MobiBot: Personifying Telepresence Communication

By

Banan S. Bamoallem

A thesis submitted to the University of Strathclyde

for the degree of

Doctor of Philosophy

Department of Design, Manufacture and Engineering Management

University of Strathclyde

Glasgow, UK

May 2017

Abstract

The smartphone has become an intrinsic part of daily life, taking the role of a trusted companion in the context of communication technology. A persistent and widely documented issue in the domain of embodied technology, however, is the lack of natural interaction. As communication takes place not only through speech, but also through gestures such as facial expressions, gaze, head movements, hand movements and body posture. This research believe these are needed to fully support non-verbal communication and make interactions more engaging and efficient. In this research, This research focus on a telepresence (TP) robotic system (MobiBot) that affords the ability to convey non-verbal behaviours such as gesture and posture that can make the interactions more natural and life-like. Our expletory study focused specifically on the head rather than any other body part as it is a rich source of information for speech-related movement. This investigated the value of incorporating head movements into the use of telepresence robots as communication platforms by means of evaluating a system that manually reproduces head movement as closely as possible. Then, expanding the consideration of the physical embodiment of the system to include the head, shoulders and proximity, this research proposes a new protocol for the translation of the vocal stream into gesture to generate human-like behaviour and support more natural interaction within the embodied technology system. A modified version of the Undefined Technology of Acceptance and Use of Technology model (UTAUT) has been used to explore social and cognitive experience when using the system. Subjects' acceptance of the gesturally-supported video communication using for the MobiBot system was examined by comparing of the different methods of interaction on video calls using the MobiBot TP system. The comparison was between mimicking movement (where the operator replicates the movement by pressing buttons) and automated movement triggered by the user vocal stream. This was carried out in order to evaluate their effect. The results of the comparative analysis indicated that the mimicking interaction of the MobiBot system for video calls was preferred by the users over the vocal-triggered automatic interaction movement method. Evaluation and feedback of the movements incorporated suggests a mix of both vocal-triggered automatic and mimicking movements, using fewer large movements and more small and steady movements, is optimal. In addition, a set of guidelines was developed using the findings from both studied, for ‘personifying’ telepresence conversations and development of such systems. This research, in general, demonstrated significantly greater benefits from incorporating movement with such systems.

Acknowledgements

My profound gratitude goes to Allah, who made all things possible and whose continuous blessing was on me.

Many people have helped over the last weeks and months and years to get me and this project completed. I could not have completed this research without them. They have made valuable suggestions, which I have incorporated, and I would like to take this opportunity to express my gratitude to them.

I would like to express my deep gratitude to my academic supervisor, Dr Andrew Wodehouse, and Mr Gordon Mair for their guidance, suggestions and invaluable encouragement for my project and its development

I am awed by the help and patience by Dino Bertolaccini while working on programming MobiBot over the last couple of years. I would like also to extend my gratitude to my colleague Wenbin Zhong for participating in programming MobiBot.

I must also acknowledge the Ministry of Higher Education and Umm AlQura University in Saudi Arabia for sponsoring me and providing me with this opportunity to study for a PhD.

I thank the Saudi Arabia Cultural Bureau in the United Kingdom and North Ireland, for their support and valuable advice during my study.

I am so grateful to my great to my family and friends, words alone cannot express what I owe them for their unconditional love, encouragement and support, which enabled me to complete this project.

Finally, my great love, gratefulness and appreciation will always go to my beloved mom and dad. I am sorry for being away from you. Thank you for the generosity of your souls.

Dedication

I dedicate this thesis to my beloved family

Declaration of authenticity and author's rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.'

Signed: Date:

Banan Bamoallem

Published Work

1. The role of the vocal stream in telepresence communication

Bamoallem, B. S. & Wodehouse, A. J. 29 Jun 2016 *PETRA'16: Proceedings of the 9th International Conference on PErvasive Technologies Related to Assistive Environments.* New York, 4 p.

Research output: Conference proceeding › Conference contribution › Paper

2. The impact of head movements on user involvement in mediated interaction

Bamoallem, B. S., Wodehouse, A. J., Mair, G. M. & Annamalai Vasantha, G. V. 1 Feb 2016 In: Computers in Human Behaviour. 55, p. 424-431 8 p.

Research output: Contribution to journal › Article

3. Design for an optimal social presence experience when using telepresence robots

Bamoallem, B. S., Wodehouse, A. J. & Mair, G. 19 May 2014 In DS 77: Proceedings of the DESIGN 2014 13th International Design Conference. p. 653-662 10 p.

Research output: Conference proceeding › Contribution to conference › Paper

4. MobiBot: Personifying Telepresence Communication

Bamoallem, B. S., Wodehouse, A. J. In: International Journal of Human-Computer Studies

Research output: Contribution to Journal › Article (Submitted)

Abbreviations

2D	Two Dimensional
3D	Three Dimensional
3G	Third Generation
4G	Fourth Generation
ADC	Arduino Digital Converter
ATT	Attitude Towards Technology
CA	Cognitive Absorption
DC	Direct Current
DoF	Degree of Freedom
ENJ	Enjoyment
FC	Facilitating Conditions
FFT	Fast Fourier Transform
FI	Focused Immersion
HCI	Human Computer Interaction
HHI	Human-Human Interaction
HRI	Human-Robot Interaction
IIS	Interaction Involvement Scale
LCD	Liquid Crystal Display
LED	Light Emitting Diode
PA	Physical Appearance
PS	Perceived Sociability
PRoP	Personal Roving Presence
SA	Social Abilities
SP	Social Presence
SPSS	Statistical Package for Social Science
SI	Social Influence
PU	Perceived Usefulness
TAM	Technology Acceptance Model
TP	Telepresence
TD	Temporal Dissociation
TRA	Theory of Reasoned Action
UK	United Kingdom
USB	Universal Serial Bus
VMC	Video-Mediated Communication
Wi-Fi	Wireless Fidelity
WoZ	Wizard of Oz

Table of Contents

List of Figure

List of Tables

Chapter 1 INTRODUCTION

This chapter presents an overview of telepresence TP robotics system, as well as its terminology. Followed by a proposed structure a general TP robotics architecture based on the existing literature. Further, the chapter discusses the current use of this technology, besides the issues associated with current products. Finally, presents the research motivation, highlights the research aim, defines the research question, and summarises the objectives and scope of the proposed system.

1.1 Introduction

Until recently, instant messaging and email were the only ways to convey information between people at a distance. They provide users with a sense of constant connectivity and politeness, because it is easy to send an email eloquently and accurately, without bothering someone with a face-to-face or video meeting. However, a downside with this is that it does not provide a high level of information richness, where both members can influence and affect the other, which further inhibits social exchange. Recently, remote communication has become more realistic with the advancements in technological equipment for teleconferencing technology. Teleconferencing systems can support meetings with real-time feedback, due to lighting, cameras and the proximity of the user.

Teleconferencing is already addressing the rapidly growing need for developing new technology in teleconferencing systems, with emphasis on the importance of the 'being there' concept. These teleconference systems range from older methods, using video and audio teleconferencing, to new web-based conferencing applications. In addition, 2D groupware and multi-users' virtual worlds have also been used for teleconferencing (Kantonen,Woodward and Katz, 2010). Though development in technology has been rapid, disadvantages still exist in systems such as conference calls. These systems are an easy and quick way to communicate, and also easy to set up without needing hardware other than a mobile, however, they are still limited in relation to audio, and require a separate channel for document sharing and bandwidth, in addition to the expensive high-end technology which is required. In addition, relationships, whether personal or professional, rely on positive social interactions which not supported yet by these systems.

It is clear that the old sender/receiver model of communication is no longer enough, this has resulted in the emergence of telepresence robotic systems as a novel solution to meet the social communication needs.

1.2 History of Telepresence

Telepresence is a new generation of robot, making it possible for interactions to be more dynamic. Therefore, it is commonly defined as a set of technologies that allow people to feel as if they are present in a location other than their true location.

> "*the phenomenon that a human operator develops a sense of being physically present at a remote location through interaction with the system's human interface, i.e. through the user's actions and the subsequent perceptual feedback they receive via the appropriate teleportation technology*", (IJsselsteijn *et al.*, 2000, p. 1)

The idea behind this technology dates back over half a century, to when Robert Heinlein introduced the telepresence concept in his short novel "*Waldo*", published in Astounding Magazine in August 1942 (Heinlein, 1969). The novel is about a genius "*Waldo F. Jones*" with a disabling disease, who builds hardware called "*Waldoes*", which allow him to perform teleoperations. Using a glove and harness, he was able to control a much more powerful mechanical hand, simply by moving his hand and fingers, using the Waldo.

Similarly, James Blish's novel (Bridge) tells the story of a worker on a moon of Jupiter, who uses a remote controlled vehicle to travel around the planet's atmosphere (Mair, 2007).

Payne (1949) implemented the concept and designed the Slave Manipulator, which was one of the oldest projects in this field in the forties (Figure 1.1). His idea was to design a mobile manipulator to handle hazardous radioactive materials at a safe distance. Despite his contribution, the system had a very limited capability, and could only reach within the range of a few meters.

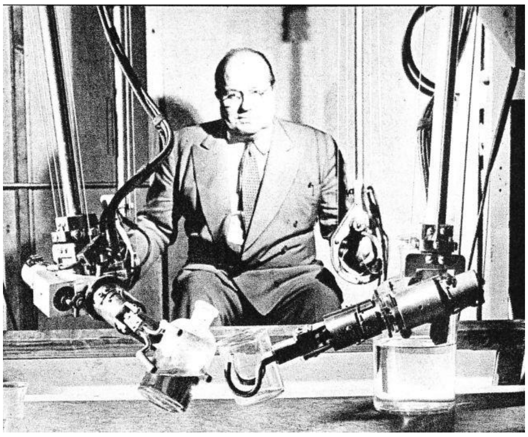


Figure 1.1. Payne's Mobile Manipulator (Courtesy of Popular Science, June 1948)

However, Marvin Minsky was the first to use the term "*Telepresence*" (TP) in the context of teleportation, on the suggestion of his friend Pat Gunkel, in a 1979 funding proposal entitled "*Toward a Remotely-Manned Energy and Production Economy*" (Minsky, 1980). In this context, telepresence refers to the manipulation of remote objects through technology that produces a sense of being physically present in a far-off location (Figure 1.2). The participants interacted with the system's human interface which received the user's action and provided subsequent perceptual feedback. Marvin Minsky explored the idea in another work published in OMNI magazine in 1980, depicting people suiting up in sensor-motor jackets in order to do their jobs remotely, where he highlighted the importance of the perception experience that resulted from this technology:

> "*Telepresence emphasises the importance of high-quality sensory feedback and suggests future instruments that will feel and work so much like our own hands that we won't notice any significant difference*." (Minsky, 1980, p. 37).

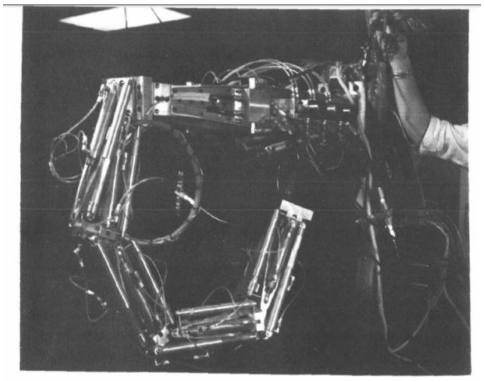


FIGURE 48.—Minsky's electrohydraulic, many-jointed manipulator arm. (Courtesy of W. M. Bennett, M.I.T.)

Figure 1.2. Minsky's Manipulator Arm (Courtesy of W. M. Bennett M.I.T)

In the following years, several telepresence projects were developed and explored the same concept. The design goal for these kinds of systems was to allow users to be physically present in a remote location when it was impossible for them to do so because of geographical distance, in order to allow users to interact with and control objects within that remote real world environment, or to access locations otherwise not available to people. Limited by the technology available at that time, these quite complex systems were mainly designed to perform a single or specific task. Frequently requiring special hardware that was both expensive to produce and solely dedicated to the task, they required skilled and highly trained individuals to operate them effectively in the remote location. Such applications ranged from remote-controlled manipulators (e.g. robot arms) to those that enabled humans to work in remote, hazardous or challenging environments, such as underwater tasks, bomb disposal, hazardous clean-ups, or for the purpose of minimally invasive surgery. Examples of this application are presented by Yoerger and Slotine (1987) for working in depths under the sea, by Codourey et al. in microscopic workstations (1997) and by Li et al. (1996) for performing space station maintenance.

1.3 Telepresence terminology

Since the early 1990s, a large number of scholarly works have been published as a result of the increasing need for development in telepresence robotic technology, which have investigated diverse disciplines such as psychology, communication, computer science and philosophy in relation to telepresence.

Lombard and Jones (2007) have identified over 1400 articles, written mostly over the previous twelve years, that have addressed telepresence. Variations in terminology have resulted from different research in the area of telepresence.

Sheridan (1992b, Thomas B. Sheridan, 1992a) defines telepresence as: "*. . . visual, kinesthetic, tactile or other sensory feedback from the teleoperator to the human operator that is sufficient and properly displayed such that the human feels that he is present at the remote site and that the teleoperator is an extension of his own body*" (p. 808). In this definition, apart from his emphasis on the importance of the connection control and sensory feedback, he defines the driver of the TP as the "*human operator*", the TP system as the "*teleoperator*" and the distant participant or environment as the "*remote site*". Based on these definitions, Schloerb (1995) in his paper, illustrated the interconnection communication system between the teleoperator and the teleoperation system (Figure 1.3). This system consists of (1) human operator, (2) operator interface, (3) teleoperator and (4) remote environment. In his article, he describes the human operator as the person who uses the teleoperation system to control and/or sense events in the remote environment. This control system consists of the operator interface and the teleoperator, with the teleoperator mainly being a machine working as a surrogate for the human operator in the remote environment. The connection between both sides (operator and teleoperator) is the operator interface, which includes a display that provides sensory input to the human operator, and controls that the human operator uses to provide information to the system. The remote environment is the real environment that is physically separated by a distance from the human operator.

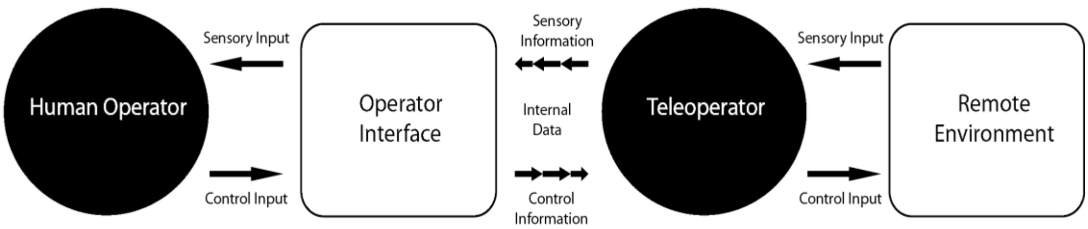


Figure 1.3. Major functional components of teleoperation by Schloerb (1995)

This system also contains the same components as those described by Ferrell and Sheridan (1967) in their definition of a remote manipulation system, as we can see in (Figure 1.4). However, they describe the teleoperator system as the manipulator. This description can also be found in Akin *et al.* (1983).

> "*At the worksite, the manipulators have the dexterity to allow the operator to perform normal human functions. At the control station, the operator receives sufficient quantity and quality of sensory feedback to provide a feeling of actual presence at the worksite*." (p. 282)

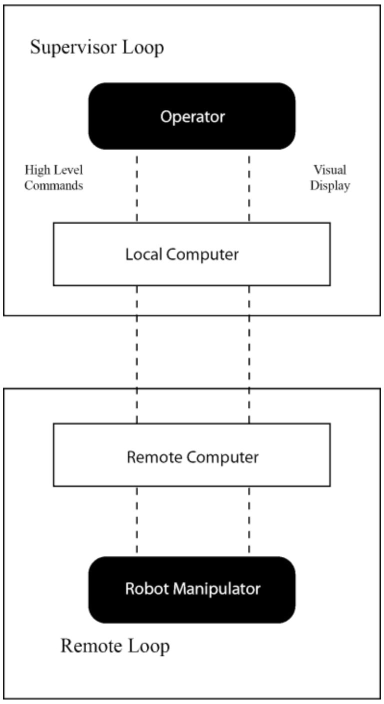


Figure 1.4. Supervisory control by Ferrell and Sheridan (1967)

Differing from the previous definitions, Kristoffersson (2013) describes the driver of the robot as the pilot and the distant environment as the local environment.

> "*The system consists of both the physical robot (sensors and actuators) and the interface used to pilot the robot. A Pilot user is a person who remotely connects to the robot via a computer interface. The pilot who is embodied in the TP system can move around in the environment where the robot is located and interact with other persons. A Local user is a user that is being situated*

at the same physical location as the robot. Local users are free to move around while interacting with the pilot user who is visiting them via the robot. Local environment is the environment in which the robot and the local user are situated" (p. 2)

Mair (1997) divided the telepresence system into three essential subsystems; the remote site system, the communication system, and the home site system. Each site consists of different equipment to control and sense the other site via a communications link between them (Figure 1.5).

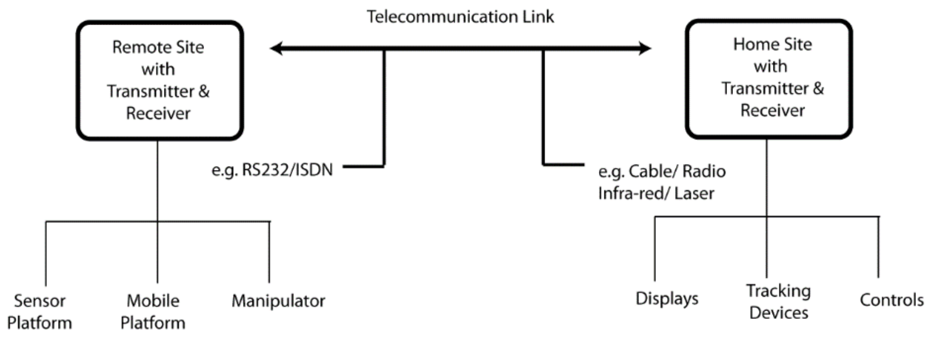


Figure 1.5. The basic elements of telepresence Mair (1997)

Cui et al. (2003) have a different view from Mair (1997), when they define the three parts as (1) A capture system to record and represent the information from the remote site, which they call the manual control; (2) a network transmission system and (3) a display system to make the local user feel as if she were somehow present in the remote scene, which is called the remote system (Figure 1.6).

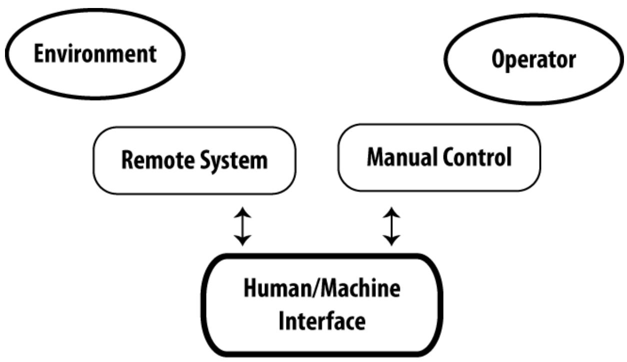


Figure 1.6. Illustration of the information flow in a teleoperation system Cui et al. (2003)

These terminologies are summarised in (Table 1.1) below. It is possible to see that there are differences between the systems, although they all have the same base architecture. This can cause confusion for the reader, as it did for us while we were engaged in our research. Based on the existing literature and as the majority of these systems have the same base architecture, we have proposed a general TP robotics architecture which might be used regardless of the nature of the telepresence or its performance (Figure 1.7). Adopting a common framework for definitions and terminology of telepresence structure will allow us to communicate and collaborate more effectively and ultimately build knowledge in this area.

Table 1.1. An overview of common telepresence terminology systems

Study	Driver Site	Driver	Distant Environment	Distant Participant	Manipulator	Control Interface
Sheridan (1992b) (1992a)	-	Human operator	Remote site	-	Teleoperator	-
Schloerb (1995)	-	Human operator	Remote environment	-	Teleoperator	Operator interface
Ferrell and Sheridan (1967)	-	Operator	Remote loop	-	Manipulator	Local Computer
Akin et al. (1983)	-	Operator	Worksite	-	Manipulator	Control station
Kristoffersson et al. (2013)	-	Pilot	Local environment	Local user	Mobile Robotic Telepresence	Computer interface
Mair (1997)	Remote site	-	Home site	-	-	Communication system
Cui et al. (2003)	Local environment	Operator/ Local user	Remote environment	-	Remote system	-

1.3.1 Proposed structure

Figure 7 identifies the two main sites of telepresence, which include four major functional components; a home site which consists of a human operator and an operator interface, and a remote site which consists of a remote human, and a telepresence or teleoperator. Both sites have computer interfaces to process the data from both sites via the communication links, which have a bandwidth capable of transmitting all of the data at a fast speed.

The human operator represents the driver (pilot) of the telepresence robot at the home site. The interface for the driver uses a set of displays, controls and sensors to provide a high-fidelity interaction between the hardware and the human, or the environment at the remote site. Also, the interface is used in order to provide continuous feedback from the telepresence robot in the remote environment to the driver.

A telepresence robot is a machine that functions as a surrogate for the human driver, and which consists of sensors and a display. The remote site includes displays which provide information inputs from sensors to control the system. The remote site is described as the real environment, or the people interacting with the robot that are physically separated from the remote operator in space (Mair, 1997, Mair, 2007, Schloerb, 1995, Thomas B Sheridan, 1992b, Thomas B. Sheridan, 1992a).

