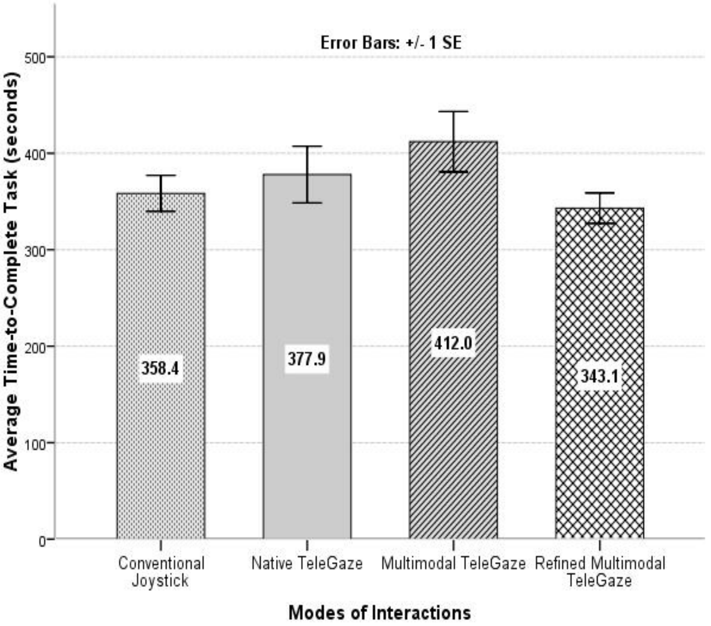


Figure 6.5: Average time-to-complete the task in all four modes of interactions.

From the results shown in the above figure, it can be observed that the refined multimodal interface is not only quicker than the other two modes of TeleGaze, but also it is quicker than the joystick target. It takes the refined multimodal TeleGaze (*M*=343.1, *SE*=26.2) 20.1% less time-to-complete the task than the multimodal TeleGaze (*M*=412.0, *SE*=43.2). It also takes it 10.1% less time than the native TeleGaze (*M*=377.9, *SE*=41.7). More importantly, it takes the refined multimodal TeleGaze 4.5% less time-to-complete the task than the conventional joystick (*M*=358.4, *SE*=26.2), which is interesting to note at this point.

Based on mean ranking, the refined multimodal TeleGaze is the most efficient mode amongst the four modes of interaction. In particular, there is a noticeable improvement in the efficiency when compared with the efficiency of the multimodal interface in the previous experiment. However, to determine the significance of this improvement statistical analysis are required. Therefore, the following testing hypothesis is constructed[45]:

H_1 *: Time to complete task is different among the different interaction modes.*

---

[45] This is the same testing hypothesis used in the previous stages of the research since the aim of the research is the same for all the different stages of the research.

The null hypothesis H_0 then, is that interaction mode does not affect the average time-to-complete task and it does take all four modes of interactions the same time to complete the task. To test the hypothesis, a one-way repeated measures ANOVA is used after checking the data for normality[46]. The results of the test show that time-to-complete task is not significantly affected by the interaction mode $F(3,8)=1.088$, $p>0.05$. Therefore, despite the fact that the refined multimodal TeleGaze is more efficient than the joystick, the difference is not statistically significant.

b. Effectiveness

Using expression 4.1[47], the effectiveness of the refined multimodal interface is calculated for each participant. The average of the calculated effectiveness is then statistically analysed against the scored effectiveness of the other three modes of interactions. To visualize the calculated effectiveness of all four modes of interactions, the average of the calculated effectiveness is plotted in Figure 6.6.

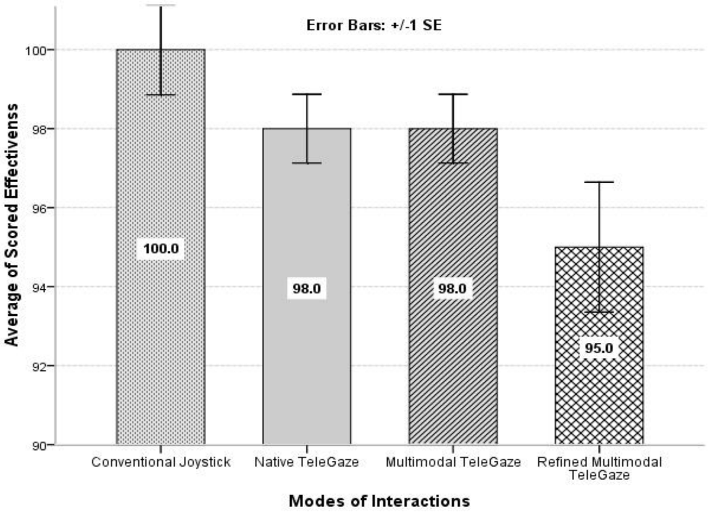


Figure 6.6: Average of effectiveness for all four modes of interactions.

From the above figure, it can be seen that on average, the effectiveness of the refined multimodal TeleGaze is less than all the other modes of interactions. The refined multimodal TeleGaze is 3.0% less effective than the other two modes of TeleGaze and

---

[46] Using Kolmogorov-Smirnov normality test, the results of testing for normality show that time-to-complete task for the refined multimodal TeleGaze is not significantly different from being normally distributed D(10)=0.2, p>0.05 (p145, [86]). The data for the joystick, the native TeleGaze, and the multimodal TeleGaze are not significantly different from being normally distributed either, as reported in the previous chapter (Chapter 5). Using Mauchly's test, the results of the test show that the assumption of sphericity is not violated $\chi^2(5)=7.793$, $p>0.05$ (p474, [86]).

[47] The effectiveness equation derived as $e=[(0.2\times c_{rds})+0.4-(0.1\times h_{obs})]\times 100$, which is presented in chapter 4 in detail.

5.0% less effective than the joystick. In order to determine the significance of the difference in the scored effectiveness, statistical tests are conducted. Using the Friedman's ANOVA[48], the results of the test show that effectiveness is not significantly affected by the interaction mode $\chi^2(3)=6.125$, $p>0.05$. Therefore, despite the fact that the refined multimodal TeleGaze scored less effectiveness than the joystick target, the difference is not statistically significant and can be neglected.

6.4.2 Inquiring Methods

a. Questionnaire

For analysis purposes, similar to the approach used in the previous experiment, the two sets of questions in the questionnaire are analysed separately. The results for the first set of questions, which inquires the interaction experience, is compared for all four modes of interactions. To visualize the results, the calculated average of the rankings for this set of questions is presented in Figure 6.7.

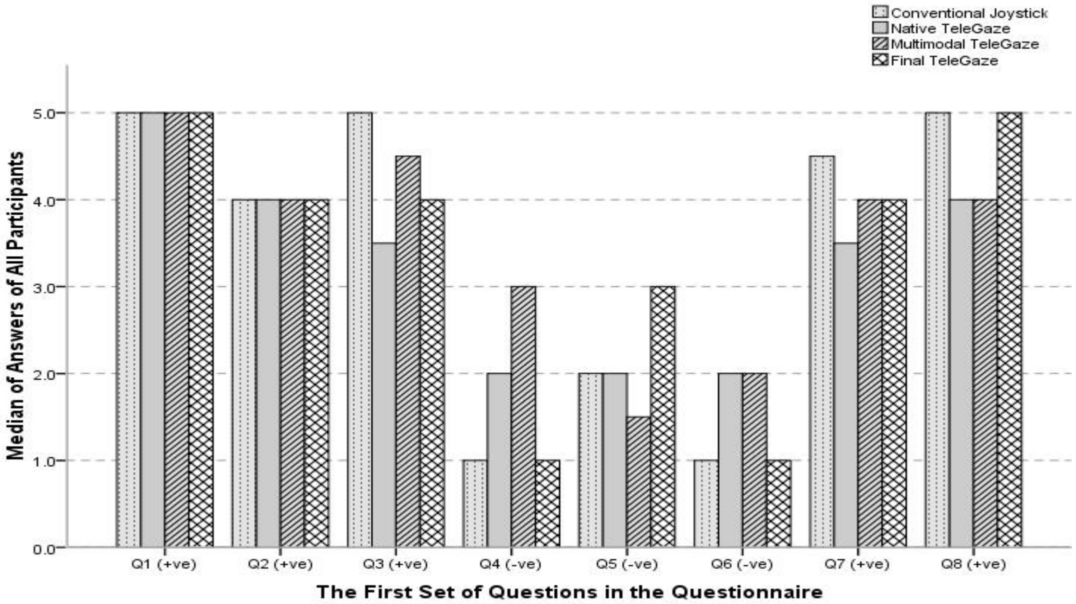


Figure 6.7: Average of the rankings for the questions in the first set of questions.

Based on the average rankings for each statement in the first set of questions in the questionnaire, the following can be inferred:

---

[48] Using Kolmogorov-Smirnov normality test, the results of testing for normality show that effectiveness of the refined multimodal TeleGaze is significantly different from being normal D(10)= 0.422, p<0.05 (p145, [86]). The results of the test for the native TeleGaze and the multimodal TeleGaze show the same (Chapter 5). Therefore, non-parametric tests are adopted to determine the significancy of the differences.

● No improvements are scored for the first two questions (Q1 and Q2). There is no room for improvements as far as Q1 is concerned, because the rankings for all three interaction modes are the maximum that can be achieved (5 out of 5). The average rankings for this question as far as the refined multimodal TeleGaze is concerned is the same. The average rankings for Q2, which inquires the capabilities that the subject expect the system to have, is the same for all four modes of interactions. This means that with all four modes of interactions, implicitly there are other capabilities that the subjects expect the systems to have.

● For Q3, the refined multimodal TeleGaze has scored higher rankings than the native TeleGaze, while it has scored lower rankings than the multimodal TeleGaze. This question inquires the level of confidence the subjects feel during the interaction. Surprisingly, these results were unpredicted prior to the experiment since it was thought that bigger action regions on this interface adds to the ease of use of the interface, and hence to the level of confidence.

● Improvements in the rankings for two out of three of the negative questions (Q4, Q5, and Q7)[49] for the refined multimodal TeleGaze can be seen against the other two modes of TeleGaze. Interestingly the results for Q5, which asks if the participants can perform any better that how they did, show higher agreement in the refined multimodal TeleGaze than the other two modes of TeleGaze. A potential reason for this might be hidden in the efficiency results for the refined multimodal TeleGaze. It might be that the participants believe that the achieved efficiency is the maximum that it can be achieved with the system. Therefore, they believe that there is no room for performing better.

● The last question (Q8) inquires the overall satisfaction level of the participant with the system. On average, the results show improvements in the refined multimodal TeleGaze compared to the other two modes of TeleGaze. Maximum ranking (5 out of 5) is scored for the refined multimodal TeleGaze, which is equal to the ranking scored for the conventional joystick.

---

[49] There statements are expressed in a negative way. Therefore, higher agreements are more negative than lower agreements, unlike the rest of the questions.

- On average, the rankings for the questions put the joystick (M=4.25) at the first place and the native TeleGaze (M=3.25) at the last place. While it puts both versions of the multimodal TeleGaze (M=4.0) between the joystick and the native TeleGaze, with a difference of only 0.25 from the joystick target on the scale of rankings (1 to 5).

To better determine the significancy of the differences in the rankings for each statement individually, the following testing hypothesis is constructed:

H_1 *: User satisfaction level is different among the different interaction modes.*

The null hypothesis then H_0 is that user satisfaction is not different for the different interaction modes. The results from a Friedman'a ANOVA[50] test show that user satisfaction is significantly affected by the mode of interaction $\chi^2(3)=7.868$, $p<0.05$. With significant differences in the user satisfaction between the four interaction modes, follow-up test are conducted for pairwise comparisons.

The Wilcoxon signed ranks test[51] is used to test the significance of the difference in user satisfaction between the refined multimodal TeleGaze and the other three modes of interaction. The results of the test show that user satisfaction is not significantly different between the joystick and the refined multimodal TeleGaze, z=-1.633, p>0.05. The results of the test also show the same for the native and the refined multimodal TeleGaze, z=-1.298, p>0.05, and the same for the multimodal and the refined multimodal TeleGaze z=-0.677, p>0.05. Therefore, despite being ranked 0.25 (on a 1 to 5 scale) less than the joystick, *the results of the test show that the refined multimodal TeleGaze meets the joystick target*, as far as user satisfaction is concerned.

As it was mentioned earlier in chapter 4, the first question in the second set of questions inquired the balance between both monitoring and controlling on the interface. The average of the answers to this question is presented in Figure 6.8.

---

[50] Friedman'a ANOVA is used because answers to questionnaire are, arguably, considered ordinal values and parametric tests are not recommended (p8, [86]).

[51] As mentioned in the previous chapter (Chapter 5), the Wilcoxon signed ranks test is used as the non-parametric equivalent to the student t-test (p552, [86]).