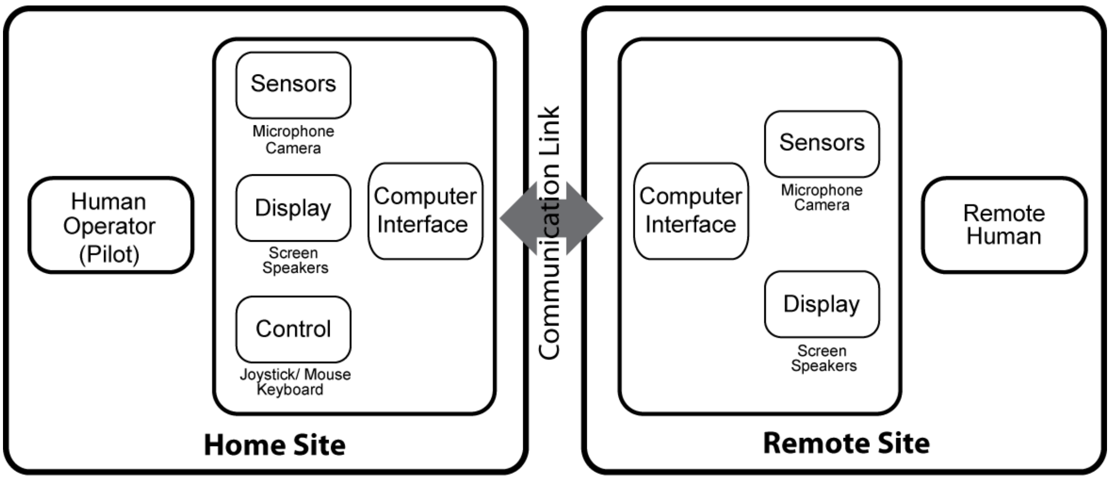


Figure 1.7. Proposed telepresence robotic structure

1.4 Current use of telepresence

As previously explained, telepresence in the context of our application means augmenting the human presence with a robot, which is operated by a human driver from a location at a distance. This basically emphasises the importance of 'being there', remotely but in a realistic manner. Smaller and faster computers, low-cost and high-resolution webcams, lightweight video screens, relatively inexpensive range sensors and nearly ubiquitous wireless communication (i.e. Wi-Fi, 3G and 4G cellular services) have transformed robotic telepresence from being solely a research concept into commercially available products (Tsui *et al.*, 2012).

It is clear that TP robots are becoming more accessible than before, as this was what happened with the telephone, radio, television, and computers. Enhancing natural interaction in a remote environment was the main reason behind introducing personal telepresence robot systems, and they have now become technologically viable. The ability to design systems to suit individual requirements is another reason for their viability. Telepresence robotics systems have started to manifest themselves in a large number of places, in the form of interaction through live video to the physical presence of the person. However, despite these advances in technology and availability, an aversion has emerged to telepresence robotic interaction due to a lack of understanding and recognition of user requirements. The next section will summarise the areas of application in domains which, education, healthcare and the commercial sector, and highlight typical emerging issues.

1.4.1 Education

The development of the internet and new cost-effective technologies has promoted an astounding growth in distance education courses and has resulted in improvements to web-based or robotic based education courses. The availability of high-speed internet access coupled with more sophisticated compression technologies have made it easier to utilise distance learning in education in different ways. The technology can equally be used for students in distance education, or in a lecture theater setting to communicate to people outside of the establishment.

Distance education is different from traditional education; students and teachers are separated by space. Compared to traditional classroom education, distance education has been characterised by mediated subject-matter presentation and mediated interaction between students and tutors. Teachers will actively use a variety of teaching strategies in distance education, according to the characteristics of the student and the class, to ensure the learning process is still active (Figure 1.8). Thus, there are two constituent elements of distance education; one representing one-way traffic from the supporting (teaching) organisation to the students, and the other, two-way traffic between the two. Besides these elements, there is a peer-group interaction between students, which also counts as an important characteristic.

1. Student- teacher interaction
2. Student- student interaction
3. Teacher- student interaction

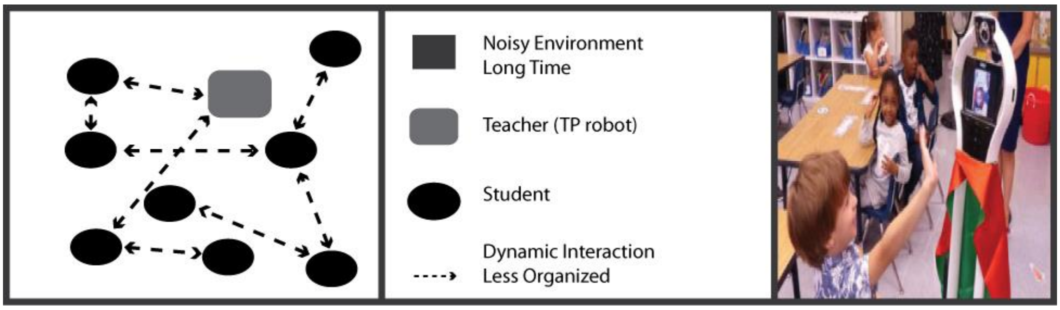


Figure 1.8. Characteristics of distant education environment

Various research has been carried out in order to understand the advantages of using telepresence in education for both higher and primary/secondary education (Patrice L Tamar Weiss *et al.*, 2001, Richards and Lara, 2011). Despite the complexity of this environment, telepresence robot use showed that both remote students and teacher were positively affected by the usage of TP. In contrast, another study suggested that telepresence systems may cause the individual to not "*be there*" psychologically, due to the lack of interaction between the users and participants, as highlighted by Fowler & Mayes (1997). As the environmental characteristic in this situation is totally different from traditional education, users of TP robots might require enhancements in terms of physical and social presence.

1.4.2 Healthcare

Differing completely from the distance learning situation discussed earlier in section (1.3.1), healthcare research was conducted in a fairly quiet environment (hospital) with a maximum of three participants (patient, caregiver nurse/ doctor) as characterised in (Figure 1.9). Telepresence robots can be used in different forms, and use terms such as telesurgery, telemedicine and telehealth, as highlighted by Bar-Cohen and Breazeal (2003). Use of telepresence was in cooperation with the professionals, aiming towards higher quality care without borders of time and distance, to realise remote medicine and healthcare with a face-to-face realistic experience, feedback and sense. TP also facilitates the provision of care for older adults living at home; enhancing their social interaction and safety, and giving caregivers some respite and support (Michaud *et al.*, 2007, Orha and Oniga, 2012, Ellison *et al.*, 2004, Findlay, 2003, Boissy *et al.*, 2007, Green *et al.*, 1995, Ashley Patrick).

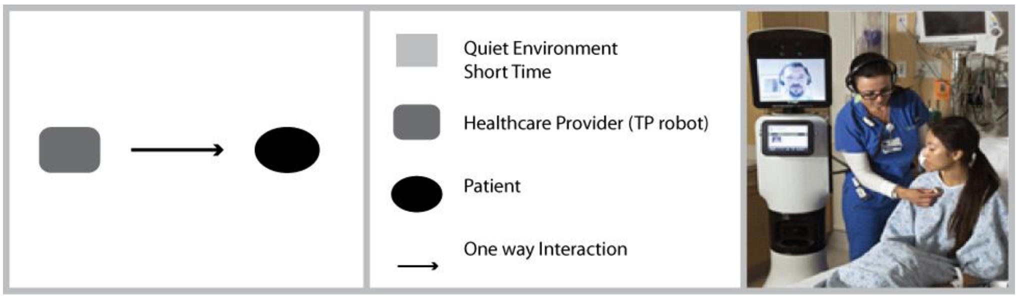


Figure 1.9. Characteristics of healthcare environment

Although their use has advantages, there are some studies that suggested that many people still have a negative attitude towards robots taking human jobs, so they still prefer to be in direct contact with their surgeon, as reported by Markar, (2012). This is due to the fact that most of these types of systems focus only on the quality of the service, ignoring the vital importance of the social design side related to the sense of presence.

1.4.3 Business

The business domain is most prominent in using telepresence, resulting from the need for consistent communication from remote locations. Telepresence has the advantage over video conferencing systems by closely replicating part of the experience of face-to-face meetings through superior audio quality, eye contact and no jerky motion (Lichtman, 2006). The situation will be similar to the healthcare one, where interactions will be fairly quiet and organised. In addition, TP supports active movement while simultaneously having a conversation, as mobility is one characteristic that differentiates telepresence robots from other domains (Figure 1.10).

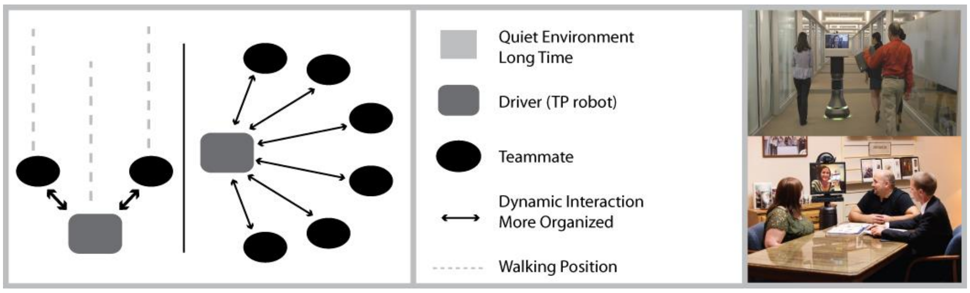


Figure 1.10. Characteristics of business environment

As with the previous domains, there are limitations as to what TP has to offer. Apart from the cost and bandwidth available, non-verbal information is the core in any communication, in particular, the eye gaze which facilitates the flow of conversation, and which is missing in TP robots. In general, the telepresence system suffers from a lack of meeting the sense of being there in real-time, which constitutes a major drawback for real interaction between the participants using a TP-robot, as reported by Tsui *et al.* (2011).

The majority of these systems were developed to try to solve technical challenges related to the remote site, such as audio and video, navigation, safety, overall design and social concerns such as direct eye contact and privacy. Ultimately, the goal is to reach a point in which the anthropomorphic features of the robot matches the nuances of human behaviours. Indeed, some research initiatives are investigating the human-like aspects of robotics, but not with regards to the home site.

1.5 Research motivation

There is no doubt that telepresence plays a big part in a whole variety of situations now, and will continue to do so due to the ever-changing nature of technology. However, the lack of human empathy and the de-personalisation in the current TP-robot system is part of the limitation, as we found from the previous sections. This might be due to the core complexity and significant costs of such an approach that render it difficult to justify for practical real-world applications. There are many different sorts of interactions included in such system; on one side, the pilot interacts with the remote human. This is a human-human interaction (HHI), while being embodied in a robot which they cannot see themselves, and at the same time, they control the robot via a computer, which is a human-computer interaction (HCI). On the other side, the remote human in the system interacts with another human –the pilot, which is (HHI), but also at the same time, they are interacting with a robot which is a human-robot interaction (HRI). As a result, very little real user experience has been gathered, processed and fed back into the design concepts and the nature of the physical platform makes it unfeasible to do so.

The relation between user and the technology need to be carefully investaged in order to create a successful experience. This can be done by basicly recognizing smart technology as an (*active participant*) in our environment as highlighted by Yung and Piebalga (2013).

Paulos & Canny, 1998 established one of the earlier works to in regard to this concept by understanding the human factors required in the development of The Personal Roving Presence or PRoP telepresence (Figure 1.11). According to Paulos and Canny:

> "*Our research is driven by the study and understanding of the social and psychological aspects of extended human–human interactions rather than the rush to implement current technological advances and attempt to re-create exact face-to-face remote human experiences*" (Paulos and Canny, 1998, p. 2).

Figure 1.11. Personal Roving Presence (Courtesy of Paulos & Canny, 1998)

Based on this fact, the PRoP designers paid as much attention to sociological research as they did to sensing and actuation techniques. They highlighted the importance of understanding human physiology, and the dynamics of face-to-face as central to TP research, in order to design realistic and thus natural interaction.

The communication issues highlighted in Paulos and Canny, study within the early TP systems are still being tackled today, and solving them remains the prime goal for contemporary telepresence robots (Tsui *et al.*, 2012). In general, TP robotics is missing the direct human contact factors which can be seen in our daily activities.

Based on these issues and in order to improve the relationships between the user characteristics and the feelings towards the robot, we need to improve our understanding of the basic telepresence rules like 'presence' within the human-human interaction. Then, this understanding should be placed within its applied context, which will help in the design phase, as it will feed the design decisions in developing strongly human-like robots in order to augment the human-robot interaction experience. Therefore, we need to understand the concept of human to human interaction and human physiology as a way to provide better understand the dynamics of human communication.

This understanding is vital to providing the most compelling overall experience for both the remote and home users. Identifying these human behavioural traits and employing new technology to implement them are the main aims for a TP-robot and for our research. Previous research has already explored this area (e.g. nonverbal actions) which was very informative, but was not based on detailed observations of human-human interaction. In addition, the focus of most of these studies was on the function or utility of the robot, ignoring the relationships between the user characteristics and the feelings towards the robot Nevertheless, we have found a great deal of research work relating to nonverbal behaviour and social robots, but we have found little research with regard to telepresence. As a result, we have found that it is appropriate to start with human communications, as we need to understand them first in order to outline the fundamental features of communication that must be supported, regardless of the communication modalities available.

1.6 Research aim

The aim of this thesis is the design and development of a new and novel approach to replicating part of human behaviour in telepresence robotic systems for more natural interaction, and to investigate the implementation of this system and its influence on users' satisfaction. In particular, this thesis examines the role of perceived cognitive social experience on attitudes towards the TP technology, which is designed for the purpose of video calls.

1.7 Research questions

Based on the aim of the research as specified above in Section 1.6, the present thesis intends to address the following research question:

Do users perceive an enhancement to the appearance of the telepresence robot interface, following the addition of nonverbal behaviour by the robot?

In order to answer this question, the following two specific sub-questions are to be answered:

1st Sub RQ: How does the enhanced appearance, following the addition of nonverbal behaviour by the telepresence robot interface, influence the users' sense of presence?

2nd Sub RQ: How do the perceived embodiment, perceived sociability, social presence, enjoyment, Temporal Dissociation and Focus Immersion influence attitudes towards the enhanced appearance, following the addition of nonverbal behaviour by the telepresence robot interface?

1.8 Research objectives

In order to answer the aforementioned main research questions, this study attempts to fulfil the following research objectives:

- Provide a framework for **understanding the mediated video display aspect** in TP by revisit **face to face interaction** in an attempt to approximate in-person experiences.
- Discovering the effect of **nonverbal behaviour** on the perceived presence experience.
- Explaining the impact of proposed **head movements** on the engagement.
- Evaluating prototype **HRI using nonverbal behaviour** from different aspects, considering both the hardware and the software issues.
- Developing prototype **HRI requirements, using nonverbal behaviour** for telepresence video interaction.
- Using the **Unified Theory of Acceptance and Usage of Technology Model** to measure the influence of the TP system on the embodiment, perceived sociability, social presence, enjoyment, Temporal Dissociation and Focus Immersion on the attitude towards the system.
- Using the **Unified Theory of Acceptance and Usage of Technology Model to comparatively** evaluate the nonverbal movement and the proposed TP systems for video calls.

1.9 Research scope

Researchers have realised that communication networks are social environments, and therefore analysing their performance must cover both engineering criteria such as telecommunication bandwidth and social criteria such as human communication. This thesis studies the behaviour of human interaction in daily life and transfers it to the digital world in relation to the interface features. In particular, the focal point of this research is an investigation into the impact of implemented nonverbal behaviour and studies the attitudes towards it in the context of video call interactions. With a particular interest in telepresence robotics, since the telepresence sector is vital for remote interaction and is rapidly growing, providing a fertile environment for users through the use of cheap and effective technologies.

1.10 Thesis Structure

For the purpose of achieving the research aim and extending the literature, this study comprises 8 chapters. This first chapter has introduced the concept underlying the study and introduced the topic. The rest of the thesis is organised as follows:

Chapter 2 reviews the literature in the field of research on social telepresence robotic systems. The chapter addresses the presence concept and characteristics and illustrates the role of engagement within the telepresence context. It also discusses issues of current social telepresence robotics and the influence of implementing nonverbal behaviour on the user presence and identifies the research gap.

Chapter 3 presents the research methodology adopted for this study, including the research approaches, research agenda and the methods and strategy followed in this thesis. The chapter also characterises the research sample, and discusses the experimental methods and methods of data analysis, with clarification of the ethical issues.

Chapter 4 demonstrates the nonverbal behaviour prototype development process. This chapter also discusses the primary study, and how the results established the foundations for the design of the main system. Further, the entire prototype development process including the system design and simulation design is discussed.

Finally, the chapter discusses the evaluation of the prototype with highlights of the outcomes.

Chapter 5 illustrates the design and development process, and the consideration of the main system design based upon the primary study outcomes. Main system designs are presented in this chapter, including all of the associated factors from the physical design, interaction methods and hardware and software requirements.

Chapter 6 elaborates on the conceptual model of the study to highlight the factors that affect the users' attitudes towards the system. The hypothesis of the thesis is discussed in this chapter. Finally, the research measurement development is presented.

Chapter 7 illustrates the system evaluation and analysis of the results, comprising the comparative study for interaction method evaluation (mimicked controlled movement as an interaction method with such a system versus vocal-triggered movement), attitudes towards the system using the Unified Theory of Acceptance and Usage of Technology Model (UTAUT) and finally, an evaluation of the Users' preference to use the system in the future, as part of the customised Unified Theory of Acceptance and Usage of Technology Model (UTAUT).

Chapter 8 discusses the research outcomes, the limitations of the research, and discusses future work, followed by highlights contributed by the study, and an overall summary of the research.

Chapter 2 LITERATURE REVIEW

The chapter presents a discussion about presence and telepresence robotic system. Followed by an overview of nonverbal behaviour as related to the sense of presence which supported by discusses of the face-to-face communication framework, and the factors associated with the coordination process. In addition, this chapter presents an overview of the current design and capability of telepresence robots. Then, it moves on to user acceptance of mediated technology by covering appearance, humanness and the uncanny valley aspects, followed by the conclusion.

2.1 Introduction

The technology advances in telepresence robot systems have opened up new usage scenarios, as it supports different functional compared to Skype and videoconferencing systems. Despite the advantages brought about by the use of advanced technologies for telepresence, which enhance the overall quality of the experience, and its functionality as a communication tool, there are still unsolved problems or issues regarding the concept of presence. This is due to academic and other research-oriented projects focusing primarily on autonomy, artificial intelligence and the ability to perform physical tasks, which might not be sufficient in improving and supporting the interaction between the TP-robot driver and home sites. Telepresence requires that users' senses be provided with such stimuli as to give them an insight into the feelings of the other person, in addition to the feeling of being present in the other location.

This chapter provides an overview of and explores the sense of presence and telepresence concept in section (2.2). It examines previous studies in section (2.3) that were associated with face-to-face communication and the supporting cues for a perceived telepresence experience. Section (2.4) discusses the communication framework and factors associated with the coordination process. Section (2.5) gives an overview of the current design of TP. Followed by User acceptance related to appearance, humanness and the uncanny valley are discussed in Section (2.6), moves on to the conclusion in Section (2.7).

2.2 Presence and telepresence

A similarity has been found between both concepts during our research, as presence is often referred to as telepresence. This can be seen in Sheridan (1992a) definition for telepresence; "*The feeling that you are actually there at the remote site of the operation*" (p. 120). This was also in line with the (McLellan, 1996, Schloerb, 1995, Reeves, 1991, Rheingold, 1991, Slater and Usoh, 1993) explanations; where they argued that the telepresence occurs when users feel that they are present in a different location other than they actually are.

It is clear, based on these definitions, that there is a subtle difference between telepresence and presence, although the concept of telepresence precedes and is closely related to the presence construct. This is mainly rooted in the site where one perceives, acts and ultimately experiences presence. This was highlighted by Marvin Minsky, when he emphasised that use of telepresence technology can make it possible for the human operator to feel the sense of being physically transported to a remote environment. Since this time, the term presence has been used to refer to a sense of transportation to a remote space created by technology. However, Steuer (1992) provided a clear distinction between telepresence and presence. He argued that telepresence is the mediated perception of an environment in which users are being transported via technology, whereas presence refers to the natural perception of an environment. In other words, presence can be defined as the extent to which the media represents the word (both physically and socially).

From the previous definitions, it can be seen that the feeling of presence lies at the centre of telepresence experience in specific and mediated experience in general. As a consequent, presence has become central to research about human-computer interfaces. This, in turn, resulted in many attempts to provide a clear conceptual definition of it. Lombard and Jones (2007) extensive literature review of presence, is count as one of the important and most recent works in the area. They highlighted the important of social presence as one of the psychological states in which the presence phenomena is an experience. Based on this, and in order to define our research area, we need to fully understand the definition of social presence. A better understanding of what it is, what encourages and discourages it in users, and its effects, will serve our work with a starting point for systematic research and theories on this topic.

2.2.1 Social presence definition

One of the early theories that drove the recent research about social presence can be traced back to the end of the 1970s; Short, Williams et al. (1976). Their widely-used conceptualisation of social presence is defined as "*the degree of salience of the other person*" (p. 65) which is, in other words; the sense of the projected presence of a person in face-to-face interaction. Short's concept of social presence is in a line with most social presence definitions, especially Biocca & Nowak (2002) where they define presence as the "*level of awareness of the co-presence of another human, being or intelligence*", and the "*feeling that one has some level of access or insight into the other's intentional, cognitive, or affective states*", (Biocca and Harms, 2002, p. 3). It can be seen that the Biocca and Harms (2002) definition was more comprehensive in the way that it covers a mediated experience in general. Although this experience is different in a number of ways from Short's concept, each of them is designed to give the user a type of "*natural", "immediate", "direct" and "real*" perception that creates for the user a strong sense of presence.

Since then, a growing community of multidisciplinary researchers has turned its attention to social presence, looking at what causes it, how the experience may be measured, and what effects it has on the media user. Because of its subjective nature, it poses a series of serious problems at both theoretical and empirical level and continues to be a challenge for researchers. The inner nature of presence has resulted in difficulties in the on-going theoretical work of conceptualisation. It has also led to a wide range of measuring tools, as there is no one specific and appropriate methodology or instrument capable of measuring it.

However, there is common ground shared by the researchers in the field, and from the definitions, it is suggested that social presence is simply the degree of resemblance to face-to-face interaction. Short et al., (1976) define social presence as the "*degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships*..." (p. 65). This means the degree to which a person is perceived as a "real person" in mediated communication. The capacity of the medium to transmit information about facial expression, direction of looking, posture, dress and nonverbal cues, all contribute to the degree of social presence of a communications medium To achieve this resemblance, the technology must manifest believable behaviour: it must establish appropriate social expectations, it must regulate social interaction (using dialogue and action), and it must follow social conventions and norms.

This insight has inspired this research to create systems capable of communicating with users in natural ways, using the natural perception of the real world as a starting point, to endow information systems with emotion and to enable them to communicate these emotions in ways that resonate with the human users. Although it will be difficult to give the same face-to-face experience with mediated interaction, it can be improved, to give a more accurate reproduction and simulation of reality than has previously been possible, as highlighted by Paulos and Canny (1998)

> "*We do not believe that we can ever replace true human-human interactions, nor is it our goal to do so. However, we do feel that it is possible to identify and distil a number of human behavioural traits or skills that are inherent to human communication, understanding, and interaction. Employing computer networking and robotic technologies to implement these traits, our goal is to ultimately provide a compelling overall experience for both the remote and local users and more importantly to create a usable system for tele-embodiment*" (Paulos and Canny, 1998, p. 296).

Thus, this research identified that starting with human communication was the best approach, as we felt we needed to understand it in order to use it to outline the fundamental features of communication that must be supported, regardless of the communication modalities available.

2.3 Learning from face-to-face interaction

The sense of presence is one of the most common psychological phenomena experienced by people in daily life, of which people are seldom aware. Ijsselsteijn & Riva (2000) described it as a default experience, which people are not used to reflect upon. This real-time face-to-face communication can be seen as the "gold standard" against which all other communication mediums are compared (Rogoff, 1991). It is dynamic and personal, where both members can influence and affect the other. Thus, it is incredibly complex; language is used to deliver news and information but in parallel, a subconscious stream of non-verbal cues such as tone of voice, facial expression, gesture, touch and eye contact tell a fuller story.

As people speak, they go through a multimodal process in their communication, involving a complex interaction between verbal and nonverbal behaviours. Speakers and listeners, using their nonverbal behaviour, can visually monitor their environment, their facial expressions change and their body posture and orientation shifts as the talk. Nevertheless, as people interact, they orient to, gesture at and manipulate physical objects in the environment they share (O'Conaill,Whittaker and Wilbur, 1993). Based on research calculations, the nonverbal channel of communication bears an estimated two-thirds of the social meaning load, leaving only one-third of all the meaning to be carried via the spoken word (Ray L Birdwhistell, 1966). These numbers suggested that people’s daily life communication depends heavily on them to express themselves, and to interpret the unspoken activities of others. Nonverbal behaviour, which includes gesture, head movement, facial expression, posture, and eye gaze are part of people’s daily communication. As explained in the next section, people draw upon nonverbal behaviour without thinking, in order to support and give additional meaning to the spoken words.

2.3.1 Gaze and Facial Expression

The face can be seen as a rich source of information about the effective status of the conversational participations, with eyes, mouth and eyebrows being considered highly expressive sources. Gaze is one of the most important communication behaviours for extracting information from the people we communicate with, and from the surrounding environment. It helps us to indicate in which direction the person is looking, the amount they are looking in that direction, as well as the physical features of the environment, as the eyes express our emotions and intentions and they help us to direct attention. However, facial expression is also used to coordinate the content of a conversation, according to Ekman and Friesen (1967). Speakers can gauge the level of listeners’ understanding by their facial expressions, which can reveal interest, puzzlement or disbelief in what they have heard. Also, listeners can get information from the speaker’s facial expression, alongside the lip expression, which can help the listeners to disambiguate the spoken content (O'Conaill,Whittaker and Wilbur, 1993).

2.3.2 Posture

Posture is often used in conjunction with other nonverbal behaviour, to determine the degree of attention. For example, the inclination and the orientation of the conversational participant's body, in particular, their trunk and upper body, where a forward leaning posture is associated with high involvement and interest.

In addition, posture can indicate the intensity of some emotional states, such as the tense posture which is associated with anger (Mark Knapp,Hall and Horgan, 2013). The position of both hands and legs are also important here.

2.3.3 Gesture

Gesture is basically the dynamic movement of a person's the head, hands and feet, and involve activities such as nodding, shrugging, gesturing, scratching and kicking formed during communication. Similar to gaze, gesture supports different communication functions, being used to coordinate conversational content, give reference and assist in turn taking. According (O'Conaill,Whittaker and Wilbur, 1993), there are two type of gesture which are speech related and speech independent. Speech-independent also was known as emblems and have a conventionalized meaning. This type does not relate to a speech but they have a verbal translation which differs from culture or subculture. For example, the gesture used to express Ok or peace in some culture is also known as V for victory in different culture.

Conversely to the first type of gesture, this one is tied to the speech and accompanies it as well. It often servings to illustrate what is being said verbally. It might the accent, word or phrase emphasises, pointing to present object. Pointing gesture might support different functions in speech, it can be used to achieve reference and also be used to manipulate or direct attention of others (Goodwin, 1981). Beside this, it can be used to describe the orientation of an abstract object in the space. With turn taking function, hand gesture can be used by listeners to signals that they want to say something so here it supports the coordinate process of the conversation.

Adding to this, the continuation of the speakers hand movement deliver that they wish to hold the conversational floor and vice versa (Whittaker and O'Conaill, 1997). Active listening usually involves a combination of nods and shakes of the head, and these gestures are vital during a conversation to indicate that the listener is engaged with the speaker. Head movements provide a range of information; they effectively give real-time listener reactions, in addition to the verbal channel. It may used to show agreement, disapproval and negation. Whereas, shoulders shrug movement which include shoulders up, shoulders flex and shoulders back might indicate a high interest as pointed by (Moore, 1985).

2.4 Communication Framework

Coordination between process and content occurs between speakers and listeners using these nonverbal behaviours, according to Whittaker and O'Conaill (1997). Content is mainly about the subject matter, and how participants build up common belief and understanding about it, whereas, the process is concerned more with the mechanisms and the management of conversation. In order to evaluate the communication process, we need to understand both these two aspects.

2.4.1 Coordination Process

Turn-taking and availability are the two aspects of process coordination. Turn-taking is concerned with the beginning and end of a conversation, and the switching of roles between speakers and listeners. It is also concerned with how participants jointly determine who will speak and who will listen. Surprisingly, most conversations are unplanned, so based on awareness of others' movements and activities, we can determine when it is opportune to initiate such an interaction (Whittaker and O'Conaill, 1997).

2.4.2 Coordination of Content

Maintaining a common understanding in conversation is the key aspect of coordinating content. It is concerned with the common events and objects that people jointly want to talk about, and which might be referred to. Speakers will have multiple communication choices, based on the common knowledge they share with listeners. The second aspect of coordinating content is feedback cues, which provide speakers with feedback on what they have said from listeners. Listeners might show acceptance, or the need for increasing their level of understanding etc. Feedback cues are important in order to identify any misunderstandings during conversation, so that they can be rectified. The last aspect is interpersonal cues, which are mainly related to social information about the participants. It allows participants to infer the emotional stance, affect and motivation of other people to the subject matter being discussed (Whittaker and O'Conaill, 1997).

Based on these theories and definitions, a TP-robot or any mediated communication system needs to exhibit recognisable human nonverbal behaviour. The next section reviews most of the current TP systems, with respect to these behaviours.

2.5 Current design and capabilities in telepresence robotics

Many telepresence robots have been developed for commercial purposes. These systems are usually marketed for remote meetings, home security or entertainment. Kristoffersson et al. (2013) made a list of different telepresence robotic systems, comparing their hardware and software specifications. Although they produced a good review about the current design capabilities of TP robots, they failed to cover the social design aspects. Based on these factors, the next section has reviewed most of the current TP robots, presenting the differences between each individual system (Table 2.1).

Table 2.1. An overview of common telepresence robotic system systems appearing in the literature

MRP system	Intended Application Area	Posture			Gesture	Gaze and Facial Expression
		Adjustable Height	Manipulation	Body Orientation	Head Movement	Screen \size
PRoP	Research	No	Laser pointer, 2DOF, hand/arm	No	Fixed screen	Not specified Small
Giraff	Elderly	No	No	No	Adjusted manually by the driver	A 14.1” screen
QB	Office	Yes	Laser pointer	No	Fixed screen	320x240 Pixel LCD screen
Texai	Office	No	No	No	Fixed screen	19” screen
VGo	Office	No	Handheld remote for local control	No	Adjustable screen	A 6 inch LCD screen
PEBBLES	School	No	Hand	No	Adjustable screen	Not specified Small
MantaroBot Classic	Office	Yes	Laser pointer	No	Fixed screen	A smartphone or tablet
RP-7	Healthcare	No	No	No	Adjusted manually	Not specified large screen
iRobot Ava	Healthcare	Yes	Yes	No	Adjusted manually	A tablet
Jazz Connect	Office	No	No	No	Adjusted manually	Not specified Small
RP-VITA	Healthcare	No	Yes	No	Adjusted manually	Not specified large screen
Luna	Personal	No	Special hands (Can be placed in any position)	No	Fixed Screen	An 8 inch LCD touchscreen

PRoP (Figure 2.1 (a)) was one of the early TP robots, as explained earlier. It is an internet-controlled, untethered, terrestrial robotic telepresence developed as a research platform. It was basically a mobile robot with video, audio and pointing devices. Users of this system were able to wander around a remote space, converse with people, examine objects and read. PRoP can support non-verbal communications by making simple motion patterns. These motion patterns express interest in a conversation, agreement with a speaker, or gain attention to ask a question in a crowded room. However, one negative point is that it has a fixed camera and height, which affects the driver's visibility of the field of view, unless the robot driver tilts the camera up. It also shows only part of the walker's body in the camera. Giraff (Figure 2.1 (b)), designed to assist elderly people. The camera and screen are mounted on a tilt unit, which allows the driver to tilt the Giraff screen up and down to control the field of view. However, it is mounted on a nonadjustable pole attached to a mobile base. QB (Figure 1 (c)) is a robot on two wheels that self-balances in the manner of a Segway (Guizzo, 2010). It has a thin telescoping 'neck' to adjust the camera height, which is a very useful function since the driver can move the head to eye-level in the remote site, whether they are standing, talking or even sitting at a table. Although the QB camera is fixed to the vertical axis, the axis itself can be rotated, solving the navigation issue, but not the communication issue. Nonverbal cues are not supporting in QB, as the screen is used only for navigation apart from displaying the remote site. In addition, it is supported with a laser pointer. MantaroBot (Figure 2.1 (d)) follows a similar design strategy, with an adjustable height with a laser pointer where the camera is mounted on the top. In Texai (Figure 2.1 (e)), the head camera points in different directions, or can switch to an auxiliary camera that shows the robot's wheels, to help in navigating, and will also allow 90 degrees of view from the centre, on the left, right and down. Besides this, it supports looking up, but tops out at 60 degrees. VGo (Figure 2.1 (f)), is another two-wheeled telepresence robotic system, which boasts serial ports to handle various attachments that can dispense medication on command from the remote operator. The screen can move remotely and the camera can change its angle and zooming capability, allowing 180 degrees of freedom (up and down) as well as 60 degrees of field of view for navigation only. The field of view is not fully covered because of the height of the VGo camera, which is mounted a fixed four feet from the floor. PEBBLE (Figure 2.1 (g)), has been designed around young hospitalised children. It is characterised by a typical wheeled robotic device, with a simple arm-like attachment, which enables the pilot user to perform a gesture similar to a hand raise. The head can move left and right, move the camera up and down, although it comes with a fixed height. Luna (Figure 2.1 (h)) was known as the first personal robot and was introduced by RoboDynamic. It is supported by a special hand for various activities, and has camera tilts with digital zoom; the panning is handled by the movement of the body, which counts as a downside in their design alongside the fixed height. Jazz Connect (Figure 2.1 (i)), has a head with the ability to rotate to give a sense of surroundings, with a fixed height of only 1 metre, which is very short compared with other TP robots. iRobot Ava (Figure 2.1 (j)), has a tablet fitted on top of the robot containing a camera, which supports the primary purpose (video telepresence). The tablet is mounted on the post on top of the robot and can rise up or down to interact with the person in front of it. It also allows the pilot to move the tablet pedestal up or down or even to rotate it to suit their needs. The torso is height-adjustable from 1 metre to 2 metres long in order to meet the user's eye level. RP-7 (Figure 2.1 (k)) by InTouchHealth has a pan-tilt mechanism on the head of the device for smooth control. Although the height of RP7 is 1.52 metres, it supports all the positions for the patient in their interactions with doctors. This is achieved by designing the neck/head component to give it fluidity in much the same way as the human head. RP-VITA (Figure 2.1 (l)) has a swivelling head to reflect the doctor's head movement, as the company claims, but it does not provide information about how this is done. No additional information is available about height, although it seems to be fixed-height. MeBot (Figure 2.1 (m)) was designed with an emphasis on being able to convey the non-verbal channels of social communication, as it is able to communicate some body posture and a wide range of head movement.

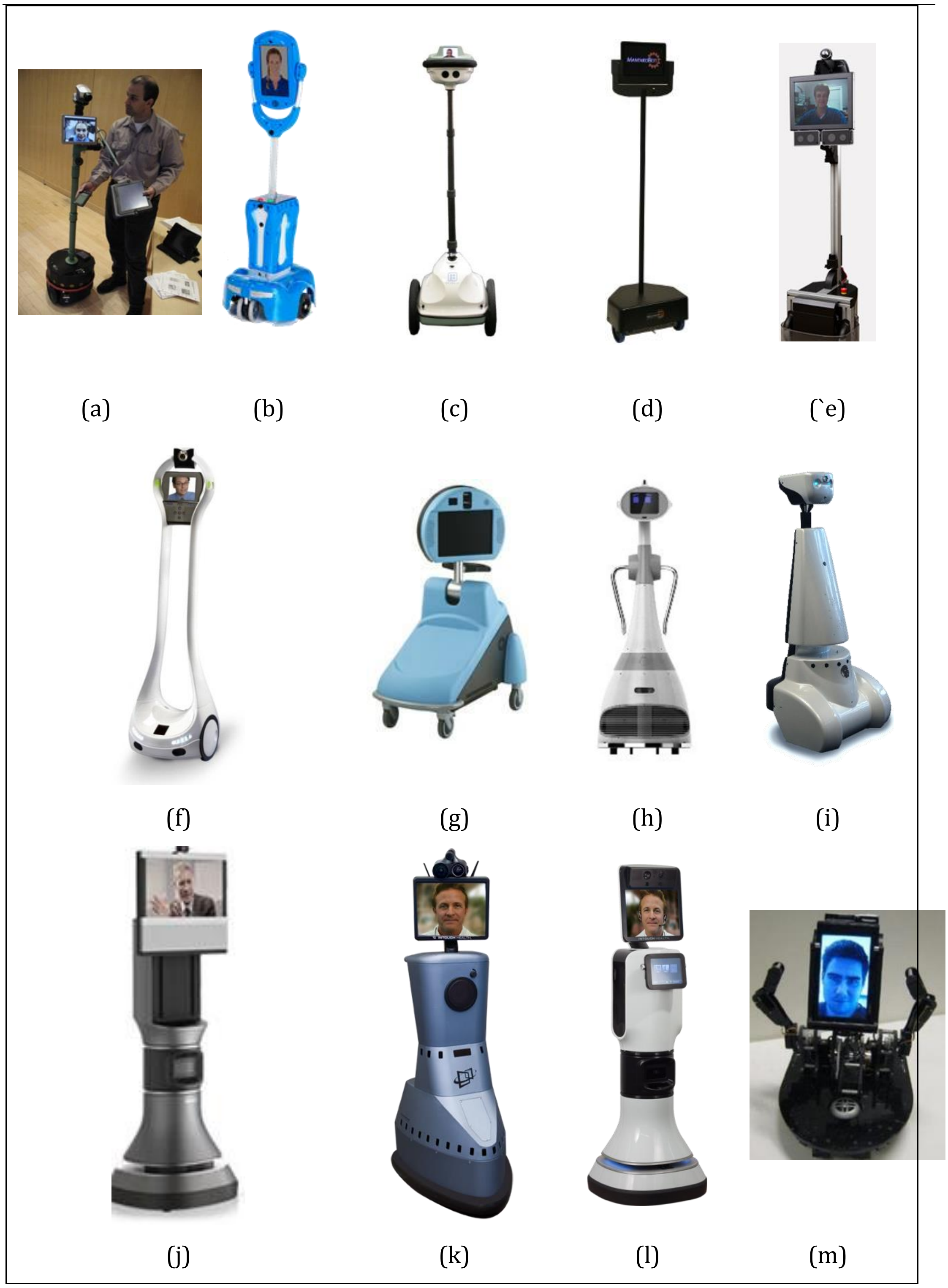


Figure 2.1. (a) PRoP, (b) Giraff, (c) QB, (d) MantaroBot, (e) Texai, (f) VGo, (g) PEBBLES, (h) Luna, (i) Jazz connect, (j) iRobot Ava, (k) RP-7, (l) RP-VITA, (m) MeBot

Although there has been development in the field of telepresence, it is clear from the table and the description above that most current design of TP robots are mainly focusing on the physical capabilities of the robots (physical presence), ignoring the vital importance of social presence via the full breadth of nonverbal behaviour content that humans experience in daily life. Guidelines for embodiment design are still lacking in focus on telepresence, which we have summarised here under the following points.

- ***Fixed height***

One potential aspect of a one-on-one conversation is to be able to look at the person's face in order to show attention, which not supported in these designs. With fixed height, the walking position is one of the issues facing most TP-robots. An adjustable height would help the users to move the head to eye-level in the local site, whether they are standing, talking or even sitting at a table. VGo has a feature that allows the user to see the drivers or the pilot on a screen, offering a simulation of closeness and of physical presence as (Tsui *et al.*, 2011) explained; "*The VGo robot's camera is mounted 4 feet from the floor. If a walker with a height of 5' 9" was in the robot's personal space (1.5 to 4 feet) or near social space up to 5 feet, then only the walker's body would be visible in the camera's field of view unless the robot driver tilted the camera up*" (p. 6).

- ***Fixed field of view and fixed screen***

The orientation and the field of view are restricted by a fixed camera. This is another of the issues facing TP in terms of the one-on-one conversation, since the walker is unable to walk alongside the robot as two people might do. The walker would be outside the camera's view and would have to move ahead of the robot. Tsui. K, Desai. M, et al. (2012) found that most of the walkers walked in front of the robot, or in front of the robot walking backwards, so that they were face to face with the camera.

In addition, users of such of a system would need to turn their head rather than their entire body to look around. An independent camera with an 180 degree of view from the centre, on the left, right and down, would provide good visibleity, and would avoid the unnatural movement of the entire robot just to change views.

- ***Articulated arms***

Users cannot touch, refer to or even manipulate objects around the physical space; however, the addition of articulated arms, controlled by the driver, could remove this issue.

- ***Small screen***

Most of these systems utilise a small screen within the design, and the image is limited by the screen size and becomes literally and figuratively two-dimensional. This is because the screen only displays the image of the person's head, which cannot fully convey their presence, as it is not possible to display the most useful aspects of nonverbal behaviour.

These are only a few points highlighting just some of the issues with current telepresence robot design. Previous studies, evaluations and design reviews of TP robotics systems have tended to focus on improving performance and reducing mobility issues for the driver rather than trying to increase social presence.

2.6 User acceptance

The challenges for Intelligent Technology and the Human Robot Interaction researchers are numerous. In our case, the challenge will be mainly related to the need to develop methodologies for eliciting user requirements in real contexts, which must meet the requirements of users on both sides of the conversation, and be acceptable in the long-term. In order for such systems to deliver their intended benefits, and for users to socialise through them, users need to accept the systems. Without the user's social acceptance, technology cannot hope to deliver whatever value it may offer, which might cause complications and be crucial in the case of our TP-robot. Acceptance is defined as the robot being willingly incorporated into the person's life. For acceptance of robots to occur, user satisfaction is essential.

Satisfaction means "*a judgment that a product or service feature, or the product or service itself, provides pleasurable consumption related fulfilment*" (Oliver, 1997, p. 13). Previous studies have investigated the impact of various human-robotic interaction factors on user satisfaction, discussing it from different angles.

- ***Appearance***

Most previous research has focused on the robot's appearance, rather than its behaviour, which is a topic for future study. The appearance of products is very important as it contributes to the sense of identity (Powers *et al.*, 2005). Careful consideration of the TP robot design is needed to minimise the negative emotional response.

Powers & Kiesler (2006) stated that the appearance of a robot has an influence on people's perception of the robot's abilities. If it looks as though it is high-tech, users will expect it to perform in a high-tech manner. In order to avoid raising people's expectations too high, Breazeal (2000) made a social robot named 'Kismet', and fashioned it in the style of a 'young creature'. She argued that people would expect a robot with a humanoid and adult appearance to be a natural conversationalist, which would be difficult to achieve in terms of software and programming. However, it was also argued that a 'young creature' would not really inspire confidence in people who were asked to use it as a robot in healthcare. A similar issue was probably the reason why another robot called 'Hopis', also designed as a healthcare robot but with a toy-like aspect, was unsuccessful. There are a number of aspects of the appearance of any robot that affect humans in different ways, which view is expanded in the next section.

The physical size of a robot can have a critical impact both on practicality and on acceptance in its role – for example, Breazeal's Kismet was made small in size in order to trigger a protective and caring reaction from human users (Powers *et al.*, 2005).

- ***Humanness***

There is debate as to how human a robot should look, with many preferring a less humanlike appearance. Mori's hypothesis about the 'uncanny valley' (1970) stated that; familiarity increases with a simulation of a human being's appearance and/or motion, until the point that the uncanny valley begins (Figure 2.2).

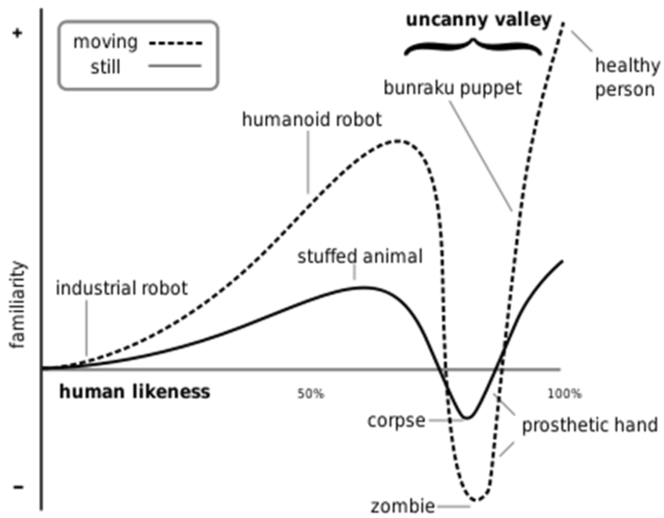


Figure 2.2. Masahiro Mori's theory about the Uncanny Valley

The term 'uncanny valley' makes reference to the sense of unease that people experience when meeting a robot that looks significantly, but not totally, human-like. A survey of 2000 people at an Expo I Switzerland revealed that as few as 19% of respondents stated that they preferred robots to look human, whereas 47% said the opposite (Arras and Cerqui, 2005). A survey carried out by Oestreicher on a relatively small number of students of social science asked about the importance of certain behavioural and appearance aspects, for robots to gain social acceptance (Oestreicher, 2007). The analysis of the results was not carried out using the most appropriate statistical methods, but a simple visual review indicates that it is better for robots to look more like machines than humans and that they needed to be safe, easy to use and reliable. The results also indicated that it was less important for a robot to have an expressive face or a personality. The degree of human-likeness required for different types of applications may vary. For example, in the case of a robot toy intended for use by children with autism, it was recommended by a variety of people (including therapists, family members and panels of teachers) that it be more mechanical that human-like in appearance, and to be controllable via a series of buttons (Robins *et al.*, 2007). Not everyone is fond of animals, so a mixture of responses are generated by robots that have animal-like features (Libin and Libin, 2004). Someone who generally dislikes animals or pets would tend not to engage with a robotic animal, but someone who was afraid of real-life animals (fear of biting) or felt that animals were unhygienic, would actually tend to engage, because the robotic animal gave them the feeling that it was 'safe'. According to research, the appearance of a human-like robot may encourage people to accept shared responsibility, which may lead to a future in roles involving job-sharing (Hinds,Roberts and Jones, 2004).