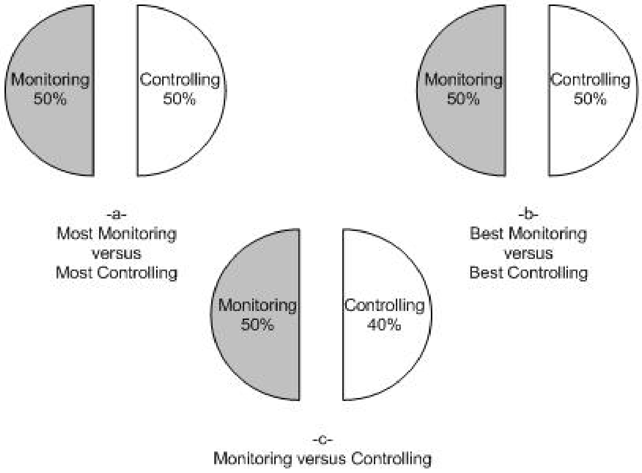


Figure 6.8: Average of the ratings for both parts of the Monitoring versus Controlling question.

As it can be seen from the figure, the ratings to both parts of the questions are interestingly ideal. Achieving the same level of both monitoring and controlling with the design of the interface is the ultimate balance that the interface can achieve. Comparing the results for this version of the interface with the results for the previous version in chapter 5, improvements can be seen as far as the quality of controlling is concerned. Since there are no differences in either the quantity or the quality of both monitoring and controlling, no statistical analysis is required.

In addition to the balance of quantity and quality of both monitoring and controlling, the second set of questions addresses other design elements of the interface. Examples of these elements are the appropriateness of the size of the action regions, which has been modified in the current version of the interface. Therefore, user ratings for this set of questions have been collected again to compare the design elements of the current version of the interface and the previous one. To visualize the differences in user rankings for these design elements, the average of the rankings for both versions of the interface is presented in Figure 6.9.

From the results presented in the figure, it can be inferred that except for the second feature, the current interface has scored either equal or higher rankings than the previous one. Interestingly, the second feature, which inquires the positioning of the camera related action regions, is the same in both versions of the interface. The fact that it has been ranked lower this time is most probably due to inconsistency in the participants answers, which is a common problem in subjective metrics. To measure the significancy of the scored improvements, statistical tests are conducted. Using the Wilcoxon signed rank test[52], the results of the test show that user satisfaction, as far as the design elements of the interface are concerned, has not significantly improved in the refined multimodal TeleGaze, z=-0.756, p>0.05. However, in general better rankings have been scored for the current version when compared with the previous version.

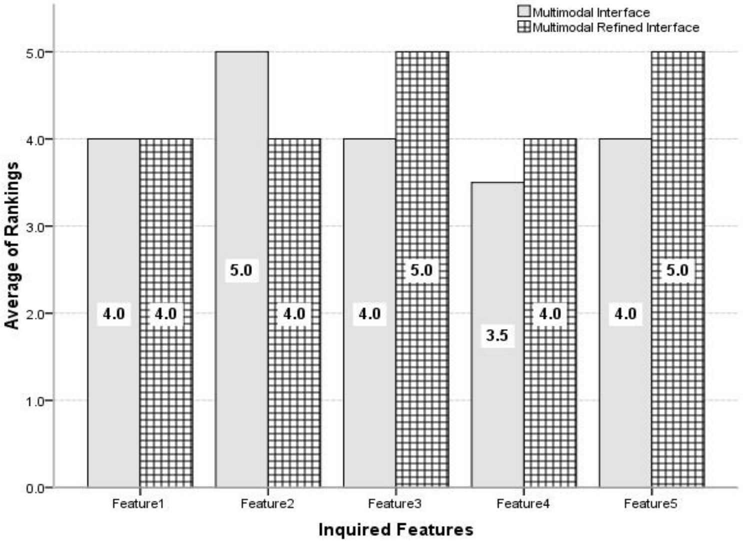


Figure 6.9: Average rankings of the design elements of the interface for both versions.

To determine the level of user satisfaction with the relocatable forward action region, a new statement was added to this set of questions. The statement inquired the participants' rankings for having the option of relocating the forward action region in terms of height. On average, the participants ranked this feature 5 out of 5, which shows that they highly appreciated the option of relocating the forward action region.

b. NASA-TLX

NASA-TLX is used as a subjective metric to measure the task workload during the execution of the task. Although it might be criticized for measuring the load after the task and not actually during the task [93], it is a well known tool for measuring task loads. Therefore, the measurements for NASA-TLX are taken for the refined multimodal TeleGaze. They are compared then, with the measurements obtained for the other three modes of interactions in the previous usability testing experiment. To

---

[52] The same reason explained in footnote 43.

visualize the results, the average of the task workload for all four modes of interactions are presented in Figure 6.10.

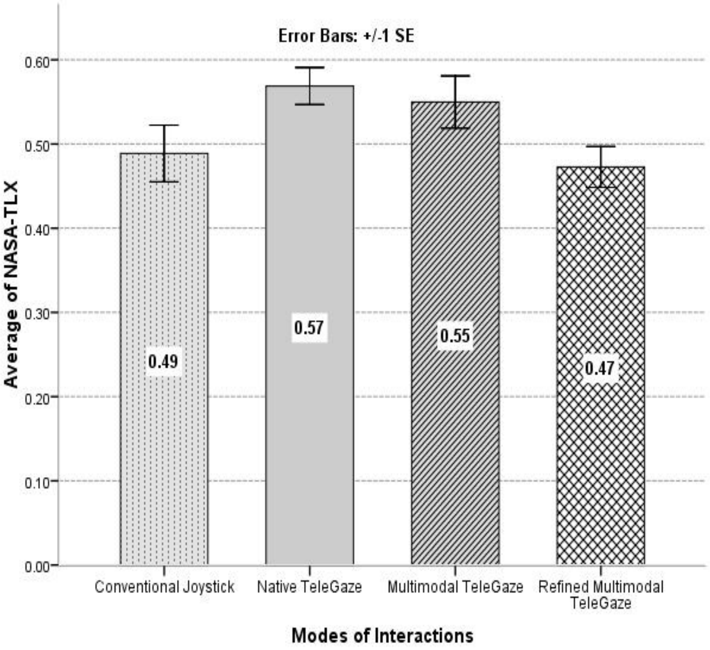


Figure 6.10: Average of NASA-TLX for all four modes of interactions.

It can be seen from the results shown in Figure 6.10 that the refined multimodal TeleGaze not only created task load less than the other two modes of TeleGaze, but also less than the joystick target. Based on mean rankings, the refined multimodal TeleGaze (*M*=0.47, *SE*=0.041) created the least task load amongst the four modes of interactions. On average, it created 2% less workload than the joystick, 8% less workload than the multimodal TeleGaze, and 10% less workload than the native TeleGaze. This shows improvements in the workload that the interaction system creates on the subject for the experimented navigational task. It shows that *the refined multimodal TeleGaze not only meets the joystick target, but it also beats the target,* as far as the task workload is concerned. However, to calculate the significancy of the differences in the workload, the same testing hypothesis presented in chapter 5 is tested, which is as follows:

H_1 *: NASA task load index is different for the different interaction modes.*

The null hypothesis H_0 then is that the task load is not different for the interaction modes. A one-way repeated measures ANOVA[53] is used to test the hypothesis. The

---

[53] As it was mentioned in the previous chapter, subjective metrics tend to be not normally distributed. However, the results of the

results of the test show that NASA-TLX is not affected by the interaction mode, $F(3,9)=2.044$, $p>0.05$. Therefore, although the refined multimodal TeleGaze created less load on the subject than the other three modes of interactions, the difference in the workload is not statistically significant.

Although NASA-TLX is normally treated as an overall value, looking into the values of each component might reveal interesting results. Therefore, similar to the approach used in chapter 5, the average of each component for all four modes of interactions are presented in Figure 6.11.

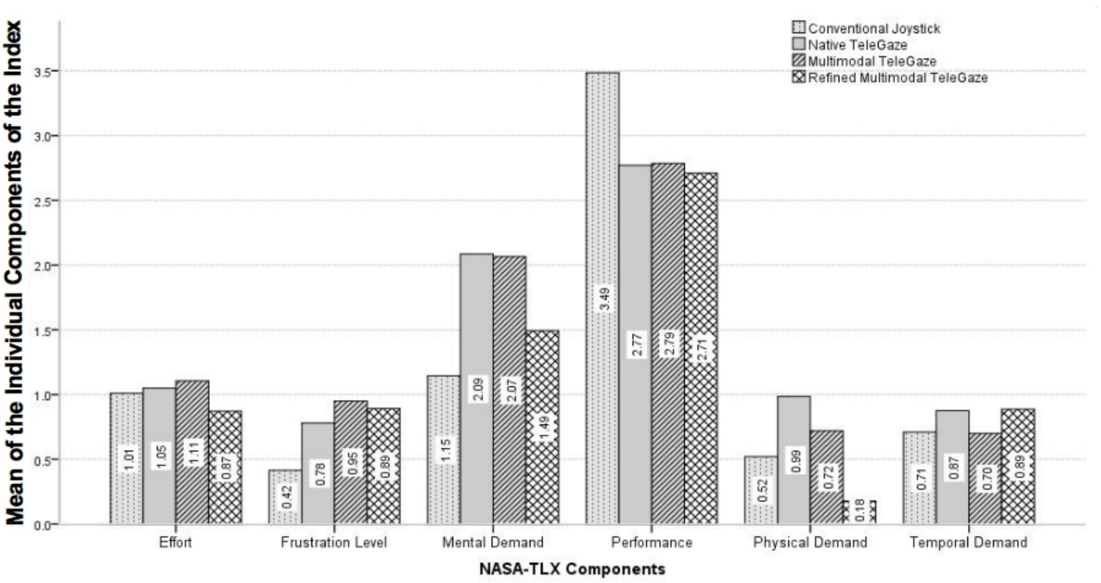


Figure 6.11: Average task load created by the individual components of NASA-TLX.

As it can be seen in the figure, in general the workload created by the refined multimodal TeleGaze is less than the other modes of interactions, except for temporal demands. This is consistent with the efficiency results since temporal demands are related to the time it takes to complete a task. As it has been mentioned in the earlier chapters, no specific time limits were set for the participants at any points in the research, except in the practising sessions. The fact that the participants felt higher temporal demands can be probably explained by their own will to finish quicker each time they repeat the task.

Kolmogorov-Smirnov normality test show that NASA-TLX for the refined multimodal TeleGaze is not significantly different from being normally distributed D(10)=0.188, p>0.05 (p145, [86]). This is similar to the results for the other three modes of interactions, which are reported in chapter 5. Therefore, the parametric test of ANOVA is used to test the hypothesis. Using Mauchly's test, the results of the test show that the assumption of sphericity is not violated $\chi^2(5)=6.443$, p>0.05 (p474, [86]).

Also it can be seen from the figure that both efforts and frustration levels are still higher for the TeleGaze than the joystick. These are expected results considering the novelty of TeleGaze in comparison with the conventional joystick. Also these results are consistent with the results for temporal demands, since these components can be highly related. Overall, despite improvements in some of the components when compared with the other two modes of the TeleGaze, the refined multimodal TeleGaze still is in competition with the joystick, with creating slightly less workload in this experiment.

6.4.3 Inspecting Methods

The findings from the inspecting methods used in the previous usability testing experiment were the main inspirations behind this stage of the research. The video recordings of what the subjects saw combined with their eye tracking data raised some issues in the design of the interface. These issues were difficult to spot without the used inspecting methods. Therefore, the same form of inspecting methods as in chapters 4 and 5 are used in this stage of the research. With the background knowledge of the design issues pointed out earlier by the inspecting methods, more focus were given to inspecting these issues in the refined design. In addition to the findings from the testing and inquiring methods, seeking the effects of the refinements in the design of the interface was the main objective of the inspecting methods used in this experiment.

Changing the default value of the camera auto-home configuration from enabled to disabled, created less panic and unpredicted situations. This finding is supported by lower values of both effort and frustration level in the NASA-TLX components scored for the refined multimodal interface (Section 6.4.2 and Figure 6.11). However, aligning the robot with the current pan angle of the camera, instead of aligning the camera with the current headings of the robot as it used to be in the previous design, turned out to be slower than expected. This is due to the hardware capabilities of the platform, as the camera pan/tilt unit is more responsive in comparison to the robot motors. Despite this fact, on average it took the participants less time-to-complete the task with the refined interface than the other three modes of interactions, including the joystick. Except this issue, no other issues are highlighted by the inspection methods in this experiment.

6.5 Discussion

The results of different evaluation metrics show improvements of the refined multimodal TeleGaze when compared with the other two modes of the TeleGaze experimented in the previous stage of the research. The refined multimodal TeleGaze not only scores higher usability than its TeleGaze interface counterparts, it also scores higher usability than the joystick target. Although not statistically significant, the refined multimodal TeleGaze scores higher efficiency, user satisfaction, and creates less workload on the subject than the other three modes of interactions. The results however, show that it has scored lower on effectiveness than the other modes.

People naturally do not repeat the same task in exactly the same manner every time. "*A slightly different perception will lead to a slightly different motor response, which in turn leads to another different perception and so on*" (p11, [92]). Therefore, considering the results for the other evaluation metrics, it is less likely that scoring less effectiveness is caused by any limitations in the interface. To summarize the differences in the obtained results of the evaluation metrics, Table 6.1 presents the mean rankings for all four interaction modes based on the evaluation metrics:

Table 6.1 : Average rankings of the interaction modes based on results of the evaluation metrics [54].