2.7 Conclusion

We believe that telepresence robots have great potential to prove useful in various areas. However, the lack of the quality of a person to person interaction through telepresence in respect to the local site has not yet been explicitly quantified. Prior studies illustrate that the role of nonverbal behaviour in communication is both complex and subtle. It shows that we need a more detailed understanding of the precise function that these behaviours play in communication. This understanding is vital from a practical viewpoint, as we need to design technologies that exploit behaviour in empowering a robot with social functionality, and we have presented some of the issues with current telepresence robots which have not yet been solved. In addition, whilst there is a great deal of work relating to non-verbal cues and social robots, there has been less research and development on how two or more interlocutors indicate an interest in initiating communication, how they then continue to interact, and how they terminate communication in respect to TP systems.

Based on these facts, this research plan to fill this gap by focusing on human to human interaction, as we believe that the implantation of some aspects of human to human interaction within TP design can affect how the human judges the interaction experience. We will use these findings as a base for our next study, by applying the human–human interaction discussed earlier to human–robot interaction

Chapter 3 RESEARCH METHODOLOGY

This chapter presents the research methodology implemented in this study. The research rational and philosophy in addition to, experimental methods, methods of data analysis, and ethical issues are described.

3.1 Introduction

This chapter presents the research methodology for the study. The research rational is presented in Section (3.2), the research philosophy is discussed in Section (3.3, 3.4, 3.5 and 3.6). The ethical part of the research discussed in section (3.7) followed by the ethic and moral implication of the system in section (3.8) and finally the chapter concludes with a summary in section (3.9).

3.2 Rational of the study

Social research is a term used to describe any research on topics related to questions relevant to social scientific fields. As Bryman (2015) stated:

> "*Social research involves research that draws on the social sciences for conceptual and theoretical inspiration. It draws upon the social sciences for ideas about how to formulate research topics and issues and how to interpret and draw implications from research findings.*" (p.3)

The rationale for doing social research varies depending on the ideas of the researcher and intellectual traditions of social sciences. There has been development in the field of telepresence robotics systems with recent advances in telecommunication technologies, systems and software, however, the lack of human empathy and de-personalisation in the current TP-robot systems is still part of the system's limitations. This is due to the focus of academic or other research-oriented projects being primarily on autonomy, artificial intelligence and the ability to perform physical tasks, which might not be sufficient in improving and supporting the interaction between the TP-robot and the participants. In addition, most research does not represent reality, and the actual capabilities we have in the real world do not match the capabilities that are used in research. Therefore, most of the research conducted on improving interactions within telepresence robot platforms adds to current research and knowledge generation as opposed to application.

Consequently, users do not get clear benefits from them in the real world as Coradeschi *et al.* (2011) explained:

> *"While these capabilities advance the state-of-the-art in the field they are usually not practical or reliable enough for commercial deployment. Sophisticated autonomous devices that require near-constant attention by the designers may advance knowledge in the field but are not suitable for real-world deployment"* (p. 3).

Improving the functionality of telepresence robots should not conflict with valuing the social interaction between the users and the robots. One of the main issues in the current telepresence system is the inability to convey a sense of social presence supported by nonverbal behaviour. The lack of the quality of a person to person interaction through telepresence constitutes a major drawback for real interaction, as reported by the users. Understanding human physiology and the dynamics of human interaction is central to TP research in order to design realistic and thus recognisable human gestures.

Thus, the proposed system offers an intuitive way for a telepresence robotic system to attempt to emulate the ideal humanlike experience that users (using nonverbal behaviour) would have when interacting with the TP. Consequently, users of telepresence could gain the benefits of this system through humanlike appearance and interaction methods, as well as fulfilling their needs for an emotional connection when interacting with the TP, alongside other features, through a technologically advanced and realistic method of interaction.

In our case, this research started by identifying the research gap throughout reviewing the literature within the research area, developing the system following an iterative design and testing approach, evaluating the engagement of the system and the attitudes to it in order to examine whether it could fulfil the requirements and address the user needs, and finally, reporting the results (Table 3.1). The research proceeded in the following stages:

3.2.1 Stage One: Iterative Development: design and testing

Understanding the users and their requirements is essential to the provision of a system that gives a compelling overall experience for the users, without neglecting the importance of analysing human physiology and the dynamics of natural gestures (Paulos and Canny, 1998). In the case of the current study, the requirements of users have to be investigated and fully considered when designing the system. Maintaining this approach in the telepresence robot is vital in enhancing the communication experience. The procedure adopted in the proposed system design (Figure 3.1) for fulfilling users' needs and the task requirements are as follows:

- ***Prototype system development: Preliminary interaction method***

This study proposed to focus on head movement replication, by developing a model which was a monitor on a stand and proposed a 'Wizard of Oz' approach for the system. This approach optimises the communication experience by bringing inanimate objects to 'life' though mimicking actions with a simple puppeteer mechanism.

The initial version of the system was developed to provide the prospective user with a more comprehensible and natural interaction through a real-time production of realistic and life-like movements of the person's head on the screen, in a way that was highly interactive and involving. A puppeteer mechanism was used as a way of mimicking a nod -up and down- as well as a shake –left and right- head movement, as it was deemed the simplest and most cost-efficient tool.

- ***Exploratory Investigation***

An Exploratory study was conducted in the current research as the initial stage in the system development. The purpose of the study was to ensure the chosen alternative design would be effective in future use, and that the system was performing its required function for the design targets in the future (Gould and Lewis, 1985).

The aim of the research exploratory test was to examine the feasibility of the system, discover the initial engagement issues of such a system and observe the user experience.

Representative subjects were involved in the exploratory study. A total of 28 users were invited to participate in one-to-one video-based interactions and to provide an initial evaluation of the system. A pre-test questionnaire was given to the subjects to collect demographics. After trying the system in two separate conditions, the subjects were required to complete a post-test questionnaire to obtain their feedback on satisfaction in their engagement in a task using the system. The study also measured the appropriateness of the head movements.

To determine interaction behaviour during the study, the participants' behaviour was videotaped and the recording reviewed and transcribed under a number of headings.

The results collected from the questionnaire and videotaped data offered an initial appraisal of the system. For example, the users made positive statements regarding the movement replication; that they were able to see movements by the person on the screen during a video call, compared to traditional on-screen movements. The results of the exploratory study guided the required system design improvements. Observing the users' experience indicated that an interaction with a socially expressive system using the Wizard of Oz approach may not be the most immersive and effective way of interaction (see Chapter 4).

- ***Further investigation***

The issues highlighted in the exploratory study which required further investigation and development was presented, with their possible solutions. These issues were considered in the revised system design. One of the main issues that users responded less quickly to the screen movement was highlighted and further discussed as a main point of the proposed system.

It was recognised during the Exploratory study that the limitations of the puppeteer mechanism as a means of replicating the movements was one of the main issues; further research was therefore carried out in order to find an alternative interaction procedure to complement the overall experience (see Chapter 5).

- ***Final proposed interaction method: final system***

A vocal stream recognition movements-based interaction method was adopted instead of the conventional puppeteer movements-based interaction, because of its life-like nature, as well as its expected influence on boosting the attitudes positively. Several movements prototypes were developed and evaluated, before producing the final version of the system. Explanations of the system development process, requirements, and functionality are given in Chapter 5.

3.2.2 Stage Two: Conceptual model and hypothesis development

Researchers are able to establish user attitudes and also any relationships existing between constructional factors and research variables via the use of theoretical frameworks. In the current study, a variant on the UTAUT or Unified Theory of Acceptance and Usage of Technology Model was chosen, to enable the identification of factors affecting the users' attitudes to general interaction using a socially expressive telepresence robot system (see Chapter 6). Additionally, for the current study, the conceptual framework demonstrates the key hypotheses which were developed with the purpose of exploring relationships amongst the UTAUT constructs.

3.2.3 Stage Three: System evaluation

In order to evaluate the system from the perspective of the expected population of representative users, two comparative evaluations for interaction method and the future usage of the system were made.

- ***Interaction method comparative evaluation***

An evaluation of Automatic Vocal-triggered movement compared with mimicking movement using the same system was carried out. The users' perceptions and preferences for either system was the source of an in-depth understanding of the contribution of the proposed system. To be more precise, it aided us in the identification of the positives and negatives of our system, and the ways in which it could enhance the user's experience.

A within-subject study was carried out with the involvement of a sample of users who were associated with a lab at the University of Strathclyde. A questionnaire was used to collect primary data, and secondary data was obtained via observations of user trials. The procedure, requirements and specifications for the study are expanded further in Chapter 7.

- ***Future preference for using socially expressive telepresence robotic systems: comparative evaluation***

A comparison of user attitudes to the socially expressive TP-robot and user preferences in relation to future system use for video calls was obtained via an evaluation study. A questionnaire was used to collect primary feedback data, and secondary data was obtained via observations of the users. The procedure, requirements and specifications for the study are detailed in Chapter 7.

3.2.4 Stage Four: Evaluating the system using the Unified Theory of Acceptance and Usage of Technology Model

The last stage of the evaluation procedure utilised the UTAUT (Unified Theory of Acceptance and Usage of Technology Model) as its basis. Integrating extra constructs into UTAUT allowed us to do an evaluation of the system, while including the factors that in this particular context, we could have expected to have an influence on the result.

User trials took place with representative users participating, and observations of the users during the trial were made, plus the users completed a post-study questionnaire based on (UTAUT) to provide additional data. See Chapter 7 for further detail.

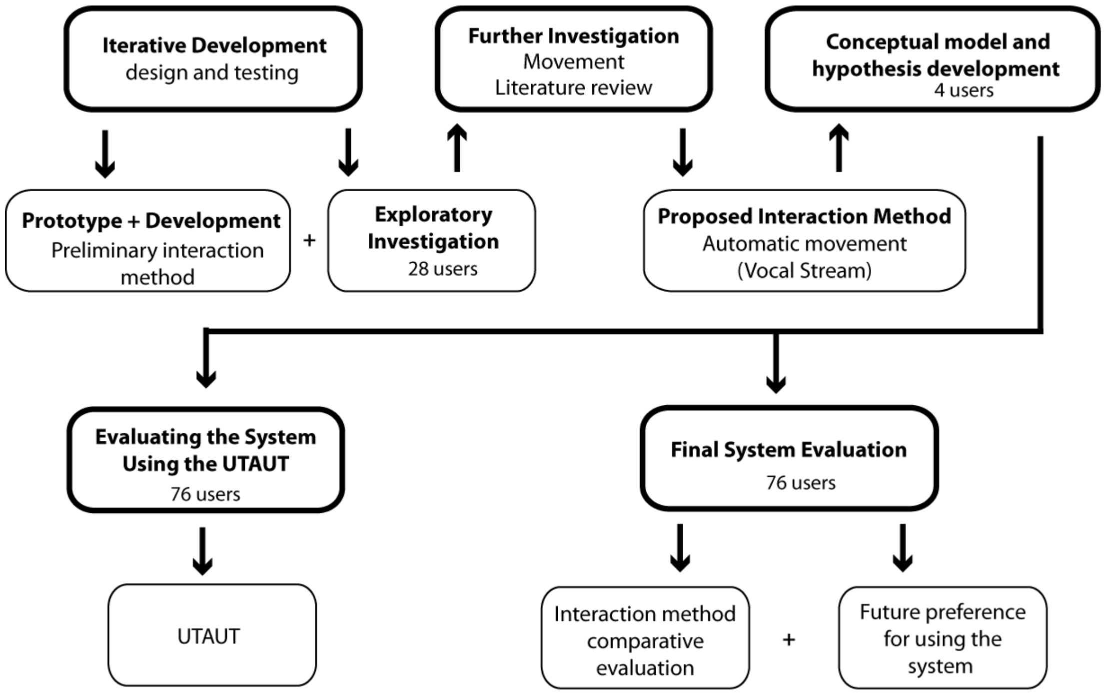


Figure 3.1. Iterative system design process

Table 3.1. Research agenda illustration

Stage 1: Iterative Development design and testing		
Phase	**Activity**	**Instrument/Analysis**
Prototype system development	Monitor on a stand and proposed a Wizard of Oz approach for the head movement replication	**Instrument**: 3Ds Max, Photoshop
Exploratory Investigation	User engagement user trials in the Lab	**Instrument:** Observation + engagement questionnaire **Analysis**: Qualitative & Quantitative
Further investigation	Further investigation issues from Exploratory study	Books, Journals, etc.
Final Proposed Interaction method: Final System	Consolidation based on 1 lecturer in human-cantered design + 2 programmer evaluating proposed movements	**Instrument:** Observation + oral questioning **Analysis**: Qualitative & Quantitative
Stage 2: Conceptual model and hypothesis development		
Phase	**Specification**	**Instrument/Analysis**
Literature review	Developing the theoretical framework and research hypotheses	**Instrument**: Unified Theory of Acceptance and Usage of Technology Model literature
Stage 3: System evaluation, comparative studies		
Phase	**Specification**	**Instrument/Analysis**
Interaction method evaluation	Testing Automatic Vocal-triggered movement MobiBot system against controlled movement interaction methods	**Instrument:** Observation + UTAUT questionnaire **Analysis**: Qualitative & Quantitative
System evaluation	**Testing the Automatic Vocal-triggered movement MobiBot system**	**Instrument:** Part of UTAUT questionnaire + user future preference **Analysis**: Qualitative & Quantitative
Stage 4: UTAUT Evaluation		
Phase	**Specification**	**Instrument/Analysis**
User trials	Engagement + attitudes towards technology based on UTAUT	**Instrument:** UTAUT questionnaire **Analysis**: Qualitative & Quantitative

3.3 Research philosophy

Understanding the philosophical character of research is a key consideration for researchers, as if can affect the value of the outcome. Sarantakos (2012) stated that understanding and interpretation of world reality by researchers will have an effect on both the processes used in the research and the findings and results of the research. Making the right choice of research strategy and research techniques to be employed is closely associated with the researcher's own understanding of the nature of the research.

Philosophers often debate the key matters of epistemology and ontology, because different approaches to methodology development are required by different research. There are four dimensions that can be used to explain the different approaches; methods and techniques, ontology, methodology and epistemology. An awareness of this philosophical assumption is necessary, due to its ability to improve the quality of the research, and increase the creativity of the person carrying out the research, as it can increase both the quality and contribute to the creativity of the researcher. This will be explored further in the next section.

- ***Ontological:*** This is related to the nature of truth in social science phenomena that influence research processes, and has two main categories (Easterby-Smith,Thorpe and Jackson, 2012) See (Figure 3.2)
    1. Objective ontology which is a more physical science approach in which the observers aim to discover what is there.
    2. Subjective ontology where the aim is to understand people's interpretations and perceptions.

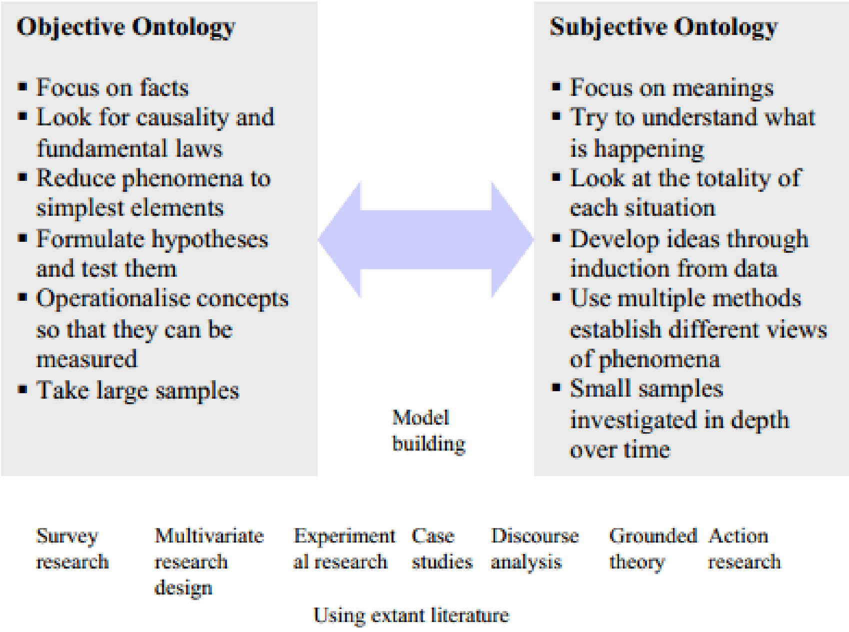


Figure 3.2. Choice of research methods related to ontology (Beech, 2005)

- ***Epistemological:*** This is also known as the scientific approach, in which the researcher will make assumptions and have a view on how research should be conducted. This will be based on which approach should be followed to test hypotheses, using which measurement techniques. It is a "*general set of assumptions about the best ways of inquiring into the nature of the world*" (Easterby-Smith,Thorpe and Jackson, 2012), (p. 60) which is generated from the different views we have gained through our background, education, personal and professional experiences. Four key epistemologies or paradigms in social sciences are:
    1. Positivism
    2. Critical realism / Relativism
    3. Interpretivism / Social Constructionism / Phenomenological Approach
    4. Action Research
- ***Methodology*** is a "*combination of techniques used to enquire into a specific situation*" (Easterby-Smith,Thorpe and Jackson, 2012, p. 60). In addition, there might be an alternative such as deductive, inductive and co-operative inquiry. Deductive methodology often starts with data rather than literature, whereas

inductive methodology starts with literature and finally, co-operative inquiry is seen in action research where researchers are involved to a high degree.

- ***Methods*** “individual techniques for data collection, analysis, etc.” (Easterby-Smith,Thorpe and Jackson, 2012, p. 60). These methods result from the specific epistemology researchers have decided to follow. In each epistemology, there are commonly used techniques and methods to help researchers to conduct their study, and they also have an impact on what he or she can see and find. These methods and techniques vary from quantitative to qualitative methods such as statistical testing, experimental, secondary data analysis, case study, observation, interviews and participation (Ates, 2008).

3.4 Ontology choice exploration for this study: Objective ontology

The research will be under the objective ontology (Table 3.2), since this research are trying to explore if implementing some nonverbal cues will provide more realistic and thus recognisable human gestures in interactions with telepresence robots. Furthermore, it is the communicative richness of these cues that offers opportunities for richer and more natural behaviours for telepresence robot users. This is because this research is aiming to understand the meaning of social phenomena (presence), and aiming to examine the best way to implement these, by focusing on facts and clarification rather than meanings. We will be as independent observers, searching for explanations within human behaviours with regards to their interaction with different TP systems. To achieve this, this research will test different aspects with regards to human communication within TP systems. Ates (2008) clarified this point:

> “*We interpret the material world and the interpretations we make might be essential to understanding social phenomena. Still the material components of reality as such will not change due to our interpretations of them*.” (p. 4)

Table 3.2. Summary justifying our choice of objective ontology of research

Relevant characteristic	Form of Ontology	Relevance of area to this research	Application in this research
Focus on facts	Objective	Though we have control over which kind of nonverbal cues we will use, we don't have control over actual interaction and activities. It is believed that the truth holds, regardless of who the observer is.	To some extent
Look for causality	Objective	This research aims to clarify whether adding some values to TP robots has a link with a sense of presence for the local site.	Yes
Reduce phenomena to simplest elements	Objective	The sense of presence broken into different categories from the literature.	Yes
Formulate hypotheses and test them	Objective	Hypotheses are formulated and will be tested out. The aim is to discover if implementing some nonverbal cues will provide more realistic and thus recognisable human gestures in telepresence robot interactions.	Yes
Operationalize concepts so that they can be measured	Objective	Concepts are operationalized by constructs and quantitative and qualitative measuring of constructs considered.	Yes
Take large samples	Objective	Better to have small scale with observations, as large-scale questionnaire or surveys will not provide evidence on how those cues would affect user's sense of presence.	No
Focus on meanings	Subjective	The broad aim is to understand the meaning of presence and its related issues through experiments and observations. The truth is not deemed to vary depending on the observer.	To some extent
Try to understand what is happening	Subjective	This research aims to understand different categories of presence with mediated communication, by understanding it in relation to human-human communication.	Yes
Look at the totality of each situation	Subjective	This study aims to clarify how to modify the TP interface in order to achieve a sense of presence.	Yes
Develop ideas through induction from data	Subjective	The starting point is literature rather than data.	No

3.5 Epistemology choice for this study: Critical realist paradigm

At the beginning of our study, we thought that we would follow the positivism paradigm, as this research were mainly standing as observers of what was happening, after simplifying the phenomena down to their smallest elements. The positivist paradigm can provide sufficient coverage of a situation which has been analysed using large samples, as stated in Esterby-Smith, Thorpe and Lowe’s book (2012), which can be discussed later. However, it would have been difficult to obtain or even to identify a large enough number of respondents to generalise the results, which is the focal point of research quality in the positivist paradigm. In addition, there is a need to use highly structured interviews, which would not be possible in our case, since we planned to use experiments. On the other hand, the interpretivist paradigm is meant to start with data rather than a literature based theory, or hypotheses to be tested, which again did not match our case. Although there have been various studies and data collection conducted in the same general topic area as our study, they were not related to the same technology. Another con related to the interpretivist paradigm is that it would be time-consuming for us to collect sufficient data from studies, and then generate ideas based on induction from the qualitative data generated by observations (Ates, 2008).

We found that it was better for us to follow the critical realist paradigm for several different reasons. Firstly, this paradigm attempts to use mixed methods - unlike the other paradigms. As the basis of this study was the performance of the system, the use of mixed methods would contribute to the research quality by strengthening the constructs, internal/external validity, and give more credibility and strength to the outcome of the research. According to Esterby-Smith, Thorpe and Lowe (2012) “*one should attempt to mix methods to some extent, because it provides more perspective on the phenomena being investigated*” (p.71). It is a crucial point in any research to choose a feasible and doable research design. Thus, this kind of paradigm added more strength to our research, unlike the positivist paradigm. It gave us a middle view between positivism and interpretivist.

Quantitative analysis methods are empirical methods for investigating events, in that data are analysed and displayed numerically (Given, 2008). This method (as the name implies) has its basis in the measurement of quantities, using mathematical and statistical tools, and the output is usually graphical or tabular. Qualitative analysis methods use descriptive measures and involve a different type of data, with results that are usually words rather than numbers. This method is often used to identify context surrounding a subject as it is based on broader questions than quantitative methods, and can be used to analyse behaviour (Maxwell, 2008).

Our study, therefore, used a mixture of quantitative and qualitative methods in order to obtain relevant data. We chose an analytical questionnaire to explore any relationships which may have existed between groups and to make comparisons and understand behavioural cause and effect. This method had the advantage of enabling us to collect standardised data from larger groups of individuals, quickly and consistently. Our intention was also to use the questionnaire in combination with observation in order to gather qualitative data and to allow more in-depth exploration of the areas of interest to the researchers.

3.6 Methodology choice for this research: Inductive approach

Our research began by mainly focusing on a literature review, and from there we developed a conceptual framework as a starting point. During our literature research concerning presence and the different elements associated with it, we increased our awareness of various aspects of the problem we were trying to solve. We decided to use an understanding of human communication within the context of real life situations as a guideline and basis for the conduct of our work.

3.6.1 Research method: experiment design

The purpose of the using experiment was to gather real-time data. As experiment utilises a similar structure to the nature of reality. Using used a one-group in two-condition design will help in evaluating the experience of the subject when interacting with a socially expressive system.

3.6.2 Techniques for this research: questionnaire and observation

We aimed to understand the situation under specific circumstances, by using observation of the subjects in our experiments. For this study, quasi-experiments were used in order to analyse and understand the situations; it resembled an experimental design, but lacked random assignments.

For most of the study, we conducted experiments using different systems in different situations, as described by Tsui *et al.* (2012). In their study, they designed an experiment for different situations; one involving a conference room meeting and the other one for a casual hallway conversation while walking. These experiments were supported by a pre-experiment and post-experiment questionnaire and observation. In our case, we used the same method of questionnaire and observation as our choice of approach.

- ***Questionnaire***

A survey was designed in order to measure and identify the influencing factors in the overall design and highlight any issues. The survey design was based on our findings from the literature review. According to Fink (2013), the questionnaire method provides an opportunity to collect information from more than one person in a short space of time, and this is usually used alongside other methods to achieve standardisation and consistency.

- ***Pre-test questionnaire***

A pre-test questionnaire was provided to the users in both the primary study and the main study. The pre-test questionnaire aimed to collect information on group differences such as age and gender. For the main study, the pre-test questionnaire was also used to collect information regarding previous experience in using a smartphone for a video call. The questionnaire also investigated preferences in using a video call, and how often it was used each week.

- ***Post-test questionnaire***

For the primary study, a post-test questionnaire was developed to gather the users' opinion of the system, the user's sense of engagement and the perceived usefulness of the movements. The main study used the questionnaire to gauge the acceptability of the system after the trials had been conducted. Questionnaires can be regarded in some circumstances to be an indirect data gathering method, when used to gauge subjective preferences and opinions. They can, however, be regarded as a direct data gathering method when employed in evaluation of satisfaction (Nielsen, 1993).

The difference between a questionnaire and an interview is the distance it creates between the user and the researcher, which can make the data more reliable, even though a questionnaire cannot have the same degree of flexibility as an interview. The design of the post-test questionnaire for the pilot study is discussed in Chapter 4, and that of the post-test questionnaire for the main study in Chapter 6.

- ***Observation***

To give more credibility and strength to the outcome of our research, observations of all the experiments were carried out in conjunction with the questionnaires, and the data were included in the overall results. Evaluation by observation is a way of obtaining data in relation to user behaviour in user interface interactions. The observations were made at the University of Strathclyde where the experiments were recorded in video for analysis at a later date.

3.7 Ethical issues

The University of Strathclyde ethics regulations were followed during this study, with consideration being given throughout the research to the various ethical principles. The approval of the departmental ethics committee was obtained before any data was collected. Ethical considerations are a key part of any good research design, as they can influence decisions on methodology, and enrich the research. Prior to any participation in the study, each user was provided with a consent form, which contained the study aims and objectives, plus full details of what was to be expected of them, the tasks they would be asked to carry out, and what kind of information would be collected. The form also contained details of the plan to record and photograph the experiment sessions, and the users were assured of their anonymity and the confidentiality of any information, and that participation was totally voluntary. It was also made clear that participants could withdraw from the study at any time without reserve, and that they were not obliged to answer any question or questions, if they did not wish to do so, without giving any reasons.

3.8 Ethical and moral implications of the robot

Intelligent machine systems are spreading in the civilian world and transforming our lives for the better. They become more capable as they become smarter and more widespread. As that happens, they became sensitive topics, because of (The Internet of Things) which revolves around data-gathering from humans and data sharing by machines. Thus, this should be accompanied with machine ethical and morals. People have been thinking and discussing and writing about ethics, yet one of the known guidelines in robo-ethics are the "*three laws of robotics*" coined by Isaac Asimov (Gunn, 1996), a science-fiction writer, in 1942. The laws require robots to protect humans, obey orders and preserve themselves. Though these laws were simple and few, it was demonstrated by him, how difficult to apply them in various real-world situations.

As a designer, this law was considered when designing MobiBot. MobiBot system was design to follow users' vocal stream in order to enhance the interaction between both sides. In general, it will act accordingly to user inputs.

3.9 Summary

This chapter presented the research approaches, and the agenda for the research, reflecting the stages followed in conducting the study in order to answer the research questions. Finally, this chapter also presented the research strategy and methodologies.

Chapter 4 EXPLORATORY STUDY

This Chapter presents the exploratory study by focusing on conversational engagement behaviour as the conceptual framework of the study. Followed by the methodology used to conduct the study. Finally, the chapter concludes with analysis and discussion of exploratory feedback and finishes with a conclusion.

4.1 Introduction

Telepresence robotics overcomes some of the issues found in different remote communication platforms. It gives remote participants the ability to interact outside rooms, hence it extends their physical presence is as well as their communication zone. However, their social presence is not as effective as the physical one. Thus, social presence is a concept which can be used to explain, and even to predict, occurrences of the effects of the communication media on the outcomes of tasks.

Chapter two identified nonverbal behaviour as the element that contributes to social presence and is essential in order to design an effective telepresence experience. This chapter elaborates primarily on the specifications, challenges and development processes involved in the translation of requirements into a physical embodiment. The sequence is: a first study is conducted in order to test the proposed prototype from the users' perspective, and the results of the first study are then used to improve the system to meet the users' need for 'human-like' characteristics in a social robot. A conceptual framework is presented in Section (4.2). The engagement role is explained in Section (4.3), and we identify the users' conversational engagement profile. The study overview is discussed in Section (4.4), with technical equipment used in this study in section (4.5). The tools used in section (4.6), (4.7). The pilot testing presented in section (4.8), its data analysis in section (4.9) and its result in section (4.10). Followed by data analysis for the exploratory study in Section (4.11) and its results in section (4.12). Followed by a discussion section in (4.13). Finally, the chapter concludes with analysis and discussion of exploratory feedback in Sections (1.7) and (1.8) respectively and finishes with a Chapter conclusion in section (1.9).

4.2 Conceptual framework

One of the early theories that drove the recent research about social presence can be traced back to the end of the 1960s; Mehrabian's (1968) concept of immediacy, which he defined as "*those communication behaviours that enhance closeness to and nonverbal interaction with another*" (p. 302). The emphasis on interactive behaviour is followed by a more recent definition, as in Palmer's (1995) definition of social presence "*effectively negotiate(ing) a relationship through an interdependent, multichannel exchange of behaviours*" (p. 291).

These definitions emphasise reaction and interactivity as the essence or indicator of social presence. They include implicit or explicit references to some level of behavioural engagement, as they propose that an increase in the social presence is related to an increase in the engagement. Providing a remote participant with adequate nonverbal behaviour has the possibility to increase their level of engagement, as we have stated before, and contribute to the possibility of making the experience more natural.

Based on these theories and definitions, we believe that in order to improve the communication with a TP-robot, or in general in mediated communication, we need to create an engagement experience. This would typically involve a mixed-initiative, and well-coordinated process that includes non-verbal cues and signals, such as gaze and mutual attention, head and hand gestures, and verbal greetings These non-verbal cues, in turn, would lead to more effective and immediate interactions. Thus, a robot needs to simultaneously exhibit competent behaviour, convey attention and intentionality, and handle social interaction. Exhibition of naturalistic behaviour and appropriate emotions by the robot is the main core of an effective system as suggested by different studies (Rousseau et al., 1997, Bates, 1994).

As experiments with early TP robots have suggested, it is essential to understand human social aspects and human psychology (Paulos and Canny, 1998). This understanding is fundamental to increasing the overall functionality of a TP-robot. For example, understanding the arm and hand mechanisms within human interaction will lead to an improvement in prototyping simple gestural mechanisms within a TP-robot. Thus, we decided to start our research by understanding the engagement processes within human to human interaction as an essential basis for understanding human to robot interaction; we view this approach as a valid means to test our theories about engagement as well as to produce useful technology results.

The next section will explain engagement and the associated behaviours in more detail. We will also, cover some of the recent projects on increasing social presence in TP robots.

4.2.1 Why engagement?

We have identified that engagement counts as one of the most important factors affecting the sense of social presence. This can be seen in various researchers, which added it to their assessment for social presence. Schloerb (1995) proposed the assessment of social presence, by assessing the observers' interaction within a mediated environment. Also, Freeman and Avons (2000) stated that there is a positive relationship between social presence and engagement for the observers in a mediated environment. He based his thought on Freeman and Avons.

> "*When an observer is presented with a high-fidelity reproduction of the physical world a compelling perception of being in the depicted scene is elicited. It is the perception which define as presence, a sense of being there in a display scene or environment*" (Freeman and Avons, 2000, p. 1)

In order to achieve or improve engagement, we studied different nonverbal cues that are delivered in face-to-face communication and analysed the way people evaluate the meaning behind the spoken words. Tones, postures, gestures, head movements, eye gaze, and pointing or in general non-verbal cues in communication play a significant role and add more information or meaning to the conversation. In a recent study about social robots, Tojo *et al.* (2000) suggested that adding non-verbal channels like facial expression, pointing and posture to a robot will improve the level of conversational turn-taking. More natural behaviour and more natural utterances will result than when in conversation with a static robot.

4.3 Characteristics of engagement

The process of establishing, maintaining and ending the perceived connection between two or more participants during interactions they jointly undertake is what is called engagement by (Grosz and Kraus, 1996). Sidner (2004) described it as collaboration activity, where participants try to maintain the connection between them by various means, besides the conversation that supports them, and then ending the engagement or opting out of it. This claim implies that engagement is not just a part of the conversation.

To establish any connection between people, engagement will be part of the collaborative process that occurs, so it counts as a natural social phenomenon of human existence.

It has been said that it is completely possible to engage another, and maintain the engagement process, without conversation but not without communication. Relying on gestural language is a way to establish and maintain the engagement as a joint goal. In face-to-face communication, non-verbal behaviour plays a significant part where conversations are present, which will be discussed in more detail in the next section.

4.3.1 Behaviours Indicating Engagement in Human Interaction

Engagement is supported by three kinds of behaviours as stated by Sidner,Lee and Kidd (2005). These are the spoken linguistic behaviour (the use of conversation), collaborative behaviour (ability to collaborate on a task) and nonverbal behaviour that convey the connection between the participants. Nonverbal behaviour is one of the fundamental factors in human interaction; in face-to-face interaction at a near or even far distance, nonverbal behaviour is the means of communicating beliefs, intentions, and desires. As McNeill (1992) has stated; conversation and nonverbal behaviour are naturally co-occurring in most human-human encounters, they are tightly intertwined in human cognition. Properly engaging oneself in a conversation requires one to have internalised how to deal with the protocols and techniques that have evolved in human society, and how to turn the result into linguistic action. For example, if a speaker continues to talk while looking away from the hearer(s) into blank space for a long time, this conveys contradictory information about not being interested in the subject of the talk.

We are mainly concerned with conversational nonverbal behaviour in face-to-face conversation, as this provides significant evidence of a connection between the participants. In addition, it is the first step towards building a realistic human-like companion with rich visual expressiveness. Conversational nonverbal behaviour can indicate different contents as Sidner *et al.* (2005) explained.

Some nonverbal behaviour supplements the content of utterances in which the gestures occur, especially those using the hands (Cassell, 2000). Whereas, head movement, eye gaze and those involving body stance and position indicate how the conversation is proceeding, and how engaged the participants are in it.

- ***Head movements***

As we have stated before, in face-to-face conversation, people involved in the conversation move their heads in typical ways. Various patterns have been discussed by many researchers from a variety of disciplines, and suggest that these movements have many functions and are determined by many variables (Ray L Birdwhistell, 2011, Hadar *et al.*, 1983, Iwano *et al.*, 1996, DeCarlo *et al.*, 2004, Hadar,Steiner and Rose, 1985b, Pittenger,Hockett and Danehy, 1960, Graf *et al.*, 2002, Rosenfeld, 1978) (Table 4.1).

Table 4.1. Head movement patterns

Pattern	Function	Study
Nodding head once with ‘ordinary’ speed.	A listener might be communicating polite involvement.	Birdwhistell, 1970
A rapid double nodding.	A listener might be communicating real interest.	
A very rapid triple head nod.	A listener might be communicating impatience, and indicate to the speaker to discontinue.	
A still position.	Associated with pauses by speaker and listening.	Pittenger et al., 1960; Birdwhistell, 1970
A vertical movement of head.	Affirmation, agreement and giving responses by listener.	Iwano, et al.,1996; Heylen, 2005
Speaker faces up to see partner.	When the speaker wants to get a response from the listener.	
A head nod by listener following a direct question.	Most probably means 'yes’.	Birdwhistell, 1970; Rosenfeld, 1978; Heylen, 2005
A head nod by listener in the middle of the other's sentence.	A signal of impatience.	

A short and intense 'rapid' head movement made by speaker.	Associated with stress in the words being said.	Hadar et al., 1985; Graf et al., 2002
A rapid head movement.	Accompanied by primary peaks of loudness.	

We can see that simple behaviours such as head movement can have many and very complex functions, which are determined by many variables, which goes to highlight how important they are. Therefore, we proposed to focus on measuring engagement within conversational behaviours, as they provide significant evidence of a connection between the participants as supported by different studies (Sidner *et al.*, 2005, Cassell, 2000).

4.4 Comparative study: Simulated vs. Static

As we have previously stated, telepresence in the context of this application means the replacement of a human presence with a robot, which is operated by a human driver from a location at a distance. Increasing presence is crucial in designing an effective robot system, thus it becomes the goal of most research projects in the area of telepresence robots. A review of the literature on interaction engagement and communication behaviour indicates that nonverbal behaviours play an important role in message production and in engagement in a variety of different situations.

Two different video call conditions were tested, in order to evaluate the effectiveness of the replicated head movement as an interaction method with such a system, compared to video-mediated static. The main objective was to evaluate the users' engagement in two conditions. A within-subject design was used, with twenty-six representative users participating in the study, which was carried out in the Private room in the University of Strathclyde.

4.4.1 Methods

- ***Participants***

The participants took part in the experiments in pairs; with each experiment involving up to 13 pairs, making a total of 26 subjects. 3 pairs were from different genders, and the rest were from the same gender (3 female pairs and 7 male pairs). The total number of experiments was 26, 13 with each condition. Each pair consisted of a Picker and a Guesser. The ages of the subjects varied between 18 and 40, and their subjects represented a variety of university majors. Subjects should be able to speak English fluently was the only inclusion criteria, whereas the exclusion criteria were subject with hearing or visual impairments and Nonfluent English speakers.

- ***Recruitment***

i. Sampling frame

A nonprobability, convenience sampling strategy was employed for data collection, as subjects were selected because of their convenient accessibility and proximity to the researcher.

ii. Sampling techniques

After supervisor and department confirmation, an advertisement was distributed electronically using the advice centre newsletter email. The email included details of the purpose of the study, and information regarding participation and what this would entail. Besides this, hard copies of the advertisement were posted around the university campus to ensure as large a number of volunteers as possible. In order to overcome the bias resulting from sampling and coverage, the sampling strategy ensured the distribution of the advertisement to cover different university levels, and different subjects were included.

iii. Sample Size

The purpose of the study was to investigate the influence that telepresence has on communication, by examining the sense of engagement as a part of social presence which can be improved further by adding more non-verbal cues and could be useful in the development of a model for human social behaviours to implement in a telepresence platform. Therefore, there was no defined number of participants to be recruited, rather the researchers aimed to identify and recruit as many as were willing to participate. In total, seventy-six subjects were recruited.

4.4.2 Study design

Our overall position was that making telepresence systems socially expressive, by affording them the ability to convey their operators’ non-verbal behaviours such as gesture and posture, could make remote interactions more present, and more engaging.

As we specifically wanted to learn about how head movements affect the collaborator's experience, we decided to run a quasi-experimental study.

This type of experiment utilises a similar structure to experimental design, however, it lacks the key ingredient which is random assignment and blinding/ it did, however, help in evaluating the experience of the subject when interacting with a socially expressive system. This quasi-experimental study used a one-group in two-condition design, to give a better understanding of how head movements might affect conversational engagement.

i. The simulated condition

Participants interacted through a video call, where the face and head movements were represented on the screen, and the screen itself replicated the head movements of the person on the screen.

ii. The video-mediated condition

The screen was in neutral and still (non-moving) poses during the whole interaction, and participants interacted through a video call where the face and head movements were represented on the screen only.

This study involved a single questionnaire, with repeated measurements were taken twice in two conditions on one group of subjects. After performing the two tasks in the two conditions, the subjects answered the set questionnaire.

Our experimental design involved one manipulated independent variable, which was whether or not the screen produced the head movement, and physically nodded. This variable was generated by two different scenarios and conditions as previously stated; video-mediated on-screen movements and video-mediated on-screen and in-space movements. The dependent variables involved both self-reporting and observational measurements during the completion of a task.

4.4.3 Pilot study task

At the beginning of the research, we decided to carry out our research in the education field, and we thought that using a creative problem-solving task would be the best choice. Our choice was based on the fact that students need to stimulate their creative thinking during their study.

Although the important effect that visual signals have on communication outcomes has been supported by different studies, a conflict in the task outcomes has been highlighted. Some studies found that for problem-solving tasks, face-to-face, and audio-only interactions do not differ in terms of task outcome (Chapanis *et al.*, 1972) (Williams, 1977). In contrast, with design tasks and social tasks involving negotiation or conflict resolution, performance was better in face-to-face or mediated communication than audio-only (Olson,Olson and Meader, 1995, Fish *et al.*, 1992, Short,Williams and Christie, 1976). Therefore we decided to go with a study done by Doherty-Sneddon *et al.* (1997) who examined communication and task performance in face-to-face, copresent, and video-mediated communication (VMC). They used the Map task in their study, and found that high-quality VMC did not appear to deliver the same benefits as face-to-face, copresent interaction. Thus, we decided to carry on the same experiment with the same task to examine if we could improve the VMC using nonverbal behaviour. For this study, we used the same maps as for the Anderson *et al.* (1991) (The HCRC map task corpus), as it was intended to study the dialogue generated during this task (Figure 4.1). Participants worked in pairs, with one person acting as the Instruction Giver, and the other as the Instruction Follower.

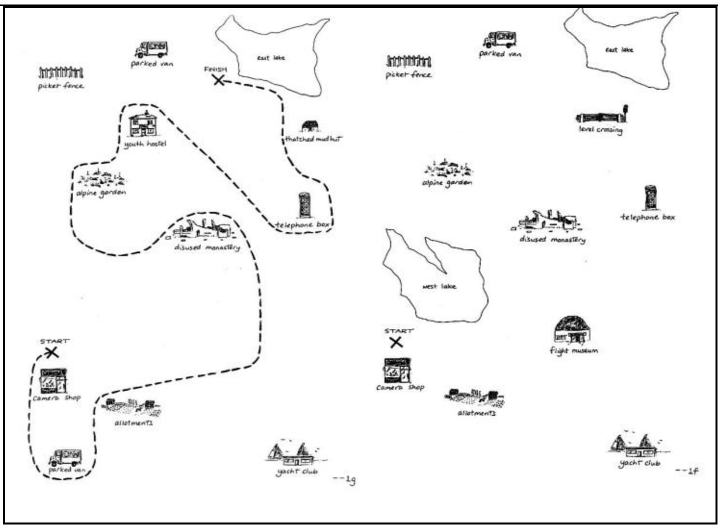


Figure 4.1. Map from (The Design of the HCRC Task Corpus)

4.4.4 Adapted task

In order to avoid the effect of the issues with the original task, we decided to switch to something more enjoyable, and something that the subjects had more in common with each other. Therefore, we decided to start with a study done by Mutlu (2009), who examined communication and task performance in human-robot communication using a guessing game. We chose to use the guessing game task because it has the following characteristics which are beneficial to our evaluation:

- It is intended to allow study of the dialogue generated during the task
- It sparks an active conversation between both sides
- It takes under about thirty minutes to finish

However, the format was adapted according to our context of use:

- In our experiment, we gave the Picker the opportunity to explain the item he/she had picked, as we were looking to have a dynamic interaction between both sides, which would be difficult to achieve if we used the original game design.
- In the original game design, both participants had all the items in front of them as they were facing each other. This situation was not part of our experiment design, and therefore we placed the items in front of the Picker only.

We devised an experimental task in which two participants were to play the guessing game. In the game, one of the players (the Picker) would choose an item - without identifying it to the other player – from among eighteen items (printed onto A5 size cards) placed in front of him/her. The other player (the Guesser) would try to guess which item the Picker had chosen by asking the Picker a set of questions that could be answered with “Yes” or “No”. However, the Picker could give the Guesser some hints if they wanted to do so. By this method, both sides would have an equal chance to interact and exchange words with each other (Figure 4.2).

For the purposes of the task, we were careful to select items that would be in common use in the UK, in a balanced set of materials, shapes sizes and colours. Both participants were provided with detailed instructions on how the game should be played.

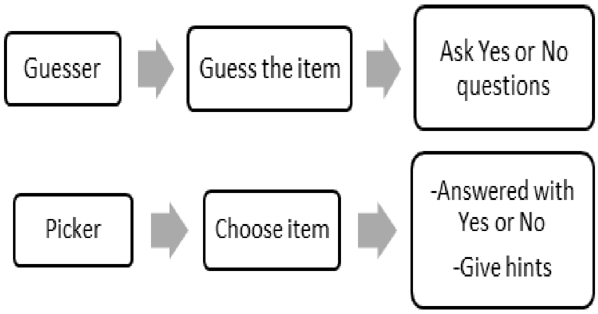


Figure 4.2. Experiment Procedure

4.4.5 Procedure

The purpose of the experiment was to gather real-time data under two different conditions: video-mediated and video-mediated with in-space screen movement. At the beginning of the session, we provided the participants with a brief description of the purpose and procedure of the experiment, but we deliberately concealed the primary purpose of the experiment. The participants were given an introduction on how the experiments would operate; and they were instructed how to play the Guessing game, allocated a role as a Picker or Guesser, and told what they needed to do. After that, we took the participants into the experiment room to start the task. We then asked the participants to fill in a pre-experiment questionnaire on their affective state and some background data, such as gender and age etc. (Appendix A).