	Conventional Joystick	**Native TeleGaze**	**Multimodal TeleGaze**	**Refined Multimodal**
Efficiency (Time-to-Complete Task)	358.4 (*MR*= 1.00)	377.9 (*MR*=0.95)	412.0 (*MR*=0.87)	343.1 (*MR*=1.05)
Effectiveness (Overall Goal Achievement)	100.0 (*MR*= 1.00)	98.0 (*MR*= 0.98)	98.0 (*MR*= 0.98)	95.0 (*MR*= 0.95)
Satisfaction Questionnaire (1-5 Likert Scale)	4.20 (*MR*= 1.00)	3.25 (*MR*= 0.77)	4.00 (*MR*= 0.95)	4.00 (*MR*= 0.95)
Overall Task Load (NASA-TLX)	0.49 (*MR*= 1.00)	0.57 (*MR*= 0.86)	0.55 (*MR*= 0.89)	0.47 (*MR*= 1.04)
Relative Average Rankings	1	0.89	0.92	1
Final Rankings	**First Place Repeated**	**Third Place**	**Second Place**	**First Place Repeated**

From the results presented in the table, it can be seen that the average of mean ranking for the refined multimodal TeleGaze is equal to the average of mean ranking for the joystick target. This ranking is calculated based on the results for all the quantitative

---

[54] The results of the firs three modes in the table are the same from the previous chapter (Chapter 5), and the results for the refined multimodal TeleGaze is calculated based on the same principle explained there. The values for the joystick are still used as the datum for ranking the other modes of interactions.

evaluation metrics, with equal weights for each individual metric. It can be seen that the mean ranking for the refined multimodal interface is higher than the joystick for two of the metrics. However, lower rankings for the other two metrics results in the same rankings for both modes in question.

As the results of the statistical tests have shown earlier in sections 6.4.1 and 6.4.2 , non of the changes in the results are statistically significant. The usability index calculated for the TeleGaze against the joystick target depends on the significance of any differences. With no statistically significant differences in the results, the parameter values are the same for both the multimodal and the refined multimodal interface. Therefore, the value of the usability index for the refined multimodal interface equals 1, which is again the value of the usability index for the joystick target.

As a result, the refined multimodal TeleGaze has achieved the same level of usability of the joystick target depending on both mean rankings and statistical analyses of the results. Therefore, it can be further emphasized that *TeleGaze, with the refined multimodal interface, meets the joystick target as a means for mobile robot teleoperation. Furthermore, it enjoys the advantage of freeing both hands of the operator from the teleoperation task, which the joystick lacks.*

6.6 Interface and Design Principles

With a novel interface that achieves the same usability scores as a conventional one, it is time to revisit the design principles. The design principles for the TeleGaze interface are adapted from heuristics developed by other researchers in the fields of HCI and HRI [75], [19]. Therefore, the TeleGaze interface claims are checked against the heuristics adapted for TeleGaze. The answers to the claims are further supported by the findings and the results of the evaluation metrics used throughout the research. Table 6.2 presents the heuristics adapted for TeleGaze and the interface claims. Where applicable, the claims are supported by findings and result figures.

Table 6.2 : Checking the current design of the interface against the design principles adapted for TeleGaze.

Heuristics Adapted for TeleGaze	TeleGaze Interface Claims and Proofs
Is the interface interactive?	Yes. Providing equal amounts and quality of both controlling as well as monitoring is significant interactivity (Figure 6.8)
Is the interface responsive?	Yes. Achieving the same level of efficiency as a conventional joystick shows that the interface is responsive enough for the intended task (Figure 6.5)
Is the interface consistent?	Yes. Using the same design language for robot related actions, camera related actions, and even interface related actions shows the consistency of the interface.
Is the interface informative?	Yes. The robot status feedback, including the pan/tilt angles of the camera in real-time is sufficient information for the user to perform the task (Figure 5.2).
Is the interface intuitive?	Yes. The positioning of the camera related action regions along the edges of the interface is believed to be intuitive. User ratings prove this (Figure 6.9).
Is the interface elegant?	Yes. The relocatable combined forward/turning action region is presented in an elegant way that mimics conventional steering (**Figure 5.2**).
Is the interface familiar?	Yes, up to a certain limit. The novelty of the interface might work against the familiarity of the interface in comparison with conventional interfaces.
Is the interface flexible?	Yes. Having an inspection and an interaction mode is considered a significant flexibility in the interface (Figure 3.10). Also the relocatable forward action region adds to the flexibility of the interface (Figure 6.2).
Is the interface clear?	Yes. Using different geometric shapes shows the clarity of the interface and the functionality of the action regions. User ratings prove this (Figure 6.9).
Is the interface user friendly?	Yes. "*Easily commanding the robot as well as reporting execution information*" is considered human-friendly communication [2].

The claims of the interface might seem slightly overambitious. However, the findings from the evaluation metrics support most of the interface claims. The only heuristic that cannot be answered with a full *yes* is the familiarity of the interface. This is mainly due to the novelty of the interface and the TeleGaze system. Therefore, unfamiliarity of the interface cannot be used against such a novel interface at this stage. This apart, the interface substantially complies with all the design principles tailored specifically for TeleGaze.

6.7 Conclusions

By the end of the previous stage of the research, TeleGaze met its joystick target in terms of usability, from all considered point of views. However, some interesting findings from the usability testing experiment inspired this stage of the research. These findings highlighted some design elements in the interface that showed potentials for further improvements. Therefore, this stage of the research has been carried out to undertake some refinements in the interface and measure their affects on the usability of TeleGaze. At the end of this stage of the research, and based on the findings from the usability testing experiment, the following key conclusions can be drawn:

- Fine tuning some of the design elements might affect some aspects of the usability of the interface. However, the interface has reached a point where no further improvements can be achieved substantially, regardless of the quantity and quality of any refinements. Although the refined multimodal TeleGaze achieved higher efficiency and user satisfaction than the joystick and its earlier versions, the differences in the results are not statistically significant. Therefore, in terms of overall usability index, the refined multimodal TeleGaze has not improved from the multimodal TeleGaze interface experimented in the previous stage.

- Some findings, at a particular stage of the design process, might direct the design towards a particular direction, with highly promising achievements. However, exploring the proposed direction is necessary to prove or disprove the findings. When the new direction does not turn out to be as expected, redirecting the design process is the next step. In the case of TeleGaze, the findings from the previous stage inspired this stage of the research, with highly promising improvements. However, the results from the experiment show that the improvements are not statistically significant. Therefore, the findings that inspired this stage of the research turned out not to be as expected.

By the end of this stage, *TeleGaze can confidently claim that it has met the joystick rival in terms of usability as a means for mobile robot teleoperation. It can also claim the advantage of freeing the hands of the operator from any controlling tasks required for teleoperation*.

Chapter SEVEN

Conclusions and Future Work

The work presented in this thesis is a novel attempt to answer the research question both from theoretical and practical points of view. Original knowledge on a novel means of mobile robot teleoperation has been obtained through three consequential phases of design, evaluation, and refinements. The main focus of this work has been the design and usability of a novel interface for human-robot interaction (HRI). With this focus, the work has produced empirical results on the feasibility and usability of mobile robot teleoperation through eye gaze, TeleGaze. In this chapter the research conclusions with critical discussion, the originality and novel elements of this work and directions for some future works are covered.

7.1 Conclusions and Critical Discussion

The first part of the research question inquired the feasibility of controlling a mobile robot from a remote location solely through inputs from human eyes. From the work presented in this thesis it can be concluded that the answer to this part of the research question is positive. The usability experiment presented in chapter 3 showed that it is possible to perform a navigational task with the native TeleGaze the same as with a conventional joystick. This was achieved with the significant advantage of

TeleGaze, which is totally freeing both hands of the operator from the interaction. However, at this stage of the research the same performance level and the same user satisfaction level of the joystick were not achieved. Therefore, initially the answer to the second part of the research question, how comparable TeleGaze is to other means of HRI, was negative. It was not possible initially to teleoperate a mobile robot using TeleGaze with the same level of performance and user satisfaction as of a joystick in the same navigation task.

To better quantify the answer to the second part of the research question a set of evaluation metrics composed of testing, inquiring, and inspecting methods was designed. Based on observations and results from the first phase of the research reported in chapter 3, some further refinements were made to the TeleGaze interface in the second phase of the research. To overcome the Midas-Touch problem observed in this phase, a novel multimodal TeleGaze was designed. Using the set of evaluation metrics and more sophisticated navigational tasks explained in chapter 4, usability experiment was carried out for the multimodal TeleGaze. Results of this usability experiment showed that the multimodal TeleGaze delivered the same level of performance and user satisfaction as a conventional joystick. Although, TeleGaze still held the advantage of totally freeing both hands of the human operator from the teleoperation.

Both parts of the research question were answered by the end of the second phase of the research with the multimodal TeleGaze. However, some interesting observations during the usability experiment grabbed attention and looked promising. This led to the third phase of the research by further refining the interface and repeating the usability experiment. The results showed further improvements in the performance of the multimodal TeleGaze when compared to all previous modes of interaction experimented in this work, including the joystick. With the results obtained in this phase, the answer to both parts of the research question has become clearer. TeleGaze achieved the same level of performance and user satisfaction of a joystick with the advantage of totally freeing both hands of the human operator. This is a very interesting conclusion due to the fact that most of the participants had prior interaction experience using joysticks but no experience using eye tracking.

Care must be taken in generalizing the above conclusions however, if TeleGaze is to be used for a wider rage of applications and interaction scenarios. To critically review these conclusions, it is necessary to revisit the research boundaries set at the beginning of the work. Also the navigational task and the experimental environment used in the usability testing experiments require further discussion.

Mollenbaeh argues that "*certain types of tasks are more suited for gaze interaction than others"* [22]. "*Identifying these and creating solutions which employ the particular strengths of eye-tracking is the key to using this rapidly advancing technology*", the argument continues. Where research shows improvement of eye driven interactions for certain tasks, it shows otherwise for other tasks [113]. In an application environment such as the one used in this research, subjects are likely to favour TeleGaze over a conventional joystick for mobile robot teleoperation. This is due to the convenience of moving while looking. However, these results are obtained for this particular application environment where there is a certain level of flexibility in the requirements of the task. Precise driving, for example, is not a major concern in such application environments, which makes it easier for TeleGaze to meet its target. In applications where precision is a major concern however, it is likely that the joystick would be more effective than TeleGaze.

Due to physiological reasons, people mostly have better control over their hands than their eyes because gazing behaviours are difficult to control in dynamic environments [23]. Research shows that "*hands-free control requires a heavy investment in operator training, and this aspect of achieving successful operation should be considered strongly before application areas are further explored*" [21]. This poses a major limitation on generalizing TeleGaze for other applications where precise movements are major concerns and not free hands. If users feel that the same level of precision of a joystick cannot be achieved with the TeleGaze interface, then it is likely that they prefer their hands to be in control. This is an issue of compromising one benefit for another. A trade-off of control precision for free hands, which in some applications, is likely to undervalue TeleGaze and the privilege of hands-free mobile robot teleoperation.

Hands-free mobile robot teleoperation is an advantage if users need their hands to perform other tasks while driving the robot. As Sibert and Jacob argue, "*eye gaze interaction is a reasonable addition to computer interaction and is convenient in situations where it is important to use the hands for other tasks*" [73]. However, the current eye tracking systems pose certain restrictions on the movements of the operator. This means that free hands are less likely to be used to perform other tasks.

In the usability experiment of TeleGaze, the participants were forced to use their hands to write down the readings from the paintings while using TeleGaze. This is a likely scenario where during teleoperation operators have to be engaged in other tasks. Hence, the advantage of free hands can be appreciated. On the other hand, if freeing hands is not necessary for a particular application then, hands-free teleoperation is less appreciated. Consequently, TeleGaze will be less valuable as a tool for mobile robot teleoperation. However, the advantage of an intuitive interface, which is easy to learn and recall, is hard to be beaten by joysticks regardless of the application and the interaction scenario.

7.2 Originality and Novel Elements

Research on using eye tracking as a means of control, to aid or substitute other means of control, is mostly addressed towards disabled people. This can be justified by the limited options that this stream of people have when it comes to means of control. Most disabled people who suffer from spinal cord injury, amputation, or quadriplegia face difficulties using their hands for controlling purposes [33]. Therefore, any level of control when compared to no control at all is an advantage for the user. Consequently the interaction experience that eye tracking creates for these users is most welcomed. In most cases, not because it provides a better experience or means of control, but because it makes the interaction experience happen in the first place.

One element of originality of this work lies in the fact that it is aimed at mainstream people. People who have the choice of using both conventional forms of control, such as the common joystick, and the novel form of control, that is TeleGaze. The work presented in this thesis has demonstrated that this approach is more interesting and challenging. Interesting because it is aim is to deliver the same level of control as a conventional joystick to the potential users. Challenging on the other hand, because it is aim it to achieve the same level of user satisfaction and trust in the system as of a conventional joystick. This element of originality defines where this work stands in comparison to other works that have used eye tracking to substitute conventional means of control.