Each pair completed two Guessing games under different conditions, which lasted approximately 15 min. We focused only on the Guesser side as it communicates more than the Picker side. The Guesser has to ask questions, give an explanation and give answers, where the Picker side has only to say “Yes” or “No” and give hints. Guessers sat in front of the modified rig that held the 12” monitor with a small web camera affixed to the top. Following the game, participants were given a 5 to 10-minute break to complete the Interaction Involvement Scale (IIS) questionnaire. This questionnaire took about 5 min to complete.

1. <u>The simulated condition</u>

For the mediated conditions, subjects were not able to see each other physically, but could work collaboratively together and were able to hear and talk to each other without wearing headsets. A screen was erected between them, adjusted to block any direct view of one another's faces, but they could see one another's faces through the webcams placed on top of each screen. Both sides had a camera set up to record their interactions.

2. The video-mediated condition

This was similar to the first condition; however, the researcher took a seat beside the screen displaying the head movements of one side, and reproduced those movements on the other side's screen as closely as possible.

The procedures were the same under both different conditions; the researcher assisted each subject with the placement of the webcam image, ensured that the subjects were comfortable, and then asked them to begin the test. For both conditions, the researcher monitored the environment, to ensure that everything was working as planned.

The sessions took place in two rooms, one equipped with a laptop only, which was where the Pickers were sitting. The second room was the experimental room for the Guesser, where we place a modified rig for the 12" monitor and video camera (Figure 4.3).

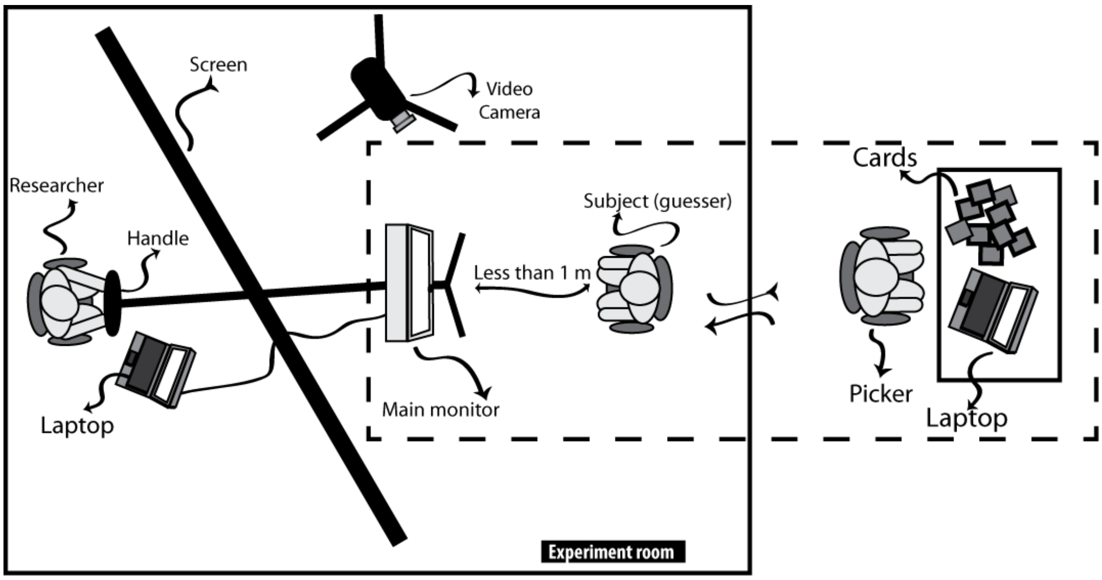


Figure 4.3. Experiment Design

4.5 Technical equipment

Based on the previous findings, we developed a model, which is a monitor on a stand; the design process for it is described as follows.

4.5.1 Design process

- ***Abstract 3D design***

The design process started with a 3D design for the expected model (figure 4.4). This consisted of a vertical base attached to a horizontal pipe. One end of the pipe held the screen monitor and the other end was a handle to deliver the chosen movement. The screen monitor base was attached to a spiral spring, thereby allowing the movement of the monitor to be extended up and down.

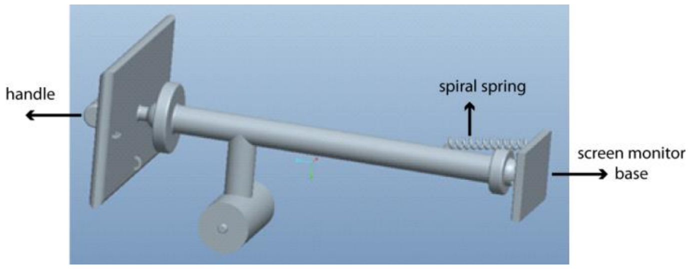


Figure 4.4. Abstract 3D design for the Model

- ***Skeleton Prototype***

Computer scientists and designers often use a skeleton prototype to try out ideas for the design of a robot, prior to committing to the final design in respect to finish and exterior shaping. A scale model is also often used, and these can be useful for prototyping movements and behaviour which may not be possible for a full-sized robot. Therefore, and before building the full-sized model, a skeleton prototype of our model was used to closely mimic the movements in physical space. Such prototypes can be quickly and easily produced using woodworking methods, and can be assembled in a matter of days, with little waste of material, as can be seen from (Figure 4.5). Using the Puppeteer mechanism in the monitor's stand, which is basically a way to bring inanimate objects to life by mimicking actions with simple hand movements, the researcher was able to manually replicate the head movements for the person on the screen.

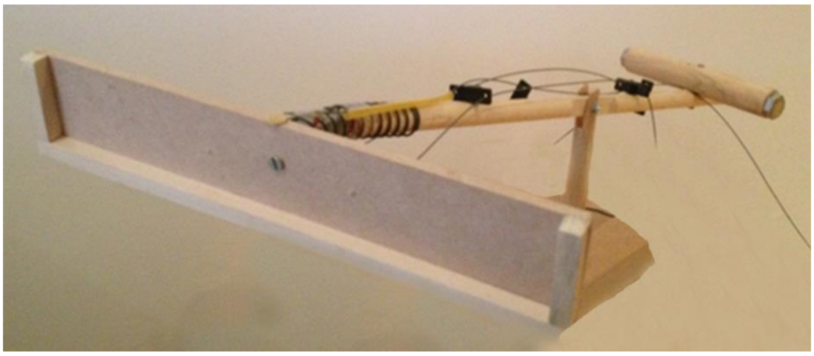


Figure 4.5. Skeleton Prototype

4.5.2 Wizard of Oz

A method for dynamically experimenting with and prototyping a system, but using human manual intervention instead of automation is known as WoZ, or the Wizard of Oz technique. This methodology was initially developed by HCI researchers, working in the realm of natural language and speech interfaces (Kelley, 1983). They used it as a way of understanding how systems should be designed, prior to the maturation of the underlying response generation or speech recognition systems. In contrast to the original use, the WoZ method is particularly useful in the field of human-robot design and interactions (HRI), (Maulsby,Greenberg and Mander, 1993, Riek, 2012). The aim is not to make-up immature technologies, but to use WoZ as a means to explore the many and varied possibilities for the design of movements and behaviour of robots, without the complications of automation. The WoZ technique often involves users, who have a major part to play in the exploration and collaborative design, in real-time.

Based on the previous model, we designed the following model (figure 4.6). Using Wizard of Oz (WoZ) technique, the researcher was able to replicate the basic movements of the head. These movements are nod -up and down movements- which are used to show agreement with what is being said, and shake –left and right movement- to express disapproval and negation.

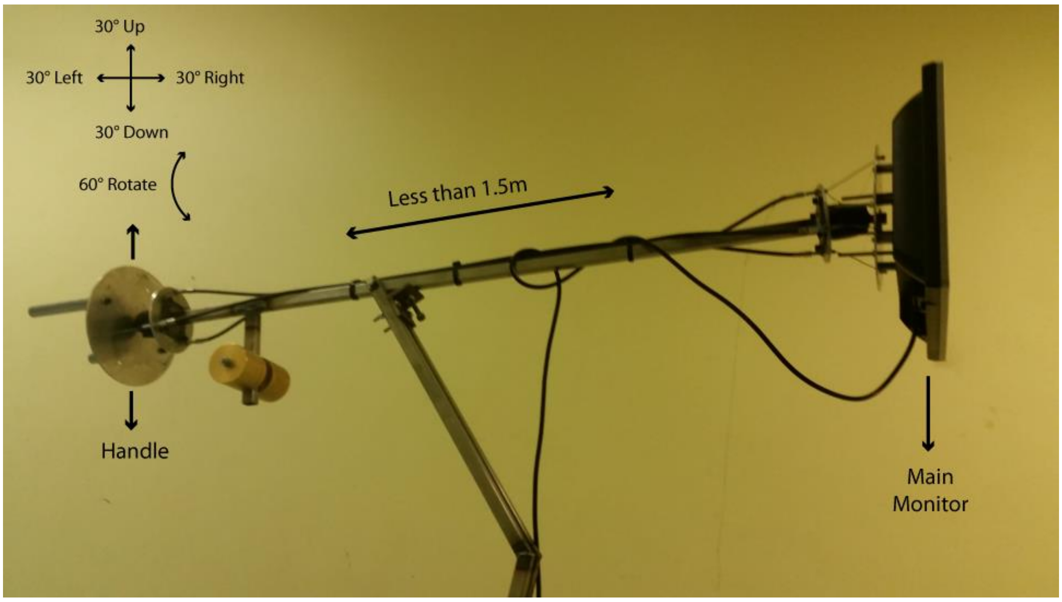


Figure 4.6. Resulting model

4.5.3 Hardware

- 12" screen: used in the study for all participants in the second condition; mounted behind a black curtain.
- Video recordings: a digital video camera recorded the Guesser side in all conditions, and was placed 1 metre away from the speaker in the recording studio. This was used to record the dialogue and the interactions of the participants, and covered three subjects (screen, body and face and the task document).
- Laptop: 1 laptop was used for video call conditions (first condition) from the Picker side.
- High-definition video webcams were mounted at the top of the screens for both conditions.

4.5.4 Software

- Skype application was used for the video call.

4.5.5 Tools

- Questionnaire and video recording.

4.6 Development of the questionnaire

4.6.1 Conversational engagement

One of the challenges in our study was to decide how to measure the impact of the in-space movement on the participant's interaction. Several studies have looked at how the structure of conversation changes with communication mode (Mark Cook and Lalljee, 1972), (Rosenfeld, 1978). Measures of conversation structure include some nonverbal behaviours such as the number and length of speaker exchanges, and the number of pauses and interruptions in a free speech (Hadar,Steiner and Rose, 1985a). A common finding in the literature is that face-to-face conversations result in more turns, shorter lengths of turn and more interruptions than audio-only or video-mediated dialogues (O'Conaill,Whittaker and Wilbur, 1993, Sellen, 1995). The interpretation of these findings has been that nonverbal behaviour or visual signals are important, thus face-to-face communication is less formal, with more interruptions (simultaneous speech) and fewer formal handovers of turns (Beattie and Barnard, 1979, Ellis and Beattie, 1986, Rutter and Stephenson, 1977). The underlying assumption behind these differences may result from technical limitations, or a lack of nonverbal behaviour as Short, Williams and Christie (1976) highlighted that the lack of head turning and directional gaze in many video-mediated communication (VMC) systems may affect turn-taking behaviour.

Therefore, it is to be expected that systems that are careful to preserve important non-verbal cues will make a difference in overall interaction behaviour. A related question is the extent to which the media difference will influence participants' interaction involvement. Recent studies suggest that socially expressive media can significantly improve viewers' interpretation of the action, and it can be seen to be more engaging and likable than a static one (Adalgeirsson and Breazeal, 2010, Sirkin,Ju and Cutkosky, 2012, Mutlu *et al.*, 2009). It is conceivable that the media could have some observable influence on cognitive activity such as interaction engagement.

Conversational engagement has not attracted much scholarly attention, and few studies up to now have made a deep level of investigation into this field to extend the way in which engagement is measured. This is particularly the case in conversational engagement in conjunction with a telepresence robot.

4.6.2 Measurement of conversation engagement

Once we had identified the factors, we needed to decide how we could assess them. As previously stated, Biocca & Harms (2002) explained that we needed to go beyond technology assessments, and more into the realms of psychology and sociology. We, therefore, decided that the project would involve a variety of means of information gathering and experiments, using questionnaires and video recordings of practical interaction experiments.

- ***Observational rating***

The interest in interpersonal interaction has led to the development of different observational rating systems of nonverbal engagement, the most well-known and established ones are (Coker and BURGOON, 1987, Laura Knarr Guerrero, 1994, Laura K Guerrero, 2005). The Guerrero system was developed using Coker and Burgoon items as a guide to measure specific engagement behaviours that could be rated by coders. We have adopted the use of this system in our study, as it the most up-to-date version of the various rating systems.

The Guerrero observational rating system (2005) was developed to rate behavioural cues of engagement in human dyadic interaction, and thus, we argue, it can also be used with human-mediated interaction. The system comprises six scales that are necessary for measurement of engagement. These are; immediacy, expressiveness, alter centrism, interaction management, composure and positive affect. (1) *Immediacy* dimension behaviours measure the physical proximity between two individuals; (2) *Expressiveness* dimension behaviours communicate the level of energy, activity, and enthusiasm toward the conversation partner; (3) *Altercentrism* dimension behaviours reflect the degree of focus on the conversation partner during interaction; (4) *Interaction management* dimension behaviours support a smooth flow of conversation; (5) *Composure dimension* behaviours reflect an absence of nervous body movement or the presence of confidence; and (6) *Positive affect* dimension behaviours include smiling, laughing and other behaviours that reflect good feelings about the interaction and the partner. If an individual is showing a greater immediacy, greater expressiveness, better interaction management, more altercentrism, lack of concern about others, greater negative arousal and more positive behaviours, it can be reasonably assumed that that individual is highly integrated in their feelings, thoughts, and experiences with the on-going interaction - a highly involved individual.

We believe that the Guerrero (2005) system is uniquely suited for measuring a wide range of nonverbal and verbal indicators, to determine the degree to which an individual is actively involved in a real-time conversation, as in our case. Another study (Norris *et al.*, 2014) added that although the system was designed to include six different dimensions, it could be altered to use particular dimensions according to the focus of the research, without impacting measurement reliability and validity. This can be seen as an advantage in our present study where the immediacy dimension is not applicable because of the nature of the game technology used in our research; there is no physical contact between participants. In addition, this type of method needs at least two people to measure the reliabilities of the findings, which might not be affordable. To overcome the issue of needing an extra opinion to achieve reliability, we decided to focus on the quantifying behaviours only. Hence, we examined gaze within immediacy, nods within alter centrism, smiling within positive affect and lack of random movement within composure.

Based on the previous section the following hypotheses were generated

Simulating head nodding in a telepresence conversation will increase engagement communicated by more eye gaze (H1), more altercentrism seen by more nodding (H2), positive feeling between both sites can be seen by more smiling and laughing (H3), smaller number of answers (H4a) and questions (H4b), and finally, composure by lack of random movement (H6).

- ***Self-reporting***

Cegala (1982) conducted one of the early studies around interaction engagement, and identified three dimensions of engagement: responsiveness (that is, mental alertness to the situation), perceptiveness (that is, ability to make attributions about one's or others' behaviour), and attentiveness (that is, awareness of factors impacting interaction).

This questionnaire was chosen because it is widely regarded as the most relevant and comprehensive instrument available to assess individual's personal tendencies, relative to engagement in communication settings, and it is also used extensively in literature (Duran and Kelly, 1988, Sidelinger *et al.*, 2008, Norris *et al.*, 2014).

As mentioned earlier, the interaction involvement scale (IIS) consists of attentiveness, responsiveness, and perceptiveness. Although each of these factors can be examined independently, an overall score can be gleaned in an effort to assess one's overall tendencies toward interaction. To our knowledge, few studies have made a comparison between conversation engagement measures (self-reporting and observational rating measures), especially when a cross situation comparison is made between two contrasting settings such as the ones under study here. Therefore, it is of interest to compare conversational engagement of participants (objective measure) and their subjective ratings of satisfaction about engagement in the task. Based on this, we hypothesise that a high score in conversational engagement (objective) will be associated with simulating head nodding in a telepresence conversation (H7).

In general, our experimental design involved one manipulated independent variable; whether the screen produced movement or not. The dependent variables involved objective and subjective measurements.

4.6.3 Data Collection

Subjective measures were used to obtain ratings of the subjects' satisfaction in their engagement in the game, using the interaction engagement scale, and to evaluate the appropriateness of the movements. This was used to answer H7.

i. Engagement

Upon completion of the task, the subjects were required to complete a post-test questionnaire, which was a version of the interaction engagement scale (Appendix A).

Response options for each IIS item range from 1 (not at all like me) to 5 (very much like me). Responses to some items were reverse coded, and high scores refer to high interaction engagement with communication.

ii. Effectiveness of movements

Part of this questionnaire was a version derived from Sidner,Lee and Kidd (2005) (4 items). This measure was given to the participants at the end of the second session, as it measured the appropriateness of the head movements. Response options for each item ranged from 1 (Disagree) to 3 (Agree).

In addition, we used open questions to gather information about our subjects' experience (Appendix A).

Objective measures were used to obtain behaviour during sessions. We videotaped the sessions for both conditions, and reviewed the recordings, transcribing them under a number of headings as discussed earlier (section 2.1.2), to determine interaction behaviour during the sessions.

i. Behavioural observations

We measured participants' task performance through capturing the number of questions they asked and the answers they gave in order to identify the Picker's choice. We also measured the engagement score for all Guessers for the two conditions. As discussed earlier, all sessions were videotaped to support the analysis of the objective measures (Appendix A).

4.7 Psychometric properties

Any tool should show basic psychometric attributes in order to be statistically acceptable. Addressing validity and reliability analysis will give an idea about the weakness or the strength of the measure.

4.7.1 Reliability

This is a reference to the internal consistency, stability or repeatability of a questionnaire (Jack and Clarke, 1998). Cronbach's alpha is a statistic that is most often used to represent reliability: it uses internal correlations to show whether items in a questionnaire or survey are measuring the same construct (Bowling, 1991, Bryman and Cramer, 2005, Jack and Clarke, 1998). If the items are demonstrating a good level of internal consistency, for a developing questionnaire, a Cronbach's alpha value of over α 0.70 would be expected, and for a longer-established questionnaire, α 0.80 (Bowling, 1991, Bryman and Cramer, 2005). Common interpretations of α values are presented in (Table 4.2) below.

Table 4.2. Cronbach's Alpha value interpretation (George and Mallery, 2016).

Excellent	Good	Acceptable	Questionable	Poor	Unacceptable
α value >.9	α value >.8	α value >.7	α value >.6	α value >.5	α value<.5

4.7.2 Validity of the questionnaire

The term 'validity' is a reference to how appropriate, meaningful, correct and useful a researcher's inferences (Fraenkel,Wallen and Hyun, 1993). In this questionnaire, selected items had been used previously in other studies which had proved to have a good degree of validity. Nevertheless, testing the questionnaire before starting the process of data collection is a fundamental step (Boynton and Greenhalgh, 2004, Oppenheim, 2000). The section below describes the procedure followed for testing the questionnaire in this study.

- ***Face validity***

Face validity is understood to be one of the weaker categories of validity, but it is of importance, as good face validity can help to ensure a good rate of response. Our test for face validity entailed a review of the overall appearance of the questionnaire, and making sure that the questions appeared to fit a logical sequence. The wording, sequencing, clarity, suitability and lack of ambiguity were all checked as part of this test (Brink,Van der Walt and Van Rensburg, 2006).

- ***Content validity***

This is concerned with the content, and defines that it should be logical and comprehensive in relation to the different domains covered by the questionnaire. Two lecturers and a technician from the university carry out this check. We had drawn and adapted the major portion of our questionnaire from existing questionnaires in our field of study, so further refinement of the items was required. A further test of content validity was therefore undertaken by three Ph.D. students and colleagues, who were able to tell us whether the questions easy to comprehend, clear and appeared to be in a logical sequence (Brink,Van der Walt and Van Rensburg, 2006).

4.8 Pilot testing

A pilot study allows the determination of the process by which the feasibility of the main study stages can be assessed, and supports the management of users and data optimisation issues. The resources can also be specified, ranging from time to budget to the technical aspects and challenges, which can be pinpointed prior to the main study (Thabane *et al.*, 2010).

A pilot study was undertaken in the current research as an initial stage in the system development. The aim of this pilot study was to examine the feasibility of the study design in the proposed research environment. The objectives were:

1. Developing and testing the adequacy of research instruments
2. Identifying potential difficulties which might occur using proposed methods

3. Assessing the proposed data analysis techniques to uncover potential problems

4.8.1 Pilot study sample

Representative subjects were involved in the pilot study, as recommended by Thabane *et al.* (2010). A total of seven users were invited to participate, and provide an initial evaluation of the system. Our subjects in the pilot study were asked to interact using English to enable us to carry out later video analysis, and to avoid any complications might affect the experiment outcome.

4.8.2 Data analysis

The analysis included demographics, hard copy questionnaires completed after the two conditions of the experiment, and video data collected during the experiment.

4.8.3 Task

Map game as described earlier. All maps consist of landmarks or features, labelled with their designated name. For our task, there were two maps, with a number of common features present on both maps, and also a number of different features. Features which were not common differed in one of three ways:

1. They were present on only one participant's map, or they were in differing locations.
2. Name Change features were identical in form and location, but had different labels on the two maps.
3. Some features appeared twice on the Instruction Giver’s map.

Therefore each map would differ within each pair, which provided opportunities for communicative difficulties to be overcome as stated by Doherty-Sneddon *et, al.* (1997). The differences were equivalent across all the map pairs, and the maps were A4 size with the landmarks portrayed as line drawings. Each map had a starting point, marked on both maps, and a finishing point marked only on the Instruction Giver's map. The start and the end point were adjacent to a common feature. Participants were made aware that there might be differences between the maps, and they were told that the Instruction Giver's map was the only one with the route marked. Thus, the task for the Instruction Follower was to reproduce the route on their map as accurately as possible.

4.8.4 Findings

Using a t-test, analysis revealed no significant differences ($p<.05$) on any of the measures we had used. However, the results of the pilot study did guide the required study design improvements. One of the problems that became apparent in the pilot study was that the video camera only recorded the subject's face, and missed off the monitor they were facing, and also the rest of their bodies. In the actual experiment, we set up the video recorder to cover three things; monitor, body, and face.

In addition, one of the things that had been reported by subjects in the pilot study was that the second task was a bit difficult to solve. Subjects had not enjoyed the task, as they were frustrated by the amount of difficulty. In addition, we noticed that subjects were tending to focus only on the map itself, and not making eye contact with their partner; they were focusing too closely on using the instructions they had been given to solve the task.