Another element of originality is the main focus of this work. As Dix and colleagues say "*the interface is not something that can be plugged in at the last minute*" (p3, [32]). Most previous works on using eye tracking to control robotic agents have focused on addressing eye tracking problems. Different algorithms have been developed and HRI has been used as a context for testing these developed algorithms [12], [61]. This work therefore, is original in the sense that it has focused on designing an interface for HRI using eye tracking. The TeleGaze interface that has been developed in this work is not original as far as associating regions to certain actions is concerned. However, one of the novelties of the interface is using transparent action regions on top of the live images streaming from the on-board vision system. This novel approach has helped in achieving optimal use of space and intuitive interface design. The design of an intuitive and natural interface for TeleGaze has been the dominant direction of this work and one of its elements of originality.

This work is also original in the sense that it has conducted an extensive evaluation and experiment design to test the usability of the proposed TeleGaze interface. Alvarez-Cortes and colleagues report that a quick scan of many user interface articles "*reveals that only one third of the articles include any type of evaluation*" which is too low of a percentage [28]. This work has brought together a set of multidisciplinary evaluation metrics to evaluate TeleGaze against the design principles and its target. The experimental environment, the navigational task, and the set of evaluation metrics have played a major part in the originality elements of the work. No evaluation at this level of details has been done prior to this work as far as using eye tracking for HRI is concerned.

Another original element of this work is adding an accelerator pedal as a novel multimodal approach to overcome the Midas-Touch problem. This approach is not original as far as adding additional forms of control to overcome limitations of eye tracking is concerned. Speech [106], mouse [38], key strokes [87], blinks and facial muscle movements have all been experimented as additional modes of interactions used with eye tracking. Using a keyboard, for example, raises the issue of hand-eye coordination [18]. While, using an accelerator pedal with eye tracking raises the issue of foot-eye coordination. Foot-eye coordination has been experimented in some contexts [42], but has never been experimented in the context of mobile robot teleoperation. Therefore, the novel use of an accelerator pedal adds to the number of multimodal interfaces used to overcome the limitations of eye tracking.

In addition to the individual originality elements mentioned above, this work has produced a novel interactive system. TeleGaze, as an interactive system, is novel in using inputs from human eyes for mobile robot teleoperation. The design and the usability testing of the TeleGaze interface using multidisciplinary evaluation metrics add to the novelty of the system. To the best of the author's knowledge, no similar work in combining the individual elements of this work has been conducted and reported in the literature prior to this one. Some works however, triggered by the early publications on this work, have started and cited this work [23].

7.3 Directions for Future Work

7.3.1 Speed Control in TeleGaze

The purpose of the pedal used in the multimodal TeleGaze was only to substitute the dwell time technique used earlier in the native TeleGaze. The functionality of the pedal has been limited to issuing commands in-line with the action region that contains the point-of-gaze (POG). It is a digital pedal that has been functioning as either *on* when pressed or *off* when released. Speed has been out of consideration in this research as mentioned in the research boundaries in chapter 1. However, control over speed is an interesting, and sometimes necessary, aspect of control that can be integrated into the TeleGaze system.

One appealing solution, as far as speed is concerned, is to add additional action regions to the TeleGaze interface. Multiple action regions associated with different values of the same action can be used as different speeds for the action. For example, three forward action regions can all issue the forward command but each at a different driving speed than the other two. This approach, although interesting as it might seem to be, will add significantly to the overall number of action regions placed on the interface. Consequently, the interface becomes more cluttered with extra action regions which affects the intuitiveness and naturalness of the interface.

An alternative appealing solution therefore, is to use an analogue accelerator pedal to provide this granulation of level of control. Similar to how conventional accelerator pedals function in driving cars, the level of the pressure can determine the magnitude of a particular action. This can be the linear velocity of the robot in the case of the forward action while it can be the rotational velocity in the case of the turning actions. Alternatively, it can be interpreted as the proportional relation between the values of both the linear and the angular velocities which determines the turning radius of the robot. When it comes to controlling the camera, the level of the pressure on the pedal can be interpreted as either the pan/tilt speed or their values. Therefore, the effect of the level of pressure of the accelerator can be handled in a contextual way according to the active action. Other possibilities also are likely to appear when integrating an analogue accelerator pedal to TeleGaze.

In this case, as far as the usability testing of TeleGaze is concerned speed needs to be added to the competitive joystick in order to obtain comparable results. It is likely that more control over speed can be achieved with the multimodal TeleGaze than a joystick. With the multimodal TeleGaze only one analogue contextual pedal can be used to add granulation of control to all the actions. It is more difficult however, to add pressure sensitive behaviour to all the buttons on a joystick. This might turn out as another interesting advantage of TeleGaze when compared to a joystick.

7.3.2 Goal Setting by Gazing

To move towards a target using TeleGaze, the operator is required to issue the sequence of commands that generates the necessary robot movements. Since the operator looks through the eyes of the robot, any visible target to the robot is visible to the operator. When a target is visible to both the operator and the robot, gazing can be used to set the target as a goal for the robot. This fact can be exploited to add some autonomy elements to TeleGaze. Using TeleGaze in its current format, the operator is required to take responsibility of all kinematic calculations. In the suggested use however, all the operator has to do is to set a goal and all kinematic calculations is done by the robot [42].

Using the multimodal TeleGaze, it is interesting to design a sister interface that enables goal setting by gazing. With a clear and visible target, a painting on the wall for example, the operator can gaze at it on the interface and press the pedal to set it as the goal for the robot. The robot then should start navigating[55] to the target as long as the distance between the robot and the target is within a predefined threshold. Although this capability needs to be autonomous, the operator should have a supervisory control authority [3]. This enables the operator to either override the goal or cancel it at all and gain back full control when necessary. Furthermore, this version of TeleGaze should not substitute the TeleGaze that provides the operator with full control. This addition to the interface should only work as a mode of TeleGaze that is activated and used when needed for ease of navigation.

To add this functionality as it is proposed here, a number of other key functionalities are required. Path planning for example is a major requirement if the robot is to navigate autonomously from one point to another. Automated obstacle avoidance is another major requirement which is more challenging. Having all these functionalities to operate together smoothly to achieve an overall goal is more complicated than having them working individually [66]. Avoiding obstacles, for example, should not conflict with navigating towards the target. If the robot needs to get out of the calculated route to avoid obstacles, the target is likely to get out of the vicinity of the robot. Therefore, the robot should keep tracking of the target while avoiding any obstacles. An appealing approach is to dedicate the camera to track the target regardless of the moves that the robot has to perform in order to avoid the obstacles. Another approach is to keep tracking all the moves of the robot while avoiding obstacles in order to reverse them afterwards.

---

[55] Navigation requires the robot to know a)- where it is? b)- where it has to go? and c)- how to get there? at any time. (Chapter 1, Footnote 2).

Motivated by this idea, the author has done some initial investigations along this work. A sister interface to the native TeleGaze has been developed that enables target selection by gazing. The work has focused on a moving target, such as a walking person, rather than a static target of a mounted painting for example. For this purpose, a combined TeleGaze and automated person following algorithm has been developed. The version of TeleGaze in concern enables the operator to select a person, visible to both the robot and the operator, as a target. The goal of the robot then is to follow that person while keeping a predefined distance from him/her. This has been easier to implement as no path planning neither obstacle avoidance are necessary in this context. The distance between the target and the robot is small enough for the robot to keep tracking the person without the need for path planning. Also due to the short distance and the person's awareness it is less likely that the robot needs to avoid any obstacles while following him/her. Therefore, to start with the system, path planning and obstacle avoidance have been left out.

Due to time limitations, this part of the work has not included usability testing or evaluations of the interface. However, evaluating such capabilities of TeleGaze is likely to demonstrate significant advantages over a conventional joystick. A more detailed description of this version of TeleGaze is included in Appendix B.

7.3.3 Personalized Action Region Sizes

The command buttons on gaze driven interfaces are usually designed to be larger than buttons on conventional interfaces. This is due to the limited accuracy of eye tracking systems and the difficulty to fine control the gaze. However, very limited information is reported in the literature on the methods of determining the size of the buttons on such interfaces. Therefore, in TeleGaze, different approaches in different phases of the research have been explored.

The fact that the action regions on TeleGaze do not obstruct the view due to their transparency created significant flexibility in terms of size. Also due to the existence of the accelerator pedal, altering the size of the action regions for the multimodal TeleGaze has not been a big risk either. In the multimodal TeleGaze, regardless of the size of the action regions, the risk of issuing a command unintentionally did not exist. Commands were only issued when the pedal was being pressed. However, determining the optimum size of action regions plays a significant role in other contexts and for other interaction applications.

Due to differences in characteristics of people, no size is optimum for all users of any gaze driven interface. Therefore, adapting interface features for each individual plays a significant role in decreasing some elements of the task load. Findings of the usability experiments in this work further supports this argument. It was found that the same position of the forward action region does not suit all the participants, which means there is no ideal position that suits all users. Therefore, a relocatable forward action region was introduced in order to position the action region differently for each individual user. However, the same action region size has been used for all the participants for any phase of the research.

Different approaches can be used to determine the optimum button size for each individual. The size threshold can be set individually for each individual user similar to what has been reported in [33]. Setting the size threshold for each individual might not be practical in some applications. Also it is likely that most users prefer larger sizes regardless of its necessity. Larger button sizes occupy more space on the interface which might be critical in some application contexts. Therefore, a more interesting and appealing approach is to develop a size calibration mechanism that is not dependent on user decisions totally.

An accuracy calibration procedure is conducted in most eye tracking applications already. This calibration can be extended to gather necessary information on the optimum size of the buttons for each individual. This will not have a significant impact on TeleGaze only, but also on a wide range of other eye tracking applications. The fact that no such calibration mechanism has been developed as yet, despite the long history of eye tracking, might indicate the difficulty of developing such systems. However, it is one of the most important and most interesting challenges in eye tracking that requires attention and devoted efforts.

7.4 Final Thoughts

Design of interactive systems is an iterative process, and it never reaches an end as long as the system is being used. One likely reason might be flaws in the design of the system that are being highlighted through long-term interactions. Another reason is the fact that users' needs and expectations, in terms of functionalities and usability, change over time while interacting with the system. That is why giant system designers, such as Microsoft for example, refine and modify their systems on a continuous base. Interestingly, with all the experience and resources that a giant company such as Microsoft has, there are occasions when its product does not satisfy its target users. Consequently, the product gets replaced with an alternative as soon as possible, similar to what happened recently to Windows Vista. Also user preferences vary significantly, in terms of appearance and layout. This is true regardless of the core functionalities and the aim of the system.

The TeleGaze interface presented in this work has outperformed a conventional joystick for mobile robot teleoperation. This has been demonstrated through a set of well designed experiments and usability testing with particular emphasis on a gallery-like scenario. The work has also exploited the correlation between gazing behaviours and moving intentions of human beings in this context. More interestingly, it has demonstrated that through direct interpretations of gazing behaviours, the users' motion intentions can be translated into robotic actions. However, due to the reasons mentioned above, this work does not claim that TeleGaze is the best means of HRI. It does not claim either that the final design of the TeleGaze interface is the absolute optimum design that can be achieved for such an interface.

It might be argued that a more sophisticated and complicated algorithm is necessary to achieve the required interpretations of gazing behaviours. Approaches such as artificial intelligence [28], user modelling [114], and probabilistic models [79] are likely to be suggested. Alvarez-Cortes and colleagues argue that "*AI techniques are often slow, and can make what should be an interactive interface a slow and unresponsive interface*" [28]. They also argue that "*users need to have a clear mental model of how the computer will respond to their input, and some uses of AI actually blur this model*". Benyon and colleagues argue that "*a key design principle is to design* *things so that people will form correct and useful mental models of how they work and what they do*" (p32, [19]). Therefore, one of the known disadvantages of using these approaches is that "*the models they generate tend to be black boxes, which do not allow one to understand the relationship between input data and model prediction*" [114].

Inline with these arguments, the researcher believes that the best design of an interactive system is the one that the end user understands. This is because, at the end of the day, it is the user who interacts with the system and not the designer. It is difficult to build interaction bonds between users of a system and the system itself if lack of understandability exists. This is likely to happen if the users are not sure how a particular behaviour caused a particular result. The fact that it took only one minute to explain how TeleGaze works and took the users an average of two minutes to decide that they are ready to use TeleGaze, demonstrates the understandability of the design. This work has demonstrated that there is no need for complexity to achieve the required interpretations of gazing behaviours. There is no need for complicated approaches if the aim can be achieved through simpler solutions since the best inventions are the simplest, at least most of the time.

Appendix A

Evaluation Metrics' Supplements

A.1 The Questionnaire Used in the Observational Study of the Native TeleGaze[56]

The following questionnaire was used in the observational study of the native TeleGaze in the first stage of the research. The participants drew a circle marking their preferred answer.

1- Centred regions with edge free interface or edge located regions with centre free interface?

a- Centre b- Edge

2- Enlarged top view for precise movement?

a- Yes b- No

3- Relation between camera and robot action regions?

a- Overlapped b- All Centred (Separated)

---

[56] Mentioned in Chapter 3, Section 3.5.3

4- Automatic camera home configuration?

a- Yes b- No

5- Blocking camera and continue moving?

a- Yes b- No

6- Explicit layouts of action regions?

a- Yes b- No

7- Displaying the point of the gaze?

a- Yes b- No

8- Captions?

a- Dynamic b- Static c- No-Captions

After trying each one of the prototypes, the participants filled the above questionnaire in order to show their preferences in terms of some design elements of the experimented prototypes.

A.2 The Questionnaire Used in the Task-Oriented Evaluation of the Native TeleGaze[57]

The following questionnaire was used in the task-oriented evaluation of the refined interface of the native TeleGaze. The participants rated their agreement on the Likert scale, where 1 is strongly disagree and 5 is strongly agree, by drawing a circle on the number that best represents their agreement. The questionnaire is composed of two sections. Following is the first section, which addresses the interaction experience and the design of the interface:

Argument	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
1. It was easy to learn all the features of the system.	1	2	3	4	5
2. The training and explanation provided was enough to learn the system.	1	2	3	4	5
3. It was simple to use the system.	1	2	3	4	5
4. It is easy to find the action regions that I need.	1	2	3	4	5

---

[57] Mentioned in Chapter 3, Section 3.7.2

5. The system has all the functions and capabilities I expect it to have.	1	2	3	4	5
6. I believe I became familiar with the system very quickly.	1	2	3	4	5
7. I felt confident using the system to accomplish the task.	1	2	3	4	5
8. I can do more complex tasks with the current capabilities of the system.	1	2	3	4	5
9. I will get more out of the system with more training and experience (i.e the system is learnable and more training will give you better performance).	1	2	3	4	5
10. The positioning of the action regions is due to my satisfaction in terms of edge, centre, right or left.	1	2	3	4	5
11. The size of each action region fulfils the purpose.	1	2	3	4	5
12. I like using the interface of this system.	1	2	3	4	5
13. I would use the system to navigate a mobile robot.	1	2	3	4	5
14. I would recommend the system for other people.	1	2	3	4	5
15. I would like the system to be developed further as it is worth it.	1	2	3	4	5
16. Overall, I am satisfied with the system.	1	2	3	4	5

The second section in the questionnaire, inquired about the possible sources of limitations if the participants have spotted any. The participants were asked to rank the most likely source of limitations in the following table:

Reasons for Limits and Problems	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
17. The design of the interface. For example the layout and positioning of the action regions.	1	2	3	4	5
18. The way the system works. For example issuing a command by looking at a region for a third of a second.	1	2	3	4	5
19. The eye tracking equipment and technology. (i.e a more convenient eye tracking equipment will solve most of the problems)	1	2	3	4	5
20. Any other reasons you might think of. (Please explain or discuss) …………………………………………………………………… …………………………………………………………………… ……………………………………………………………………	1	2	3	4	5

This questionnaire was filled after completing the navigational task in the task-oriented evaluation of the refined interface of the native TeleGaze. As it has been mentioned in Chapter 3, in addition to this questionnaire other evaluation metrics were used also (Chapter 3, Section 3.7.2).

A.3 The Questionnaire Used in the the Multimodal TeleGaze[58]

The following questionnaire was used as part of the evaluation metrics used in the evaluation of the multimodal TeleGaze:

Glossary of modes:

Mode 1 of Interaction is: ………………………………. Sequence:………..,

Mode 2 of Interaction is: ……………………………..… Sequence:……..….,

Mode 3 of Interaction is: ……………………………..… Sequence:………..

Rank your agreement with the following statements using a 1 to 5 scale, where 1 is *strongly disagree* and 5 is *strongly agree:*

Mode 1 of Interaction:

Statements	Your Rankings				
1- I found it easy to learn how to use the system:	1	2	3	4	5
2- The system has all the capabilities I expect it to have:	1	2	3	4	5
3- I felt confident using the system:	1	2	3	4	5
4- It was difficult to get familiar with the system and how it works:	1	2	3	4	5
5- I don't think I can perform better than how I did:	1	2	3	4	5
6- I don't see the system appropriate for this kind of applications:	1	2	3	4	5
7- I would like more chances to try more difficult tasks:	1	2	3	4	5
8- Overall I am satisfied with the system:	1	2	3	4	5

The same above set of questions were filled by the participants after each mode of interaction for all three modes of interactions.

The following set of questions inquires the design of the interface and not any particular mode of interactions. Therefore, it was filled at the end of the whole experiment after all three modes of interactions.

1- *(To answer this question, draw a circle as close as you agree to the term on one end of the statement.)*

The aim of the TeleGaze interface is to provide you with two capabilities simultaneously:

Monitoring and ***Controlling***

---

[58] Mentioned in Chapter 4, Section 4.7.2 and Chapter 5, and Section 5.7.2

Which one of the capabilities do you think the interface provided **MOST**?

Monitoring	*	*	*	*	*	Controlling

And which one of the capabilities do you think the interface provided **BEST**?

Monitoring	*	*	*	*	*	Controlling

2- The positioning of the action regions are meant to be intuitive. How do you rank the positioning of the action regions from this point of view?

Robot controlling action regions:

1	2	3	4	5

Camera controlling action regions:

1	2	3	4	5

3- Different geometric shapes are used to differentiate between robot controlling action regions and camera controlling action regions. How useful did you find this approach?

1	2	3	4	5

4- Which functionality and/or part of the interface did you ***like*** most?

………………………………………………………………………………………

And which one did you ***dislike*** most?

………………………………………………………………………………………

5- How appropriate do you rank the sizes of the action regions?

1	2	3	4	5

6- Overall how do you rank using the interface:

1	2	3	4	5

A.4 The Rating Sheet, Weighting Sheet, and Definitions of the NASA-TLX[59]

RATING SHEET

● **Mental Demand:** How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex?

Low High

● **Physical Demand:** How much physical activity was required (e.g., pushing, pulling, turning, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low High

● **Temporal Demand:** How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Low High

● **Effort:** How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low High

● **Performance:** How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?

Low High

● **Frustration:** How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Low High

---

[59] Mentioned in Chapter 4, Section A.4, Chapter 5, Section 5.7.2, and Chapter 6, Section 6.4.2

The definitions of the components are obtained from the instructions manual of the index provided by NASA[60]. The above rating sheet and the following weighting sheet were filled by the participants for each mode of interaction.

Effort Or Performance	Temporal Demand Or Frustration	Temporal Demand Or Effort
Physical Demand Or Frustration	Performance Or Frustration	Physical Demand Or Temporal Demand
Physical Demand Or Performance	Temporal Demand Or Mental Demand	Frustration Or Effort
Performance Or Mental Demand	Performance Or Temporal Demand	Mental Demand Or Effort
Mental Demand Or Physical Demand	Effort Or Physical Demand	Frustration Or Mental Demand

The overall index is calculated based on the instructions provided in the same manual mentioned above. The index can be filled and calculated using either a computerized version or a paper based version. It was believed that the paper based version is easier to use and understand. Therefore, this version was used instead of the computerized version.

---

[60] The instructions manual and more details on the index can be found at http://humansystems.arc.nasa.gov/groups/TLX/index.html (Last Accessed on 14th April, 2010)

A.5 The Questionnaire Used in the Refined Multimodal TeleGaze[61]

The same questionnaire used in the evaluation of the multimodal TeleGaze was used in the evaluation of the refined multimodal TeleGaze with some modifications. The following is the questionnaire in the form that is used in the evaluation of the refined multimodal TeleGaze.

Sequence of Readings: ………………………………

Rank your agreement with the following statements using a 1 to 5 scale, where 1 is *strongly disagree* and 5 is *strongly agree:*

Statements	**Your Rankings**				
1- I found it easy to learn how to use the system:	1	2	3	4	5
2- The system has all the capabilities I expect it to have:	1	2	3	4	5
3- I felt confident using the system:	1	2	3	4	5
4- It was difficult to get familiar with the system and how it works:	1	2	3	4	5
5- I don't think I can perform better than how I did:	1	2	3	4	5
6- I don't see the system appropriate for this kind of applications:	1	2	3	4	5
7- I would like more chances to try more difficult tasks:	1	2	3	4	5
8- Overall I am satisfied with the system:	1	2	3	4	5

1- *(To answer this question, draw a circle as close as you agree to the term on one end of the statement.)*

The aim of the TeleGaze interface is to provide you with two capabilities simultaneously:

Monitoring and ***Controlling***

Which one of the capabilities do you think the interface provided MOST?

Monitoring	*	*	*	*	*	Controlling

And which one of the capabilities do you think the interface provided BEST?

Monitoring	*	*	*	*	*	Controlling

---

[61] Mentioned in Chapter 6, Section 6.4.2

2- The positioning of the action regions are meant to be intuitive. How do you rank the positioning of the action regions from this point of view?

Robot controlling action regions:

1	2	3	4	5

Camera controlling action regions:

1	2	3	4	5

3- Different geometric shapes are used to differentiate between robot controlling action regions and camera controlling action regions. How useful did you find this approach?

1	2	3	4	5

4- Which functionality and/or part of the interface did you ***like*** most?

………………………………………………………………………………………

And which one did you ***dislike*** most?

………………………………………………………………………………………

5- How appropriate do you rank the sizes of the action regions?

1	2	3	4	5

6- How do you rank the functionality of the moving-forward action region?

1	2	3	4	5

7- Overall how do you rank using the interface:

1	2	3	4	5

Appendix B

TeleGaze for Mobile Robot Person-Following

B.1 Introduction

In the effort of developing natural means for human-robot interaction (HRI), significant amount of research has been focusing on Person-Following (PF) for mobile robots. PF, which generally consists of detecting, recognizing and following people, is believed to be one of the required functionalities for most future robots [115]. Therefore, it is becoming an increasingly popular research topic in the field of robotics with significant progress towards robust and reliable implementation of this functionality [116].

Research in this field is mostly directed towards fully automating this functionality, which makes the challenge even more tedious. Focusing on this challenge leads research to divert from other challenges that coexist in researching any PF system. A natural PF functionality consists of a number of tasks that are required to be implemented in the system. However, in more realistic life scenarios, not all the tasks required for PF need to be automated. Instead, some of these tasks can be operated by human operators and therefore require natural means of interactions and practical balance between automation and operation.

In order to highlight all the tasks that are believed to exist in any PF system, a novel PF taxonomy has been introduced by the researcher. Also, in order to provide a natural means for HRI, TeleGaze is used in the implementation of the introduced PF taxonomy. The work detailed in this thesis, was inspired by previous studies involving using a robot to perform a PF application. This application was subsequently extended to include eye-gaze operation at the start and stop of the person-following operations using eye-gaze input. A description of this system is therefore included here.

In this appendix, the PF taxonomy, few interaction scenarios and the integration of TeleGaze into the PF taxonomy are covered. This chapter is mainly based on the researcher's publications on this topic. However, it has been included here as an example for a realistic implementation of TeleGaze. Also as a direction for likely future work on goal-directed implementations of TeleGaze, which is believed that it adds substantially to the functionalities of TeleGaze. Goal-directed TeleGaze is suggested in Chapter 7 (Section 7.3.2).

B.2 Terminology Definitions

Before digging into the PF taxonomy and the different tasks that are involved in developing any PF system, it is necessary to clarify and define some terminologies that are widely used. This is necessary due to the fact that the terms *tracking* and *following* are used in the literature to refer to the same meaning and/or different meanings interchangeably [115], [116], and [117]. Therefore, in order to standardize the use and the meaning of these two terms in PF applications and future writings, it is necessary to define them in this context.

Tracking is going to be used in the taxonomy to refer to the set of actions taking place in order to keep the POI in the vicinity of the robot without altering the physical position of the robotic platform. This might include digital, optical and physical actions of only the active vision system of the robot and not the whole robotic platform. Digital and/or optical zooming, for example, might be used to keep the appearance of the POI in the scene at a certain ratio of the whole scene. Also pan/tilt might be used to keep the POI in a certain area of the scene.

Following, on the other hand, is going to be used in the taxonomy to refer to the set of actions taking place in order to keep the POI in the vicinity of the robot by altering the physical position of the robotic platform. This, in its basic form, consists of the four common actions of forward, backward, left, and right. This task requires distance information to keep the robot at a desired distance of the moving target while avoiding accidents that might occur if getting too close to the target.

B.3 Taxonomy of Person-Following

The challenge of keeping track of the Person-Of-Interest (POI) is believed to be the main challenge in any PF application. This challenge is mostly addressed through modifying or developing object tracking algorithms used to keep track of the POI [118]. Or, in some cases, to cope with variations in the interactions' conditions, fusion of cues and algorithms is used to address the problem [119]. However, a complete PF system is not limited to this challenge only.

Regardless of the complexity of the applications and the likely scenarios, a complete PF system consists of a number of tasks that each might raise a number of challenges during the course of interaction and the implementation of the PF functionality. The aim of the PF taxonomy introduced here is to highlight the tasks involved in developing any PF system. All the tasks presented in the taxonomy are required to be implemented in a natural form of HRI regardless of the application context. In addition to the tasks themselves, the taxonomy presents a number of likely interaction scenarios in the form of Loops-Of-Interactions (LOI), where each loop consists of a number of tasks. The complete PF taxonomy is illustrated in Figure B.1.

Notice the difference between Person-Following (PF) as the entire system and person-following (pf) as an individual task in the overall system. The ideal LOI is presented in the taxonomy with thick-continuous lines starting from task one and ending with task eight. However, different loops in the taxonomy represent different interaction scenarios that are likely to happen in any PF application. Although, for instance, it is most likely that task two will start once task one is accomplished, task eight might start instead after task one if a wrong person is registered. Therefore, the LOI that consists only of tasks one and eight is a likely interaction scenario in real life PF applications. The mentioned scenario explains the importance of the taxonomy and how a PF application needs to address more than just the problem of tracking and following the POI. Examples of other interaction scenarios can be found in the publications.