The subjects' continuous behaviours, such as smiling and eye gaze, were counted simply using electronic stop watches, whereas the rest of the behaviours were counted by the number of times they were exhibited, such as the number of questions and answers, nods and adaptors or twisting behaviours. We also noted that interaction length varied between 2 to 7 min due to player style and personal self-confidence; we found that shy, taciturn player's interactions were shorter than those of talkative players. This could have caused an issue with our data, as the score for the number of behaviours would have varied as well. To overcome this issue, we converted the score to the number of behaviours per 1 minute by dividing the score by the time taken which was also used as a method for the adopted guessing game task later.

4.9 Data analysis

The quantitative and qualitative data were analysed separately, and interpretations leading to answering the research questions were integrated. The following section explains our approaches to analysis of the data revealed in each phase of the current research.

4.9.1 Qualitative data analysis tool

The SPSS software package (version 19 for Windows) was used to analyse the questionnaire data.

4.9.2 Treatment of the missing data

The current thesis has adopted the method of 'mean imputation' in order to deal with the problem of missing data. This method, which is frequently adopted, replaces empty data point with a mean value.

A scan for missing data was carried out on each data set, and in cases where a subscale calculation was required by a construct, an imputed mean was inserted. The main benefit of this approach to missing data is that it does not result in a change to the distribution of the data, and is not reliant on guesswork – the disadvantage is that it can result in a reduction in overall variance (Tabachnick,Fidell and Osterlind, 2001), which is usually only an issue if there is a great deal of data missing, which was not the case with our data. The imputed means used to replace the missing data points were calculated from the data for each condition, rather than using the overall data mean.

4.9.3 Questionnaire

- ***Descriptive analysis***

In order to decide which type of statistical analysis is appropriate for the obtained data, it is important for the researcher to understand the main features observed in the data. Therefore, descriptive analysis was carried out as a first stage in all the analysis. Means and standard deviations were calculated for the questionnaire and video analysis data.

Many statistical tests also rely on an assumption of normality, so for this reason, a normality test was also carried out on the data. A normal distribution is one that has a symmetrical bell-shaped curve with a mean of 0 and a standard deviation of 1. In order to establish whether the data were actually normal, the Shapiro-Wilk test was performed with a null hypothesis that the sample did not deviate from the normal, which would be rejected with a p-value <.05

Skewness was also considered as part of the assumption of normality, which was measured by obtaining the z score, which is a measure of skewness divided by its standard error. A value between +1.96 and -1.96 at 0.05 level is judged to be significant statistically (Cramer and Howitt, 2004).

- ***Inferential analysis***

A comparison of two means from the within-subject test group was also made, using a paired *t*-test. In this test, the two means are typically representative of two related but different conditions. This test is intended to show whether statistical evidence exists between paired observations that for a particular result, the difference in the means is other than zero (Hsu and Lachenbruch, 2008).

4.9.4 Quantitative data analysis

A coding by themes was used as a method to analysis the qualitative data. By combining them under specific categories and then highlight the similar passages of text with a code label. This will make it easier to be retrieved at a later stage for further comparison and analysis.

4.10 Results

4.10.1 Descriptive analysis

The same group of users was used to evaluate the three objectives of this study. Table 4.3 describes the participants, and includes age ranges, numbers, gender, and proportions of responses. The participant ages were with a wide range of 18 to 40 (25.92 ± 5.86), with over one-third (31%) being aged between 18 and 20 and nearly one-third (27%) aged between 26 and 30.

Table 4.3. Characteristics of the users within the study

Characteristic	Participants	Percentage %
Age (years)		
18-20 years	8	31%
21-25 years	5	19%
26 to 30 years	7	27%
31 to 35 years	5	19%
36 to 40 years	1	4%
Gender		
Male	17	65%
Female	9	35%
Total	**26**	

4.11 Comparative study: Video Simulated Screen vs. Video-Mediated Static Screen

Two different video call conditions were tested in order to evaluate the effectiveness of the replicated head movement as an interaction method with such a system, compared to video-mediated static, from the representative users' perspective. The aim of testing both interaction methods was to examine the face channel, which contains some of the important nonverbal cues such as movements and expressions, focusing on the effectiveness of replicating head movement which includes nodding and head orientation.

Users were given the same tasks, but with different choices of guessing items, for each condition (simulated and static). The task allowed the user to interact with a partner, using the developed model (a monitor on a stand), which was felt to be adequate for the user to make decisions about each condition. Figure (4.7) shows the user interacting with the developed system.

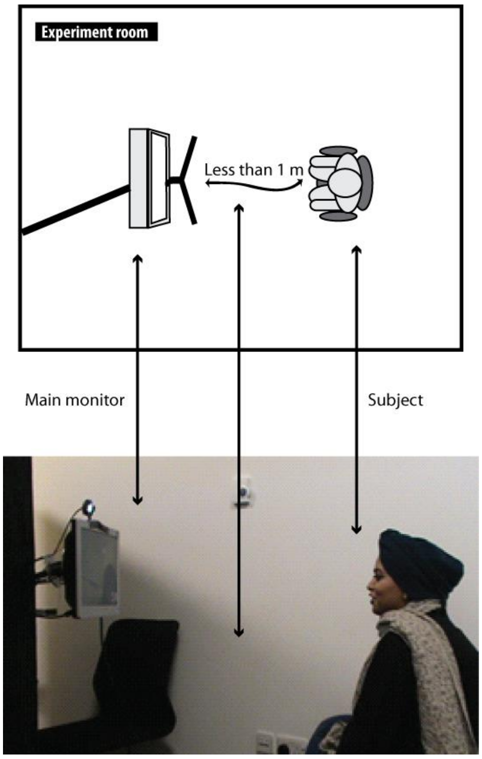


Figure 4.7. Experiment Setup

4.11.1 Preparation of engagement rating data for analysis

Behaviours were examined prior to testing the study hypotheses to search for any sign of systematic patterns consistent with errors or missing data. Based on this examination we found three different systematic patterns in our data. The first case had little or no incidences of head nods; adopters and twisting behaviours observed. We also faced a problem of missing data from the game scenario, some recordings did not provide a full view of the player, either because of the player covering the head with hands or scarf, or because the player leaned too far forward situating them partially or completely out of view of the video camera, resulting in no rateable nonverbal behaviour. As a result, we prevented these items from being used in further analysis for altercentrism H2 and composure H6.

4.11.2 Questionnaire design and evaluation

- ***Questionnaire design***

After conducting the trials, the users provided their feedback on a hardcopy questionnaire. The questionnaire included 23 questions, aimed at evaluating user's conversation engagement. For the purposes of comparison, the same questionnaire was used for each condition.

- ***Questionnaire evaluation: reliability***

Cegala *et al.* (1982) reported reliabilities Alphas of .84 for the scale of the interaction engagement scale, Rubin, Perse and Barbato (1988) reported .62 and Duran and Kelly (1988) reported .66.

To measure the reliability of the questionnaire, the mean of the questionnaire was assessed for reliability using Cronbach's Alpha. The test for reliability revealed good to excellent p-values of .874 in the simulated conditions and .778 the static condition.

4.12 Results and analysis

4.12.1 Normality checking

The Shapiro-Wilk test for the simulated conditions items (number of answers and eye gaze) returned 0.0 significant value and smiling had a significant value of 0.009. All of the p-values were <0.05, which means that the null hypothesis was rejected for those items and the data were not normally distributed. However, number of questions and conversational engagement (ENG) revealed a p-value >0.05, meaning that they were normally distributed (Table 4.4).

Table 4.4. Normality checking for the simulated conditions

Items	Shapiro-Wilk		
	Statistic	Df	Sig.
Number of Questions	.906	26	**.021**
Number of Answers	.810	26	.000
Smiling	.890	26	.009
Eye gaze	.812	26	.000
ENG	.981	26	**.902**

The calculation of the z-value based on the skewness and kurtosis values showed different outcomes. The z-value of all of the items was between -1.96 and +1.96 except for eye gaze which returned a value of 3.92 (Table 4.5).

Table 4.5. Calculation of the z-value for the simulated conditions

Item	Mean		Std. Deviation	Skewness		Kurtosis		z-value
	Statistic	SE	Statistic	Statistic	SE	Statistic	SE	
Eye gaze	46.1950	3.59204	18.31586	1.792	.456	4.077	.887	3.92
Smiling	2.3278	.35416	1.93983	1.583	.427	4.060	.833	0.00
Number of Answers	1.4555	.21328	1.16818	1.509	.427	1.835	.833	0.00
Number of Questions	3.0792	.30027	1.64464	.874	.427	-.108	.833	0.00
ENG	3.5700	.09500	.13435	1.509	.427	1.695	.833	0.00

In terms of the static condition, to assess data normality the Shapiro-Wilk test and the calculation of the z-value based on the skewness and kurtosis values were used. Number of questions and engagement scale, revealed a *p*-value <0.05, meaning that there was not enough evidence to accept the null hypothesis for both types of tests. Number of answers, smiling and eye gaze did not seem to be normally distributed according to the Shapiro-Wilk test and the z-values (Table 4.6).

Table 4.6. Normality checking for the static conditions

Items	Shapiro-Wilk		
	Statistic	Df	Sig.
Number of Questions	.977	26	**.811**
Number of Answers	.884	26	.007
Smiling	.592	26	.000
Eye gaze	.840	26	.001
ENG	.962	26	**.439**

The calculation of the z-value based on the skewness and kurtosis values showed the same outcomes (Table 4.7).

Table 4.7. Calculation of the z-value for the static conditions

Item	Mean		Std. Deviation	Skewness		Kurtosis		z-value
	Statistic	SE	Statistic	Statistic	SE	Statistic	SE	
Number of Questions	3.2163	.31168	1.58924	-.003	.456	-.720	.887	**0.00**
Number of Answers	1.1863	.19792	1.00919	1.352	.456	1.777	.887	2.96
Smiling	3.2668	.80400	4.40368	3.808	.427	17.615	.833	8.92
Eye gaze	46.1724	3.71815	18.95894	1.579	.456	3.220	.887	3.46
ENG	3.6852	.10155	.55619	.515	.427	-.285	.833	**1.21**

4.12.2 Comparative analysis

In general, this investigation aimed to shed light on the influence that head movement has on video call interaction. Using a paired *t*-test, we analysed the gaze, smiling and a number of questions and answers between the two conditions. We hypothesised that subjects in the simulated condition scored highly in conversational engagement (H7), also communicated by better gaze (H1), smiling (H3) and fewer number of questions (H4a) and answers (H4b). The results revealed significant differences on the part of the measures we used for the two conditions ($p<.05$).

Smiling and number of questions in the simulated setting reported overall higher scores (M = 3.27; SD = 4.40) for smiling supporting (H3) and (M=1.19; SD= 1.01) for number of answers supporting (H4b) than in the static setting (M = 2.33; SD = 1.94) and (M= 1.46; SD= 1.17) respectively. Whereas, eye gaze and number of questions in the static setting reported an overall higher score (M= 46.20; SD= 18.32) for gaze and for number of questions (M= 3.08; SD= 1.64) than in the simulated setting (M= 46.17; SD= 18.96) for gaze and (M = 3.22; SD = 1.59) for number of questions. Also, a t-test was run on IIS scores (H7), the results showed no significant difference between the group means (table 4.8).

Table 4.8. Comparing the means of the factors in both conditions

Factor	Condition	Mean + SD
Smiling	**Simulated**	3.27±4.40
	Static	2.33±1.94
Eye Gaze	Simulated	46.17±18.96
	Static	46.20±18.32
Number of Questions	Simulated	3.22±1.59
	Static	3.08±1.64
Number of Answers	**Simulated**	1.19±1.01
	Static	1.46±1.17
Conversation Engagement	Simulated	3.68±0.56
	Static	3.68+0.56

- ***Effectiveness of the movements***

Despite these results, our qualitative data showed that most of our subjects were satisfied with the movement in the second condition. More than 53% were not confused by the movement, 60% thought the movement moved in the appropriate time, 63% found that the movement improved the communication and 57% agreed that the movement helped in enhancing the communication. In addition, 43% prefer the simulated movement condition than the static condition 38%, whereas19% found no differences between the conditions.

4.13 Discussion

The study of nonverbal communication during conversations, particularly the study of head movement, is an extremely rewarding field. Head movements effectively give real-time listener reactions and form part of the feedback loop that we all rely on to tell us how effective our communication is being (Hadar,Steiner and Rose, 1985b). This study was inspired by a desire for similar findings, which in general were achieved, as a good many of the participants reported that the simulated condition helped them in understanding and communicating much better with their partners, and the movement helped in making the conversation more efficient, which as positive findings, support our hypothesis. One subject said:

"In the second condition, I think the screen movement helped me to interact in a more effective and efficient manner, I was reminded to keep my eyes on the screen and my conversation partner when the screen moves".

This was on the line with another comment from different subject

"I prefer the conversation in the second condition more, as it helped me to better understand what my partner intends to express"

Whereas another added that

"I like when the screen replicating the movements, in which provides emphasis more about the partner's expression"

However, in the static condition, we did not find any improvement for the respondents with the use of video call and video simulated call. These results do not support the results from the literature review; this might be for technical, analysis tools and experiment design reasons summarised under main points as described in figure (4.8).

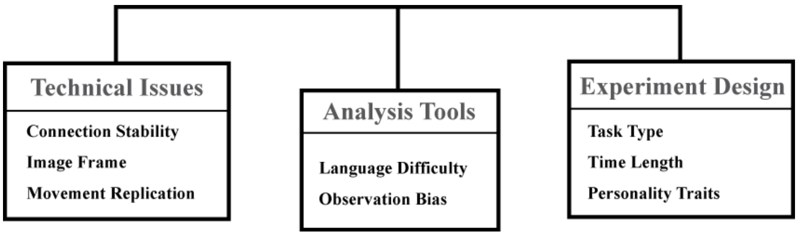


Figure 4.8. Issues with Exploratory experiment

- ***Connection stability***

  Some subjects reported that the visual and the audio signals were not particularly good, as there was sometimes a delay; this might on the surface appear plausible and explain some of the results we obtained. It is possible that the technology we used resulted in a less than ideal image transmission, due to limited bandwidth and technology constraints. As a result, resolution or frame rate may have suffered, or signals may have been delayed. A good way to test this supposition would be the use of technology which made it possible for the video-mediated conversation to look as similar to a face-to-face conversation as possible (high-quality image resolution, no transmission delays), simulating eye-to-eye contact. In fact, Doherty-Sneddon *et al.* (1997) found that video-mediated conversations tend to be more formal, with fewer interruptions and longer utterances, especially when there is asynchrony or delay in visual and audio signals.

- ***Image frame***

  We also believe the similarities between the two conditions in the quantitative data in the present study may be related to the effect of the size of the video image. Therefore, some users reported that the simulated condition caused a distraction during the interaction. An explanation for this is due to the size of the video image, as the movement was meant to be for the face image only, not the shoulders as well.

- ***Movement replication***

  It is important also to highlight that head movement as displayed in conversation, are more dynamic than we could achieve with our system. The head moves constantly while talking, and if the head is still, it tends to be interpreted as a pause or as listening (Pittenger,Hockett and Danehy, 1960, Ray L. Birdwhistell, 1983). Also, speech-related head movements range in amplitude between 170 and zero, with a range of speeds (Hadar *et al.*, 1983). Thus, we needed to produce real-time dynamic communication in order to get a significant result, which was not the case with our experiment. Perhaps participants responded less quickly to the movements because they did not convey the same sense of urgency. More accurate reproductions and simulations of head movements might give an objective improvement in remote communication.

A further investigation in terms of the technical issues will be presented in the next chapter as it is our main focus of the study.

- ***Language difficulty***

  With the interaction engagement scale, the results did not show any statistically significant improvement. Although the present study's sample was representative of college-aged students, and covered an international population, we observed that some of the students were asking for explanations when they answered the questionnaire. We can argue that it is possible that respondents did not accurately answer our questions, as many of them were international students and did not fully understand the questions.

- ***Observation bias***

  To get reliable results for observation, different papers suggested training at least two raters to recognise and rate nonverbal and verbal behaviours, until they reached a satisfactory inter-rater reliability. Also, they suggested to set numbers of rules for the raters when analysing video data to improve the accuracy and limit any mental fatigue. With our data, *t*-tests revealed no significant differences ($p<.05$) on any of the involvement parameters. In our opinion, this might be caused by different issues in our methods or our data.

Our methods required at least two persons to analyse video data to check the reliability, which we were not able to afford and were not within the study capabilities.

- ***Task type***

  Another work by O'Malley *et al.* (1996) explained that some studies cited have involved fairly open-ended discussions and debates. Simultaneous speech and interruptions tend to occur in less formal and more spontaneous circumstances. In our task, the shorter turns and high level of interruptions may have been indicative of the artificiality of the situation causing awkwardness among the participants, and difficulty in smooth turn-taking. We presume from this, that the more conversation cues there are, the more the participants will be involved in the conversations. In other words, people tend to interrupt each other, when there is a problem in regulating conversations, for example, overlapping speech between participants.

- ***Time length***

  The amounts of time the participants took to complete the task ranged from 2 to 7 min, as they had to complete a maximum of 4 guesses. This, in our opinion, was not a sufficiently large difference to have a noticeable impact on their communication patterns.

- ***Personality traits***

  An alternative explanation is that we might argue that interaction engagement is best defined as a personal trait, which is less likely to be influenced by the context in which the communication is taking place, and more to do with the personal traits of the participants. Villaume, Cegala (1988) stated that people who are highly involved understand the relational aspect of conversations rather than the content of the conversation, so a highly engaged person might ignore the channel and place emphasis on the relational signals. Other rationales may exist in the motivation for interpersonal communications and also factors such as interaction engagement.

4.14 Summary

This chapter presents a comparative study on the influence that head movement has on video call interaction. The aim of the study was to explore the potential for replicating the head movements that support engagement and feedback functions between people, which could be useful in the development of a model of human gestures to implement in a telepresence platform. This comprised evaluation of two different video call conditions to evaluate the movement as an interaction method with such a system compared to video-mediated static, which revealed that in this particular setting, the users' smiling and number of questions were improved in the simulated condition than in the static condition. Whereas, the number of questions and gaze were more significant in the static condition than in the simulated one. In respect to the conversation engagement, no differences were found between the two conditions. It can be concluded that despite the limitations, our findings at least highlight important implications for our main study (Chapter 7). These findings suggest that face-to-face interaction is complex in its own right, as it includes various behaviours that help in maintaining the connection between two people. Thus, it will be difficult to find any significant result if we only focus on one of these behaviours. In general, it should be noted that real-time communication requires more than verbal communication, facial expressions, and head nodding. It is important to complement it with other types of nonverbal behaviour such as posture.

Although the measurement and analysis methods used in this study appear to be a practical way to acquire data on the magnitude and length of a variety of gestures, we also believe that available technology can be used to further measure other specific behaviours such as eye gaze, using an eye gaze tracking system. Additionally, synchronising the head movements using a head tracking system to transfer the head movement in real-time to the movement of the screen is possible. Thus, we plan to produce precise on-screen movement by synchronising the on-screen movement with the head movement of the participants, which was one of the main issues we faced in this study. Full details about how we plan to resolve this synchronisation problem will be provided in the next chapter.

Chapter 5 PROTOTYPE DEVELOPMENT

This chapter presents the iterative design and development of the final system, and elaborates on the challenges and solutions for implementation of the system, based on the exploratory study.

5.1 Introduction

Our original aim was to improve social presence within the local site, and therefore our first study focused on improving the site interface design. However, we found that in order to improve both the interaction and social presence, we needed to produce more real-time dynamic communication in order to get a significant result.

This chapter presents the development procedure and considerations for the complete proposed system design. The chapter shows the design process sequentially, based on iterative design and evaluation methods. Section (5.2) outlines the system requirements based on the outcomes of the exploratory study. In Section (5.3), a discussion the Alpha version design presented, beginning with suggested methods for social abilities, then the proposed movements in section (5.4), using Kendon phases (5.5) and the final design specification in Section (5.6). Section (6.7) discusses the system sensing considerations, and physical design processes are presented in Section (5.8). Finally, the Alpha version and Beta versions for the movement with the limitation are described in Sections (5.9) and (5.10). A summary of this chapter presented in section (5.11).

5.2 Technical issues based on outcomes of the Exploratory study

The observations of the users during the *Exploratory study* (Chapter 4) as well as their subjective feedback were considered when designing the final version of the system. Taking the original aim of our study into account, the main points about the technical issues requiring further investigation and development regarding each site are presented briefly below:

5.2.1 Connection stability

Users indicated their dissatisfaction with the quality of the video call, reporting that the visual and audio signals were not particularly good as there was sometimes a delay. The limited bandwidth and technology constraints resulted in the video resolution and frame rate suffering and causing signal delay. A study by Kies, Williges and Rosson (1997) stated that small video resolution (169x120) did not have any effect on task performance; however, it did decrease satisfaction when compared to 320x240 image resolutions, as it might have reduced the accuracy of emotion detection (Knoche,McCarthy and Sasse, 2005).

Although video resolution cannot be controlled in our study, we can control some of the related factors.

- ***The viewing distance:*** viewing distance of the video has a direct effect on perceived quality of the video, the shorter the distance, the better the quality. This is because it becomes more difficult to resolve detail in the display if it is moved further away, even if we increase the perceived quality by increasing the number of pixels in a given area (Knoche,McCarthy and Sasse, 2005).
- ***Sound quality:*** in respect to the effect of the sound quality, different studies have highlighted that subject satisfaction with perceived video quality was improved by better quality of audio (Busso *et al.*, 2004, Beerends and De Caluwe, 1999). Although we cannot control image quality, as it is related mainly to internet bandwidth, we can ensure that we use good quality sound software to minimise the impact.

5.2.2 Image frame

The interaction with the system in the *exploratory study* was effective for task accomplishment, as the subjective feedback from the users indicated that implementing movements was useful. However, it was deemed not to be engaging and lacked presence. Users in the experiment reported that the simulated condition caused confusion during the interaction. We believe this was due to the video frame, as the movement was meant to be for the face image only, not the shoulders as well. Because of the technology we were using, we were unable to shrink the video image to only give a view of the head and neck. Instead, a view of the head and shoulders, down to the elbow was present in both conditions. As a result, subjects were mainly using visual cues from the face (e.g. gaze, expression, lip-movements) but also to a degree, global cues such as posture and gesture (e.g. shrugging of shoulders).

- ***Small Screen:*** Most current telepresence systems come with a small display for the local site, showing only the operator's upper body, or a life-size image of the operator's face (Kristoffersson,Coradeschi and Loutfi, 2013, Ishiguro and Trivedi, 1999, Nakanishi,Murakami and Kato, 2009, Paulos and Canny, 1998). Displaying a small image of the upper-body might be harmful to presence, as found by the same study, and also supported by Knoche,McCarthy and Sasse (2005). They found that acceptability is significantly lower for images smaller than 168x126, regardless of content type, due to the fact that important detail is lost on the smaller screens. However, we decided to adopt the same approach by having a small screen with only a life-size image of the operator's face, which would also be supported by nonverbal behaviour movement to minimise the issue, which is discussed in the next section.