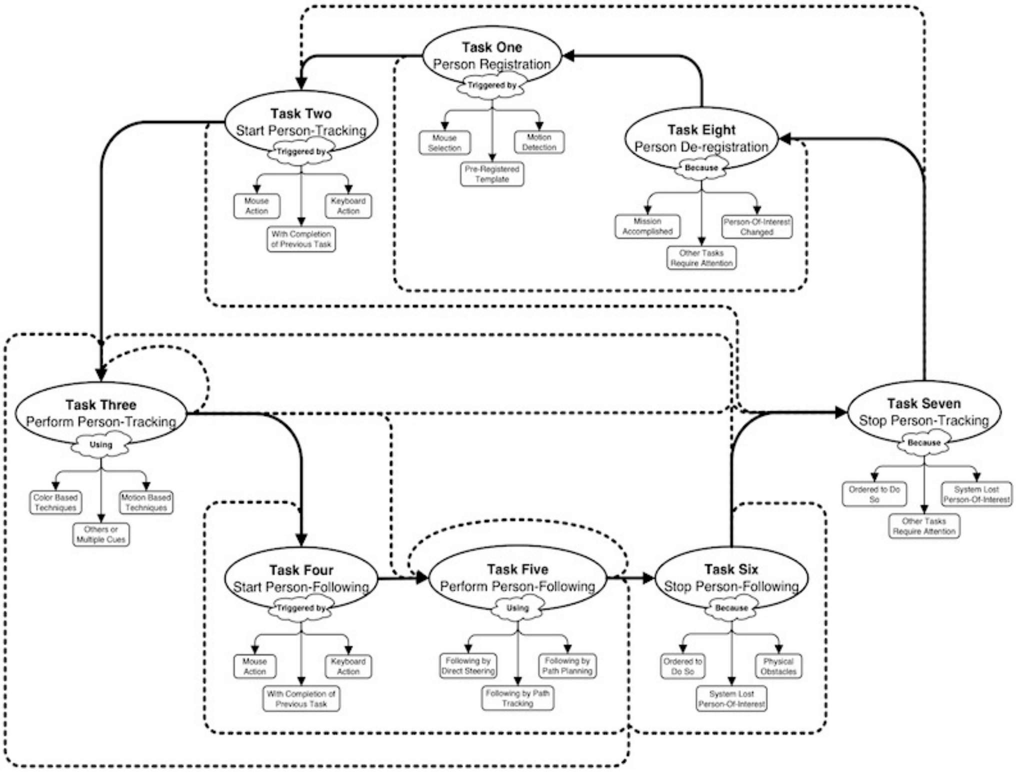


Figure B.1: The Person-Following Taxonomy for Mobile Robots.

B.4 The Implementation of the Taxonomy

The forms of information acquisition for both the system and the human operator vary depending on task requirements. The combination of autonomous and non-autonomous functionalities in one application is a common approach in developing many robotic systems [120]. Some of the tasks in the taxonomy can be either operated or automated. This means that not all the tasks presented in the taxonomy require automation. In fact, some of them make more sense when they are operated by a human operator and not automated.

One of the tasks for example that is most likely to require operation and not automation is registering the POI (task one). However, this does not mean that operating the task should be achieved in an artificial way and not considered from a natural HRI point of view. Implementing this task has been achieved in a number of different ways, as reported in the literature, so far such as using a mouse selection, people detection [121], motion detection [117], or even a pre-registered template such as a predetermined colour of the POI [118]. This task however, when operated, needs to be implemented in a more natural way of HRI interaction.

Also some of the other tasks such as starting person-tracking (task two), starting person-following (task four), stopping person-following (task six), stopping person-tracking (task seven) and finally person deregistration (task eight) can be operated in a PF application and not automated. Some of these tasks are merged into one task in some applications, such as starting person-tracking (task two) once the person registered (task one) and then starting person-following (task four) once person- tracking (task two) started. However, in a more realistic application each one of these tasks needs to be implemented once the conditions for their implementation are met and not as a group of tasks altogether. Therefore, an ideal PF application needs to deal with invoking each task separately from the other tasks in the taxonomy while it enables a natural HRI form of invoking each task. TeleGaze is used as a natural means of HRI in developing and designing a rather realistic PF system.

B.5 The Integration of TeleGaze into Person-Following

A special version of the TeleGaze interface has been designed for mobile robot PF. This version of the TeleGaze interface is based on the native TeleGaze, which uses only inputs from human eyes and not any other input devices. One of the major modifications to this version of the interface is having two different modes of operations, which are the TeleGaze mode and the PF mode. The TeleGaze mode is a conventional TeleGaze interface that enables the operator to interact with the robot using inputs from the eyes. The PF mode however, enables the operator to operate the robot in a PF mode. Once switched to the PF mode the operator is enabled to switch back to the TeleGaze mode using inputs from her eyes. The layout of this version of the interface in the TeleGaze mode is illustrated in Figure B.2 where a dedicated action region to change the mode of operation can be seen.

The TeleGaze mode, which is one of the two operation modes of TeleGaze, enables teleoperation through human eye gaze. The PF mode, however, enables the operator to change from a teleoperated mode to an automated PF mode. This mode, based on the principle of understanding the operator's intentions through eye movement data, enables the operator to select the POI by gazing at him/her for a certain period of time. Gazing at a person in the scene of the robot implicitly indicates that the operator is interested in following that person. This is a natural and intuitive implementation of registering the POI (task one) in the PF system.

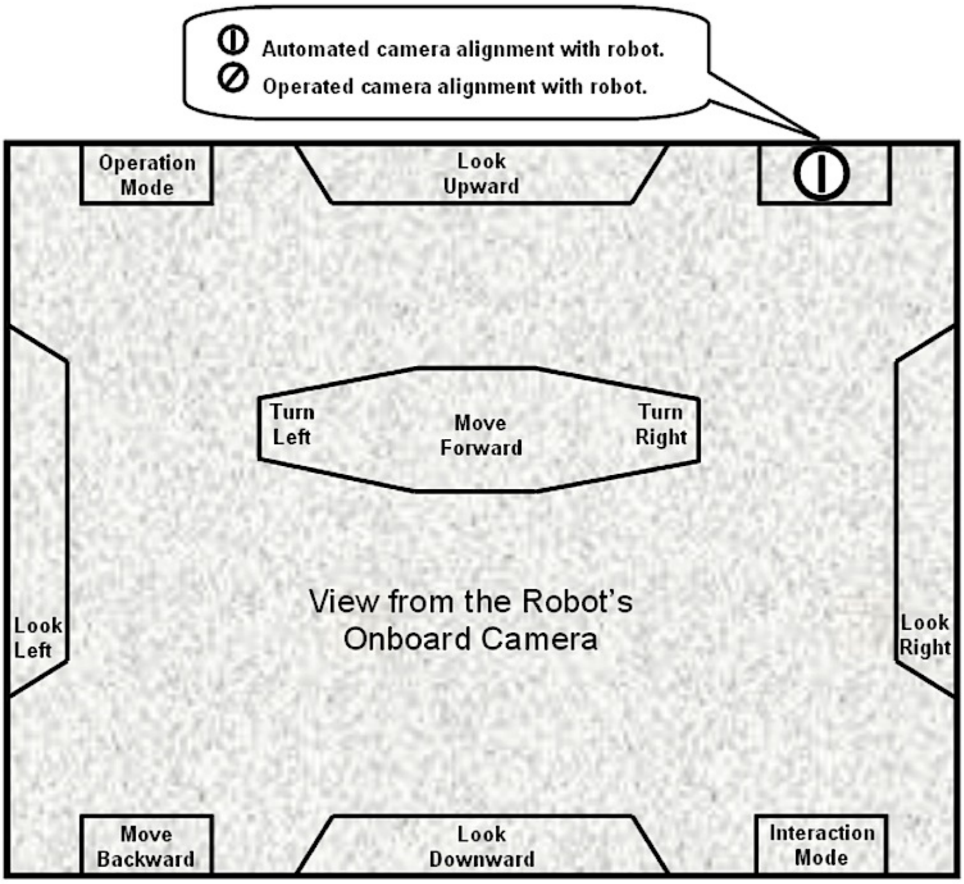


Figure B.2: The layout of this version of the interface in the TeleGaze mode.

Once the POI is registered in the system, the system informs the operator by drawing a box surrounding the POI in the scene. When this task is completed, then the system starts tracking and following this person (tasks two, three, four, and five). The dependent functionality of the system based on the interaction and operation modes via the TeleGaze interface is believed to achieve one of the basic principles of natural HRI which is implicit changes in interaction modes [122]. PF mode is shown in Figure B.3.

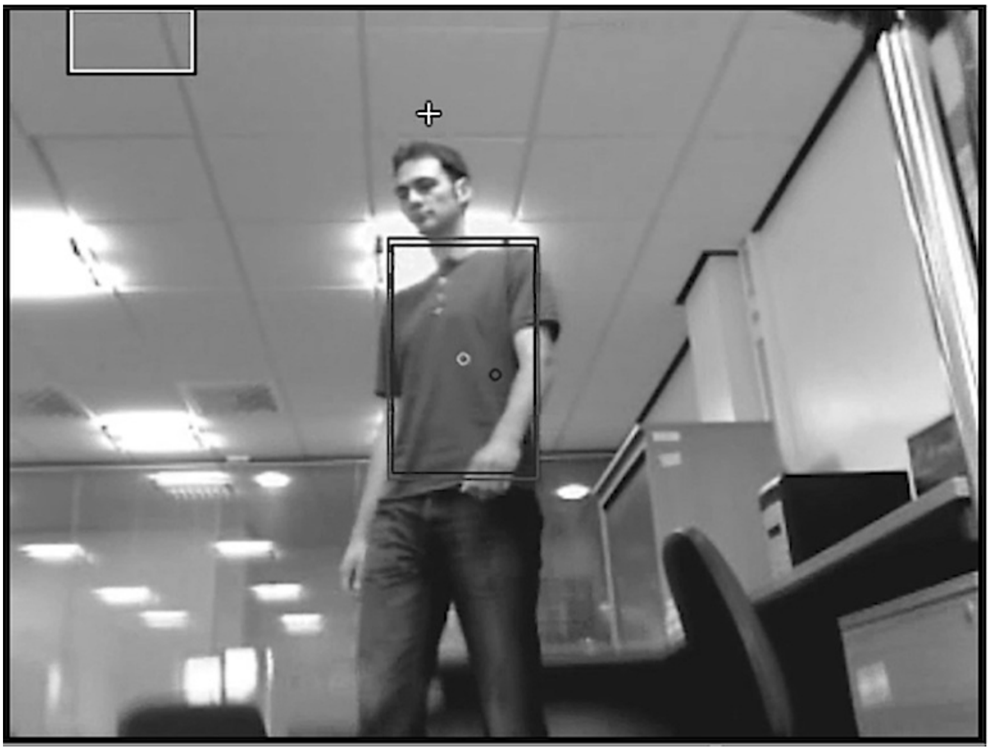


Figure B.3: A snapshot of the interface in the PF mode.

The only action region available in the PF mode is for the operator to gain back control over the robot. To do this, all that required is gazing at the action region which changes the operation mode back to the TeleGaze mode where the operator can control the robot. In other words, stop following and stop tracking the POI (tasks six and seven) and deregistering the POI (task eight). However, during the course of PF, if the robot lost the POI for any reason, it keeps looking for him/her for a period of time. If the POI was found, then it starts following him/her again (tasks two, three, four and five). If the robot failed to find the POI, then it switches back to the TeleGaze mode where the operator teleoperates the robot and the POI gets deregistered (task eight). During the course of PF if the POI is lost, the robot keeps the registration of the lost person as the POI unless the operator intervenes and changes back to the TeleGaze mode or selects a different person to be the POI.

B.6 Conclusions

Based on the presented PF taxonomy, it can be concluded that the problem space of PF is not limited to one tracking algorithm or a set of following actions. There are a number of other tasks that need to be addressed as much as these two. Therefore, this novel taxonomy of PF for mobile robots has been presented. The taxonomy shows a number of different tasks that are involved in researching any PF application. Furthermore, implementing these tasks need to be done in a natural and intuitive way in order to achieve natural HRI.

The individual tasks in the PF taxonomy depend on the interaction scenarios. Not all the tasks presented in the taxonomy might be invoked in all PF applications. However, the PF system needs to be developed so that it is capable of dealing with different tasks in the taxonomy and in different interaction scenarios. To achieve this aim, TeleGaze is integrated into a PF system. TeleGaze enables natural HRI and enables a robotic agent to understand the intentions of its human partner. The integration of TeleGaze to the PF application presented also shows an intuitive form of information acquisition for HRI applications in real life scenarios.

TeleGaze provided a goal-directed navigation of mobile robots in this implementation. Selecting the POI through gazing and switching to an automated PF capability is achieved in a natural form of HRI. In this implementation, a moving object, which is the POI, has been selected as the goal. However, this can be extended to include a static object, such as a painting on a wall, which is visible in to both the robot and the human operator. Researching this implementation of TeleGaze is believed to be of great interest to the research community and HRI applications. Therefore, it is recommended in the directions for future works on TeleGaze (Chapter 7, Section 7.3.2).

Bibliography

[1] Olsen, D. R. and Goodrich, M. A., "Metrics for evaluating human-robot interactions", *In Proceedings of Performance Measures for Intelligent Systems (PerMIS '03)*, 2003

[2] Galindo, C., Gonzalez, J., and Fernandez-Madrigal, J., "Control architecture for human-robot integration: Application to a robotic wheelchair", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol36(5), pp. 1053-67, 2006

[3] Sheridan, T. B., *Telerobotics, Automation and Human Supervisory Control*, MIT Press Ltd., Cambridge, Mass., 1992

[4] Yanco, H. A., Keyes, B., Drury, J. L., Nielsen, C. W., Few, D. A., and Bruemmer, D. J., "Evolving interface design for robot search tasks", *Journal of Field Robotics*, Vol24(8-9), pp. 779-99, 2007

[5] Farkhatdinov, I., Ryu, J., and Poduraev, J., "A user study of command strategies for mobile robot teleoperation", *Intelligent Service Robotics*, Vol2(2), pp. 95-104, 2009

[6] Fong, T. and Thorpe, C., "Vehicle teleoperation interfaces", *Autonomous Robots*, Vol11(1), pp. 9-18, 2001

[7] Duchowski, A. T., "A breadth-first survey of eye-tracking applications", *Behavior Research Methods, Instruments, and Computers*, Vol34(4), pp. 455-70, 2002

[8] Colombo, C., Andronico, S., and Dario, P., "Prototype of a vision-based gaze-driven man-machine interface", *In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*, pp. 188-92, 1995

[9] Morimoto, C. H., Koons, D., Amit, A., Flichner, M., and Zhai, S., "Keeping an eye for HCI", *In Proceedings of XII Brazilian Symposium on Computer Graphics and Image Processing*, pp. 171-6, 1999

[10] Maat, L. and Pantic, M., "Gaze-X: adaptive, affective, multimodal interface for single-user office scenarios", *In Proceedings of 20th International Joint Conference on Artificial Intelligence (IJCAI 2007), Workshop on Artificial Intelligence for Human Computing*, pp. 251-71, 2007

[11] Drewes, H. and Schmidt, A., "Interacting with the computer using gaze gestures", *In Proceedings of 11th IFIP TC 13 International Conference on Human-Computer Interaction*, pp. 511-24, 2007

[12] Bhuiyan, M. A., Ampornaramveth, V., Muto, S., and Ueno, H., "On tracking of eye for human-robot interface", *International Journal of Robotics and Automation*, Vol19(1), pp. 42-54, 2004

[13] Decher, D. and Piepmeier, J., "Gaze tracking interface for robotic control", *In Proceedings of 40th Southeastern Symposium on System Theory (SSST),* pp. 274-8, 2008

[14] Yonezawa, T., Yamazoe, H., Utsumi, A., and Abe, S., "Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking", *In Proceedings of 9th International Conference on Multimodal Interfaces,* pp. 140-5, 2007

[15] Stoyanov, D., Mylonas, G. P., and Yang, G., "Gaze-contingent 3D control for focused energy ablation in robotic assisted surgery", *In Proceedings of 11th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 347-55, 2008

[16] Istance, H., "Communication through eye-gaze: Where we have been, where we are now and where we can go from here", *In Proceedings of Symposium on Eye Tracking Research and Applications (ETRA 2005)*, pp. 9, 2005

[17] Hughes, S. B. and Lewis, M., "Task-driven camera operations for robotic exploration", *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, Vol35(4), pp. 513-522, 2005

[18] Mingfu, L., Shiqi, L., and Di, Z., "A semi-autonomous teleoperation system based on robotic hand-eye coordination", *In Proceedings of IEEE International Conference on Mechatronics and Automation (ICMA '08)*, pp. 279-84, 2008

[19] Benyon, D., Turner, P., and Tuner, S., *"Designing interactive systems: people, activities, contexts, technologies*", Pearson Education Limited, Essex, UK, 2005

[20] Jarvis, R., "A go where you look tele-autonomous rough terrain mobile robot", *Springer Tracts in Advanced Robotics: Experimental Robotics VIII*, Vol5, pp. 624-33, 2003

[21] Amai, W., Fahrenholtz, J., and Leger, C., "Hands-free operation of a small mobile robot", *Autonomous Robots*, Vol11(1), pp. 69-76, 2001

[22] Mollenbaeh, E., Stefansson, T., and Hansen, J. P., "All eyes on the monitor: gaze based interaction in zoomable, multi-scaled information-spaces", *In Proceedings of 13th International Conference on Intelligent User Interfaces (IUI '08)*, pp. 373-6, 2008

[23] Tall, M., Hansen, J. P., Alapetite, A., Hansen, D. W., Aqustin, J. S., Mllenbach, E., and Skovsqaard, H. T., "Gaze-controlled driving", *In Proceedings of 27th International Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 4387-92, 2009

[24] Barcelos, T. S. and Morimoto, C. H., "GlnX: gaze based interface extentions", *In Proceedings of Eye Tracking Research & Applications (ETRA '08)*, pp. 86-7, 2008

[25] Koh, D. H., Gowda, S. A. M., and Komogortsev, O. V., "Input evaluation of an eye-gaze-guided interface: kalman filter vs. velocity threshold eye movement identification", *In Proceedings of 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, pp. 197-202, 2009

[26] Duchowski, A. T., *"Eye tracking methodology: theory and practice*", 2nd, Springer, London, UK, 2007

[27] Nehmzow, U., *"Mobile Robotics: A Practical Introduction*", 2nd, Springer, London, UK, 2003

[28] Alvarez-Cortes, V., Zayas-Perez, B. E., Zarate-Silva, V. H., and Uresti, J. A. R., "Current trends in adaptive user interfaces: challenges and applications", *In Proceedings of 2nd Electronics, Robotics and Automotive Mechanics Conference*, pp. 297-302, 2007

[29] WowWee Robotics, "Roboraptor", *http://www.wowwee.com/en/products/toys/robots/robotics/robocreatures/roboraptor* [Last Checked on 18-04-2010]

[30] iRobot Corporation, "Government and Industrial Robots", *http://www.irobot.com/sp.cfm?pageid=109* [Last Checked on 18-04-2010]

[31] Mobile Robots Inc., "Robots for Researchers and Universities", *http://www.mobilerobots.com/Mobile_Robots.aspx* [Last Checked on 18-04-2010]

[32] Dix, A., Finlay, J., Abowd, G. D., and Beale, R., *"Human-Computer Interaction*", 3rd, Prentice Hall, Harlow, UK, 2004

[33] Tsui, C. S. L., Jia, P., Gan, J. Q., Hu, H., and Yuan, K., "EMG-based hands-free wheelchair control with EOG attention shift detection", *In Proceedings of IEEE International Conference on Robotics and Biomimetics (ROBIO '07)*, pp. 1266-71, 2007

[34] Chella, A., Pagello, E., Menegatti, E., Sorbello, R., Anzalone, S. M., Cinquegrani, F., Tonin, L., Piccione, F., Prifitis, K., Blanda, C., Buttita, E., and Tranchina, E., "A bci teleoperated museum robotic guide", *In Proceedings of International Conference on Complex, Intelligent and Software Intensive Systems (CISIS '09)*, pp. 783-8, 2009

[35] Abdollahian, G., Pizlo, Z., and Delp, E. J., "A study on the effect of camera motion on human visual attention", *In Proceedings of 15th IEEE International Conference on Image Processing (ICIP '08)*, pp. 693-6, 2008

[36] Goodrich, M. A. and Schultz, A. C., "Human-robot interaction: A survey", *Foundations and Trends in Human-Computer Interaction*, Vol1(3), pp. 203-75, 2007

[37] Wei, L., Hu, H., and Yuan, K., "Use of forehead bio-signals for controlling an intelligent wheelchair", *In Proceedings of IEEE International Conference on Robotics and Biomimetics (ROBIO '08)*, pp. 108-113, 2008

[38] Zhai, S., Morimoto, C., and Ihde, S., "Manual and gaze input cascaded (MAGIC) pointing", *In Proceedings of Conference on Human Factors in Computing Systems (CHI '99)*, pp. 246-53, 1999

[39] Dautenhahn, K., "Methodology and themes of human-robot interaction: A growing research field", *International Journal of Advanced Robotic Systems*,Vol 4, pp. 103-8, 2007

[40] Jacob, R. and Karn, K., "Eye tracking in human-computer interaction and usability research: ready to deliver the promises", *The Mind's Eye: Cognitive And Applied Aspects of Eye Movement Research*, pp.573, 2003

[41] Majaranta, P. and Raiha, K., "Twenty years of eye typing: systems and design issues", *In Proceedings of Eye Tracking Research and Applications*, pp. 15-22, 2002

[42] Noonan, D. p., Mylonas, G. P., Darzi, A., and Yang, G., "Gaze contingent articulated robot control for robot assisted minimally invasive surgery", *In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '08)*, pp. 1186-91, 2008

[43] Purwanto, D., Mardiyanto, R., and Arai, K., "Electric wheelchair control with gaze direction and eye blinking", *Artificial Life and Robotics*, Vol14(3), pp. 397-400, 2009

[44] Jacob, R. J. K., "What you look at is what you get: eye movement-based interaction techniques", *In Proceedings of SIGCHI Bulletin*, pp. 11-18, 1990

[45] Barea, R., Boquete, L., Mazo, M., and Lopez, E., "System for assisted mobility using eye movements based on electrooculography", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol10(4), pp. 209-18, 2002

[46] Barea, R., Boquete, L., Mazo, M., and Lopez, E., "Wheelchair guidance strategies using EOG", *Journal of Intelligent and Robotic Systems: Theory and Applications*, Vol34(3), pp. 279-99, 2002

[47] Shamoto, M., Kato, K., and Yamamoto, K., "An implementation of humanoid vision - analysis of eye movement and implementation to robot", *In Proceedings of 46th SICE Annual Conference (SICE '07)*, pp. 744-7, 2007

[48] Green, S. A., Billinghurst, M., Chen, X., and Chase, J. G., "Human-robot collaboration: A literature review and augmented reality approach in design", *International Journal of Advanced Robotic Systems*, Vol5(1), pp. 1-18, 2008

[49] Kuo, C., Chan, Y., Chou, H., and Siao, J., "Eyeglasses based electrooculography human-wheelchair interface", *In Proceedings of IEEE International Conference on Systems, Man and Cybernetics (SMC'09)*, pp. 4746-51, 2009

[50] Yathunanthan, S., Chandrasena, L. U. R., Umakanthan, A., Vasuki, V., and Munasinghe, S. R., "Controlling a wheelchair by use of EOG signal", *In Proceedings of 4th International Conference on Information and Automation for Sustainability. Sustainable Development through Effective Man-Machine Co-existence (ICIAFS)*, pp. 283-8, 2008

[51] Lin, C., Ho, C., Chen, W., Chiu, C., and Yeh, M., "Powered wheelchair controlled by eye-tracking system", *Optica Applicata*, Vol36(2-3), pp. 401-12, 2006

[52] Bartolein, C., Wagner, A., Jipp, M., and Badredding, E., "Easing wheelchair control by gaze - based estimation of intended motion", *IFAC Proceedings Volumes*, Vol17(1), 2008

[53] Jipp, M., Bartolein, C., and Badreddin, E., "The impact of individual differences on human information acquisition behavior to enhance gaze-based wheelchair control", *In Proceedings of IEEE International Conference on Systems, Man and Cybernetics (SMC '08)*, pp. 2591-2596, 2008

[54] Bien, Z., Chung, M., Chang, P., Kwon, D., Kim, D., Han, J., Kim, J., Kim, D., Park, H., Kang, S., Lee, K., and Lim, S., "Integration of a rehabilitation robotic system (KARES II) with human-friendly main-machine interactions units", *Autonomous Robots*, Vol16(2), pp. 165-91, 2004

[55] Yonezawa, T., Yamazoe, H., Utsumi, A., and Abe, S., "GazeRoboard: gaze-communicative guide system in daily life on stuffed-toy robot with interactive display board", *In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1204-9, 2008

[56] Mariottini, G. L., Prattichizzo, D., De Biasi, M., Rufa, A., De Capua, A., and Rossi, S.,

"Human-robotics interface for the interaction with cognitive and emotional human domains", *In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '07)*, pp. 528-33, 2007

[57] Atienza, R. and Zelinsky, A., "Intuitive human-robot interaction through active 3D gaze tracking", *In Proceedings of The Eleventh International Symposium on Robotics Research*, pp. 172-8, 2005

[58] Atienza, R. and Zelinsky, A., "Intuitive interface through active 3D gaze tracking", *In Proceedings of International Conference on Active Media Technology (AMT 2005)*, pp. 16-21, 2005

[59] Atienza, R. and Zelinsky, A., "Active gaze tracking for human-robot interaction", *In Proceedings of 4th IEEE International Conference on Multimodal Interfaces*, pp. 261-6, 2002

[60] Schneider, E., Kohlbecher, S., Villgrattner, T., Bartl, K., Bardins, S., Poitschke, T., Ulbrich, H., and Brandt, T., "Vision system for wearable and robotic uses", *In Proceedings of 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 53-8, 2008

[61] Akashi, T., Wakasa, Y., Tanaka, K., Karungaru S., and Fukumi, M., "Interactive interface with evolutionary eye sensing and physiological knowledge", *IEEJ Transactions on Electronics, Information and Systems*, Vol129(7), pp. 1288-95, 2009

[62] Pinpin, L. K., Johansson, R. S., Laschi, C., and Dario, P., "Gaze interface: utilizing human predictive gaze movements for controlling a HBS", *In Proceedings of 2nd Biennial IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob '08)*, pp. 158-62, 2008

[63] Islam, M. E., Begum, N., and Bhuiyan, M. A., "Vision system for human-robot interface", *In Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008)*, pp. 415-19, 2008

[64] Yoo, D. H., Kim, J. H., Kim, D. H., and Chung, M. J., "A human-robot interface using vision-based eye gaze estimation system", *In Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS 2002)*, pp. 1196-1201, 2002

[65] Shirakura, N., Morita, M., and Takeno, J., "Development of a human interface for remote-controlled robots using an eye-tracking system", *In Proceedings of IEEE International Conference on Mechatronics and Automation*, pp. 351-6, 2005

[66] Jarvis, R., "Multimodal robot/human interaction in an assistive technology context", *In Proceedings of Second International Conference on Advances in Computer-Human Interaction (ACHI '09)*, pp. 212-18, 2009

[67] Li, Z. and Jarvis, R., "A multi-modal gesture recognition system in a human-robot interaction scenario", *In Proceedings of IEEE International Workshop on Robotic and Sensors Environments (ROSE '09)*, pp. 41-6, 2009

[68] Maeda, Y., Suzuki, S., Igarashi, H., and Hidaka, K., "Evaluation of human skill in teleoperation system", *In Proceedings of SICE-ICASE International Joint Conference*, pp. 5339-44, 2006

[69] Roca, X., Vitria, J., Vanrell, M., and Villanueva, J. J., "Gaze control in a binocular robot systems", *In Proceedings of 7th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA '99)*, pp. 479-85, 1999

[70] Castellini, C. and Sandini, G., "Gaze tracking for robotic control in intelligent teleoperation and prosthetics", *In Proceedings of The 2nd Conference on Communication by Gaze Interaction (COGAIN 2006)*, pp. 73-7, 2006

[71] Castellini, C., "Gaze tracking in semi-autonomous grasping", *Journal of Eye Movement Research*, Vol2(4), pp. 1-7, 2008

[72] Nagamatsu, T., Kamahara, J., and Tanaka, N., "3D gaze tracking with easy calibration using stereo cameras for robot and human communication", *In Proceedings of The 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 59-64, 2008

[73] Sibert, L. E. and Jacob, R. J. K., "Evaluation of eye gaze interaction", *In Proceedings of Conference on Human Factors in Computing Systems (CHI 2000)*, pp. 281-8, 2000

[74] Coltekin, A., Heil, B., Garlandini, S., and Fabrikant, S., "Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis", *Cartography and Geographic Information Science*, Vol36(1), pp. 5-17, 2009

[75] Drury, J., Riek, L. D., Christiansen, A. D., Eyler-Walker, Z. T., Maggi, A. J., and Smith, D. B., "Command and control of robot teams", *In Proceedings of International Conference of the Association of Unmanned Vehicle System*, pp. 15, 2003

[76] Nielsen, J., "Enhancing the explanatory power of usability heuristics", *In Proceedings of ACM Conference on Human Factors in Computing Systems*, pp. 152-8, 1994

[77] Ravden, S. and Johnson, G., *"Evaluating Usability of Human-Computer Interfaces: A Practical Approach*", Ellis Horwood Books, London, UK, 1989

[78] Lin, Y., Zhang, W. J., Wu, C., Yang, G., and Dy, J., "A fuzzy logics clustering approach to computing human attention allocation using eye gaze movement cue", *International Journal of Human-Computer Studies*, Vol67(5), pp. 455-63, 2009

[79] Salvucci, D. D. and Anderson, J. R., "Intelligent gaze-added interfaces", *In Proceedings of Conference on Human Factors in Computing Systems (CHI 2000)*, pp. 273-80, 2000

[80] Xu, T., Kuhnlenz, K., and Buss, M., "Information-based gaze control adaptation to scene context for mobile robots", *In Proceedings of 19th International Conference on Pattern Recognition (ICPR '08)*, 4 pp., 2008

[81] Witus, G., Hunt, S., and Ellis, R. D., "Experiments in augmented teleoperation for mobile robots-I", *In Proceedings of The International Society for Optical Engineering (SPIE). Unmanned Systems Technology IX*, pp. 65610R-1-10, 2007

[82] Yi, J. C. , "User-research driven mobile user interface innovation: a success story from Seoul", *Interactions*, VolXVII(1), 2010

[83] Bailey, B. P. and Iqbal, S. T., "Understanding changes in mental workload during

execution of goal-directed tasks and its application for interruption management*", ACM Transactions on Computer-Human Interaction*, Vol14(4), pp. 21-8, 2007

[84] Komogortsev, O. V., Mueller, C. J., Tamir, D., and Feldman, L., "An effort based model of software usability*", In Proceedings of International Conference on Software Engineering and Practice (SETP '09)*, pp. 75-83, 2009

[85] Mitsunaga, N., Smith, C., Kanda, T., Ishiguro, H., and Hagita, N., "Adapting robot behavior for human-robot interaction*", IEEE Transactions on Robotics*, Vol24(4), pp. 911-16, 2008

[86] Andy Field, *"Discovering Statistics using SPSS*", 3rd, SAGE, London, UK, 2009

[87] Zhang, X. and MacKenzie, I. S., "Evaluating eye tracking with ISO 9241 - part 9*", In Proceedings of 12th International Conference on HCI Intelligent Multimodal Interaction Environments*, pp. 779-88, 2007

[88] Tamir, D., Komogortsev, O. V., and Muller, C. J., "An effort and time based measure of usability*", In Proceedings of 30th International Conference on Software Engineering (ICSE'08)*, pp. 47-52, 2008

[89] Mitsunaga, N., Smith, C., Kanda, T., Ishiguro, H., and Hagita, N., "Robot behavior adaptation for human-robot interaction based on policy gradient reinforcement learning*", In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 218-25, 2005

[90] Steinfeld, A, Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., and Goodrich, M., "Common metrics for human-robot interaction*", In Proceedings of 2006 ACM Conference on Human-Robot Interaction*, pp. 33-40, 2006

[91] Brooke, J., "SUS: A quick and dirty usability scale*", In Proceedings of Usability evaluation in industry*, 1996

[92] Nehmzow, U., "*Robot Behaviour: Design, Description, Analysis and Modelling*", Springer, London, UK, 2009

[93] Bartels, M. and Marshall, S. P., "Eye tracking insights into cognitive modelling*", In Proceedings of Symposium on Eye Tracking Research and Applications (ETRA 2006),* pp. 141-7, 2006

[94] Crowe, E. C. and Narayanan, N. H., "Comparing interfaces based on what users watch and do*", In Proceedings of Eye Tracking Research and Applications Symposium*, pp. 29-36, 2000

[95] Yanco, H. A., Drury, J. L., and Scholtz, J., "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition*", In Proceedings of Human-Computer Interaction*, Vol19(1-2), pp. 117-49, 2004

[96] Kuzuoka, H., Yamazaki, K., Yamazaki, A., Kosaka, J., Suga, Y., and Heath, C., "Dual ecologies of robot as communication media: Thoughts on coordinating orientations and projectability*", In Proceedings of Conferences on Human Factors in Computing Systems (CHI 2004)*, pp. 183-90, 2004

[97] Xie, M., *"Fundamentals of Robotics: Linking Perception to Action"*, World Scientific Publishing Co. Pte. Ltd., Toh Tucj Link, Singapore, 2004

[98] Scholtz, J., "Theory and evaluation of human-robot interactions*", In Proceedings of 36th Hawaii International Conference on Systems Sciences*, pp. 125-35, 2003

[99] Adams, J. A., "Multiple robot/ single human interaction: Effects on perceived workload*", Behaviour and Information Technology*, Vol28(2), pp. 183-98, 2009

[100] Hong, S., Setiawan, N. A., and Lee, C., "Multiple people gesture recognition for human-robot interaction*", In Proceedings of 12th International Conference on HCI Intelligent Multimodal Interaction Environments*, pp. 625-33, 2007

[101] Olivares, R., Zhou, C., Bodenheimer, B., and Adams, J. A., "Interface evaluation for mobile robot teleoperation*", In Proceedings of ACM Southeast Conference (ACMSE '03)*, pp. 112-18, 2003

[102] Johnson, C. A., Adams, J. A., and Kawamura, K., "Evaluation of an enhanced human-robot interface*", In Proceedings of IEEE International Conference on Systems, Man and Cybernetics (SMC '03)*, pp. 900-5, 2003

[103] King, L., "The relationship between scene and eye movements*", In Proceedings of 35th Annual Hawaii International Conference on System Sciences*, pp. 1829-37, 2002

[104] Seagull, F. J., and Xiao, Y., "Using eye-tracking video data to augment knowledge elicitation in cognitive task analysis*", In Proceedings of 45th Annual Meeting of Human Factors and Ergonomics Society*, pp. 400-3, 2001

[105] Jacob, R. J. K., "eye movement-based human-computer interaction techniques: toward non-command interfaces*", Advances in Human-Computer Interaction*, pp. 151-90, 1993

[106] Tan, Y. K., *"Eye gaze tracking and speech recognition for data entry and error recovery: a multimodal approach"*, PhD Thesis, Nottingham Trent University, 2003

[107] Vilimek, R. and Zander, T. O., "BC(eye): Combining eye-gaze input with brain-computer interaction*", In Proceedings of 5th International Conference on Universal Access in Human-Computer Interaction: Intelligent and Ubiquitous Interaction Environments*, pp. 593-602, 2009

[108] Ramp, E., Bra, P. D., and Brusilovsky, P., "High-level translation of adaptive hypermedia applications*", In Proceedings of 16th ACM Conference on Hypertext and Hypermedia*, pp. 126-8, 2005

[109] Law, B., Lomax, A. J., Atkins, M. S., Machenzie, C. L., and Kirkpartrick, A. E., "Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment*", In Proceedings of Symposium on Eye Tracking Research and Applications (ETRA 2004)*, pp. 41-7, 2004

[110] Itoh, K., Tanaka, H., and Seki, M., "Eye-movement analysis of track monitoring patterns of night train operators: effects of geographic knowledge and fatigue*", In Proceedings of XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association*, pp. 360-3, 2000

[111] Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., and Crundall, D., "Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers*", Ergonomics*, Vol46(6), pp. 629-46, 2003

[112] Underwood, G., Humphrey, K., and Foulsham, T., "Knowledge-based patterns of remembering: Eye movement scanpaths reflect domain experience*", In Proceedings of 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society*, pp. 125-44, 2008

[113] Puolamaki, K., Ajanki, A., and Kaski, S., "Learning to learn implicit queries from gaze patterns*", In Proceedings of 25th International Conference on Machine Learning*, pp. 760-7, 2008

[114] Conati, C., Merten, C., Amershi, S., and Muldner, K., "Using eye-tracking data for high-level user modeling in adaptive interfaces*", In Proceedings of 22nd AAAI Conference on Artificial Intelligence and the 19th Innovative Applications of Artificial Intelligence Conference*, pp. 1614-17, 2007

[115] Takemura, H., Ito, K., and Mizoguchi, H., "Person following mobile robot under varying illumination based on distance and color information*", In Proceedings of IEEE International Conference on Robotics and Biomimetics (ROBIO '07)*, pp. 1500-5, 2007

[116] Hu, C., Ma, X., and Dai, X., "A robust person tracking and following approach for mobile robot *", In Proceedings of IEEE International Conference on Mechatronics and Automation (ICMA '07)*, pp. 3571-6, 2007

[117] Hyukseong, K., Youngrock, Y., Jae, B. P., and Kak, A. C., "Person tracking with a mobile robot using two uncalibrated independently moving cameras*", In Proceedings of IEEE International Conference on Robotics and Automation*, pp. 2877-83, 2005

[118] Tsalatsanis, A., Valavanis, K., and Yalcin, A., "Vision based target tracking and collision avoidance for mobile robots*", Journal of Intelligent and Robotic Systems: Theory and Applications*, Vol48(1), pp. 285-304, 2007

[119] Bernardin, K., Gehrig, T., and Stiefelhagen, R., "Multi-level particle filter fusion of features and cues for audio-visual person tracking*", In Proceedings of 2nd Annual Classification of Events Activities and Relationships (CLEAR '07) and Rich Transcription (RT '07)*, pp. 70-81, 2007

[120] Carelli, R., Forte, G., Canali, L., Mut, V., Araguas, G., and Detefanis, E., "Autonomous and teleoperation control of a mobile robot*", Mechatronics*, Vol18(1), pp. 187-94, 2008

[121] Treptow, A., Cielniak, G., and Duckett, T., "Active people recognition using thermal and grey images on a mobile security robot*", In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2103-8, 2005

[122] Goodrich, M. A. and Olsen, D. R., "Seven principles of efficient human robot interaction*", In Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, pp. 3942-8, 2003

Note: The total word count for this thesis is 67934 words.