5.2.3 Movements

Assessment of the users during the exploratory study as well as their subjective feedback was considered when designing the initial version of the MobiBot system. The results did not show any evidence of the possibility of identifying any improvement by incorporating head movements. Though it attempted to replicate key movements of head, however, the manual replication for the movement caused a delayed time between the physical in screen movement and the on-screen movement. It is important to highlight that head movement, as displayed in conversation, are more dynamic than we could achieve with our system. The head moves constantly while talking, and if the head is still, it tends to be interpreted as a pause or as listening (Norris *et al.*, 2014, Cegala, 1984). Thus, we needed to produce real-time dynamic communication in order to get a significant result, which was not the case with our first experiment. More accurate reproductions and simulations of head movements might give objective improvements in remote communication.

In addition, we still believe that proximity and posture play an important role in the interaction. A study done by Nguyen and Canny (2009) stated that body language cues go beyond facial cues, in approximating a person-to-person experience.

This is because we attempt to control our facial expressions, whereas proximity and posture convey independent messages.

Taking into account this point, with the next version of the system we aimed to provide increased presence and a more natural method of interaction via the use of more real-time non-verbal behaviour replication (Figure 5.1). In general, the initial design of the revised system considered a combination of the following points:

1. A limit on the driver site view
2. Improved quality of sound
3. A limit on the viewing distance
4. A small screen at the local site
5. Automated replication of movements alongside controlled movements
6. Addition of extra postural movements

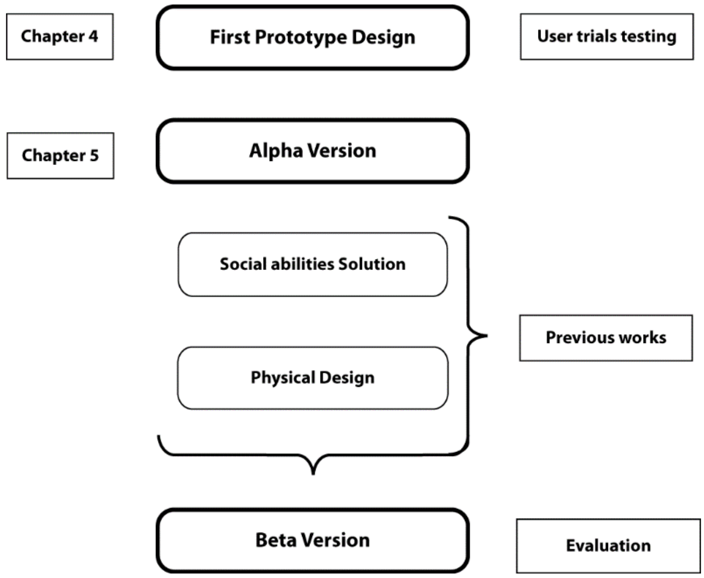


Figure 5.1. System development process

5.3 Alpha version

Our main goal was to display naturalistic behaviour by a combination of skilled behaviour, conveying attention and handling social interaction. In other words, making the communication richer and more effective, which is the basis of any effective system as suggested in the different studies (Rousseau *et al.*, 1997, Bates, 1994). To produce a true to life experience, we needed to follow the rules of human interactions and social conventions. At present, the most accurate method of achieving a life-like experience for HCI is to use multi-modal interactions, i.e. physical motion capture, physiological inputs, the normal five senses, cognitive state and emotional state. Ultimately it is the combination of models that facilitates a system's accurate production of nonverbal behaviour. Busso *et al.* (2004) recognised the need for intelligent computer interaction using a multi-modal approach. Busso *et al.* (2004) focused on analysis of the benefits and weaknesses of emotional recognition systems, based only on facial or vocal interactions. Their results demonstrated that although facial systems performed better than vocal systems, the performance of systems was dramatically improved when the two modalities were combined.

Generating a true whole body markerless motion interaction, even using such an approach is however still problematic. We need points of reference on the body, whether from sensors or cameras. Signal processing and feedback need to be achieved in real-time, and feedback in conjunction with gesture recognition thresholds are still not adequate for all domains. Thus, there is a need for further advances in machine learning techniques, or development of a new technique to deal with real-time markerless interaction for a usable experience.

Taking into account current technological capabilities, we propose to adopt the single model approach by relying on the vocal stream to enhance the natural interactions within HRI, which will be discussed in the next section.

5.4 Establishment of social abilities requirements

The challenges for the Human Robot Interaction researchers are numerous. In our case, the challenge was mainly related to developing methodologies for eliciting user requirements in real contexts, which must adhere to both site users' requirements, and be acceptable in the long-term. Defining a new method to implement human behavioural traits was one of the ideas that sparked off our platform concept and design. Our decisions for this design were informed by one of the early research works by Ray L Birdwhistell (2011), which analysed linguistic data and transcribed extensive stretches of vocal behaviour to reveal that kinesics structure is parallel to the vocalic stream.

5.4.1 Why does movement matter?

Our device had the ability to use an automatically movable screen to produce specific movements which were related to the face, as described earlier. In addition, it had forward, backward and shrugs automatic vocal-triggered movements. These movements were generated to fit in with the technological capabilities that we found feasible to implement with the current equipment. Thus, these movements did not require remote sensors attached to the person's face or body to detect their physical movements, apart from the audio sensor attached to the device. Although these movements might not replicate the exact movements of the person, different studies stated that users were sensitive to the movement of a physical apparatus even if the remote person did not move, which led to improvement in their communication (Nakanishi,Murakami and Kato, 2009). Movement has a great deal of power as a medium for expression, regardless of which type of organism or object is involved. Humans are no different from other animals in that they quickly sense perceived motion, which is a trait of which other humans, animals and makers of artifacts take advantage. Humans are sensitive to motion not just in other humans, but in non-sentient objects too. This leads humans to a tendency to treat certain inanimate as well as animate objects as though they had a sense of being and purpose (Baldwin and Baird, 2001, Dennett, 1978, Malle,Moses and Baldwin, 2001), which is often known as the Theory of Mind (Baron-Cohen, 1991).

Based on this finding we decided to use the vocal stream as our method for creating more natural gesture interactions, which is discussed in detail in the next section.

- ***Speech related movements***

Individual embodied expressions have two radically different kinds of sign activity; verbal symbols or substitute nonverbal behaviours that aid the interaction. Nonverbal behaviour can function to qualify whatever an individual means by a statement, or convey information about the actor's social attributes, about their own conceptions of others present, and about the setting (Goffman, 1999, Ekman and Friesen, 1982, Banse and Scherer, 1996). We found that this link goes beyond just conveying the information, to covering similarities between nonverbal and certain verbal behaviours of individuals, as supported by different studies (Boomer, 1965, Trager and Smith, 2009, Mehrabian, 1972, Trout and Rosenfeld, 1980).

One of the earliest microscopic analyses of the coordination between movement and speech was carried out by William S. Kendon in the early 1960s (Kendon, 1970). He was able to match nonverbal behaviour with a speech transcript in a close study of sound and film records of interactions. A rhythm was proven to exist even at the most microscopic levels (e.g., spoken syllables) where the points of change in the flow of sound are coincident with the points of change in body movement. Nevertheless, Argyle (1973) highlighted that nonverbal behaviour has a hierarchy which corresponds to different verbal unit sizes, e.g. emphasis through a change in loudness or pitch for speech can be emulated through hand or head movements. This can also be found between body movements, vocal hesitations and pauses in speech (Boomer, 1965, Mehrabian, 1977).

Condon and Ogston (1966, 1967) suggested that listener's actions were modelled on a speaker's speech stream and vice-versa, e.g. a phoneme change can be seen in a speaker's talk, which results from the small movements produced by the listener's head, eyes, wrist, mouth and fingers. They further explain that as a rule, speakers and listeners are in synchrony up to the word level, as any variation in the configuration of movement of the listener will match the variation in the speaker's configuration at word, syllable and phonic levels.

> "*If speaker and listener are in synchrony and the listener lifts a cigarette to his lips, draws on it, and lowers the cigarette again, the boundaries of the major components of this action will coincide with boundaries in the behavior flow of the speaker, but these boundaries will not necessarily also be boundaries of the larger waves of behavior in the speaker, for instance the boundaries of his phrases.*" (p.93)

This emphasises the importance of synchronisation between vocal stream and nonverbal behaviour, in regard to regulating and organising dialogue itself in group interaction situations, through sharing attentions and expectancies. In other words, this coordination provides one of the ways in which two people signal that they are open to one another, and not to others.

Based on this point, our method could only rely on the vocal stream to replicate some of the speech related nonverbal behaviour, which is detailed in the next section.

5.5 Kendon phases

As previously stated, one of the early works providing a detailed analysis of human interaction is Kendon (1972). We looked specifically at the detailed analysis to use it as a source for examples of different behaviors between listeners and speakers. This analysis is of particular interest as it provides movement phases during the interaction, with a full description of each phase, which helps to frame the outline of our system.

Three phases of movement and speech rhythm between speakers and listeners were generated as resulted of his analysis which are:

- ***First Phase (opening position)***

At the beginning of the interaction, there is an associated movement called the opening position. This phase serves to visually validate that the speaker is speaking to the right person, and for onlookers, it clarifies to whom the speech is being directed. Shared-movement rhythmicity can be seen here; a mirrored movement which only happens between the speaker and the person he addresses directly, to seize their attention.

- ***Second Phase***

As the speaker becomes more confident that they have the attention of the listener, the movement more or less ceases, apart from mouth movements, eye shifts and blinking of the eyes.

- ***Third Phase***

Finally, as a result of the familiarization between both sides, in this last phase, the listener's behavior is followed by the speaker's, and related to the variation of the pitch level of the speaker's voice.

Apart from these phases, it has been found that some facial expressions or head movements appear at specific junctures in the speech of our partners; for example, head nods and movements of hands and feet tend to occur at the end of rhythmical units of the talker's speech e.g. at pauses within phonemic clauses, but mainly at junctures between these clauses. Vocally stressed words also tend to be accompanied by movements.

5.5.1 Kendon's version for our system

Since our aim was to improve the interactions between both sites in mediated interactions, we focused on identifying the cluster of signals that distinguish a positive evaluation of an interaction partner from a negative one, as identified by Mehrabian (1968). We chose the gestures related to our research, and tried to match them with the phase by identifying which best matched the phase descriptions. Four gestures were chosen from Mehrabian (1968) list based on our design specification, and these were:

1 More forward lean
2 Closer proximity
3 More direct body orientation

In the next section, we will give an outline for movement specifications in respect to Kendon's (1972) interaction phases (Table 5.1).

Table 5.1. Proposed Movement Specification

Study	Movement	Situation	Specification
Person in front of the device	**Pan the mobile holder**	Starting conversation with the person in front of the device	Panning within range of 5°-15°
	Forward	Moving forward to approach the other side at the beginning of the conversation	The system will leave approximately 400mm and 600mm between listener and speaker, with speed range between 500mm/s and 700mm/s
Both sides	**Slight backward lean**	Leaning backward when the other side starts the speech or the person on the device	Backward lean 15°-20°.
	Upward shrug movement	lift the system head up when high pitch occurs on another side, or the person on the device	Within range of 5°-20°
	Leaning forward	If there is a pause between both sides more than 10 seconds, the system will tilt the head down	Within range of 5°-20°
	Multiple head nods	associated with long speeches from both side	individual left and right components of movement within range of 10-50mm

- ***First Phase (opening position)***

As it is called an opening position, we thought a pan by the mobile holder would give the user the same feeling as if the remote person had moved closer to the user when they began talking. In addition, a forward movement would help in maintaining the exchange of talk between two people, emphasising the beginning of the speech, and of course ensuring there were no physical obstructions to block them from addressing each other in an encounter. This movement would be a translation of the starting conversation from the person in front of the device.

While investigating the optimal range of viewing distance for desktop monitors, a study by Ankrum (1996) recommended a minimum distance of 635 mm. However, this was not the case in every situation; another study relating to viewing distances for LCD monitors found that screen reflections affected the viewing distance, which resulted in a shorter distance compared to a normal desktop monitor (Shieh, 2000).

Although this was also supported by Shieh and Lee (2007), they argue that this effect is minimal. Therefore, the system tried to keep approximately 400mm and 600mm between one side and the other when moving forwards. This forward movement range was between 500mm/s and 700mm/s.

- ***Second Phase***

The second phase is described as quiescent by Kendon (1972); the analysis did not show big movements between both sides, only slight movements. Therefore, we decided to add visibly smaller movements into this phase. A slight backward lean movement accompanied the initiation of the talk (from the person in front of the device) and a slight forward movement when the person on the device was talking, as suggested by Mark Knapp,Hall and Horgan (2013). This slight lean adjusted the viewing angle to the optimum viewing angle which is between 15° and 20° beneath the horizontal sight line.

- ***Third Phase***

As we could not indicate when the conversation would reach an end, we decided to add some movement to fit the description of the phase. This phase is described as an interchange phase, where speaker and listener mirror one another’s posture, and such posture shifts often occur synchronously. However, in our case, we replicated this with a slight upward shrug movement in relation the rise of the pitch of voice, when the primary stress points in the speech are accrued as reported by Mark L Knapp and Daly (2011). In addition, forward movement of the head happened during silent parts of the conversation where movements seemed to peak. This was within the range of 5°-20°.

If we examine these three phases, we find that there were two scenarios in respect of the vocal stream translation. The first two movements aimed to translate the vocal stream into movement by one side (Figure 5.2), whereas the rest concerned the volume only, without any differentiation between the sides (Figure 5.3).

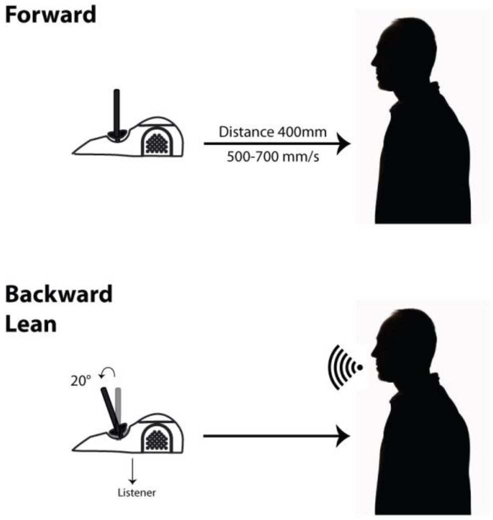


Figure 5.2. Scenario 1

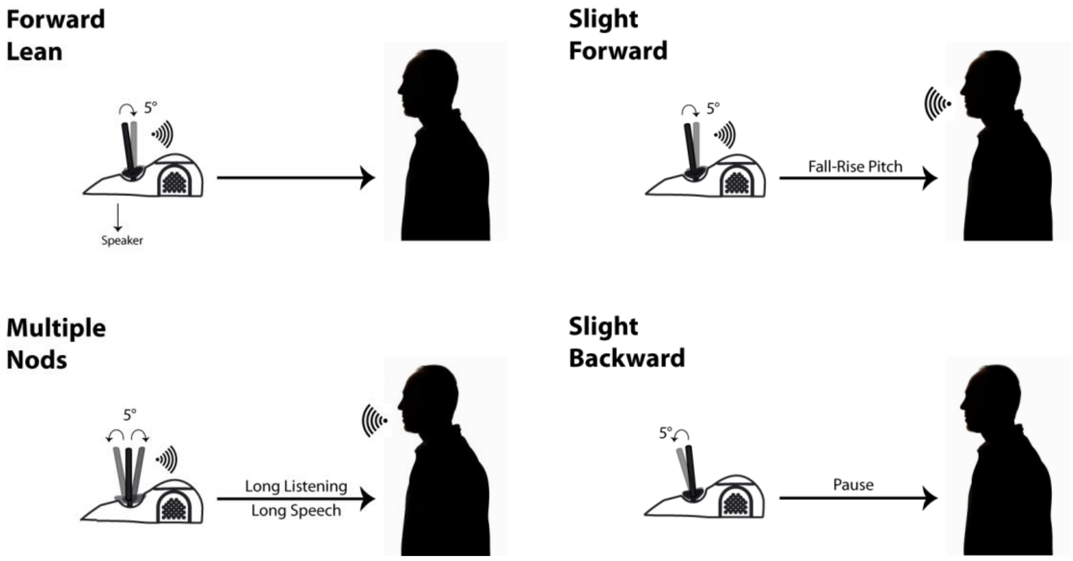


Figure 5.3. Scenario 2

5.6 Final design specification

Through the iterative design process, the final design specification is described below. Our Alpha version combined both vocal-triggered movements and user-controllable mimicked movements, giving the driver lateral screen movement, and making the whole device capable of movement from place to place.

Performance

- ***Movement***

A combination of both automatic vocal-triggered and user-controllable movements was used to give the driver, in addition to the ability to convey nonverbal behaviour, the ability to move the screen from side to side, and move the device from place to place remotely as described here:

Table 5.2. Proposed Movement Specification

User inputs	Automatic Vocal-triggered
Tilting forward/backward	Pan for the mobile holder with 15°-20°
Pan or rotation	Forward movement distance between 400mm and 600mm
Linear left/right	Forward lean Within range of 5°-20°
Linear forward/backward	Backward lean Screen backward 5°-10°.
	Shoulder shrug Within range of 5°-20°
	Nods: 15°-20°

- ***Speed***

1. A maximum forward walking speed 900 mm/s
2. A minimum walking speed 400 mm/s.

The maximum backward speed should be at least half the maximum forward speed.

- ***Payload***

The platform can carry a load up to 250g based on maximum smartphone weight.

- ***Manoeuvrability***

The platform has 2 degrees of freedom when operating and is able to move and maintain control at the same time.

- ***Navigation***

The platform must be able to determine its position relative to the person in front of it.

- ***Power source***

The platform uses a portable closed rechargeable power source (battery).

Constraints

- ***Size***

The vehicle does not exceed 300mm in length.

- ***Weight***

The vehicle does not exceed 1 kg.

- ***Shape***

The shape should be designed in a way that can accommodate any smartphone.

Control

- ***Communication***

The communication of tasks and programming takes place on the other pilot or remote site through wireless internet connection.

- ***User interface***

The user interface is in the shape of a program installed on a personal computer.

5.7 Design considerations

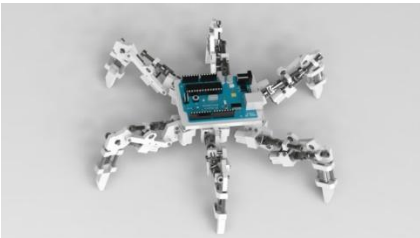
In order to build a telepresence robot that at least superficially approaches human-likeness, we needed to consider different aspects of our proposed design for telepresence robotic products that support face-to-face communication. Taking into account these specifications and the limited budget for the project, the following section explains the different aspects that contributed to the final design.

5.7.1 Drive concept

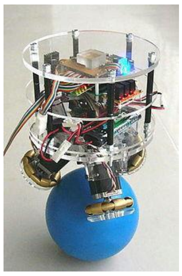
There are many different ways for mobile robots to move across a solid surface, with methods ranging from wheels and tracks, legs and single ball (Figure 5.4). However, wheels are most often used as they are better suited for most functions than other methods. In addition, most current commercially available mobile platforms employ wheels as they better facilitate rolling.

Wheels

Tracks



Legs



Single Ball

Figure 5.4. Mobile Robots Drive Potential Solutions

- ***Mecanum Drive***

The drive system was powered by Mecanum wheels because of their special motion advantage, as opposed to traditional wheels. They provide rapid motion, high mobility and 3 degrees of freedom, as they can move on any heading and turn at the same time (Salih *et al.*, 2006). All types of Mecanum wheels are based on the same general principles. A Mecanum wheel (Figure 5.5) is a combination of a central wheel which is supplemented by rollers, which are located around the wheel periphery at various angles.

The angle of the rollers is arranged so that they carry some of the load from the central wheel, and their rotational direction ensures that the entire vehicle can be moved in any direction that the driver wishes, without the central wheel having to change direction. This makes the device free-rolling, smooth to operate and highly manoeuvrable (Diegel *et al.*, 2002).

Figure 5.5. Mecanum wheels

- ***Two, Three or Four?***

Two, three and four wheels are the most commonly used, each combination having different advantages and disadvantages. A two-wheel drive has very simple controls, but reduced manoeuvrability. A three-wheel drive has simple controls and steering, but limited traction. A four-wheel drive has more complexity in the mechanics and control, but higher traction. Traditional mechanism configurations for Omni-directional robots are based on three and four wheels, but having two wheels would give the advantage of smaller size and require less equipment (Sharbafi,Lucas and Daneshvar, 2010). With more than two motors and wheels, it is expected that the robot would have more effective floor traction (de Oliveira *et al.*, 2008) i.e. less wheel slippage. However, it will have a more complex mechanism, more complex controls and traditional current consumption, and may require some kind of suspension to distribute forces equally among the wheels (Oliveira *et al.*, 2009).

Based on this information, we decided that the Mecanum drive concept, with a pair of powered wheels on a platform, combined with a free-running third wheel, was the solution to follow. This seemed to be the best fit for our design requirements, due to the simplicity of the controls, the reduced cost and the limited space we had available due to the proposed size and shape in relation to the smartphone. The following section will explain the physical design in detail.

5.7.2 Physical configuration

When designing a smartphone-based robot, an inevitable design decision is the integration of the mobile device within the overall morphology of the robot, as explained in the following points.

5.7.3 Anthropomorphic form

Humans are susceptible to the perception of certain patterns and forms as humanlike, especially patterns and forms that look like body proportions and faces (Lidwell,Holden and Butler, 2010). Two concepts were considered, which explain the reason behind using the anthropomorphic form in designing our MobiBot system. These were:

- ***The familiarity by keeping things the same***

The familiarity thesis states that our minds are attuned to use anthropomorphic forms to help us to make sense of things we do not understand, in terms that we do understand, and what we understand best is ourselves. The familiarity thesis is primarily a cognitive motivation for anthropomorphism. Therefore, the chosen image needs to be capable of being recognised as a human who is welcoming and informative, and who can add to the positive affective tone in interactions and thereby form a relationship with the viewer (Lidwell,Holden and Butler, 2010).

- ***Command and control by explaining the unknown***

Anthropomorphism is often used to explain products with new functions or technologies. In other words, when using the anthropomorphic form, the correct part must be visible and it must convey the right message without damaging the aesthetics. With reference to humans, this means a suitable and appropriate level of personification without simply selecting an imaginary or symbolic human-like form.

5.7.4 Simplicity

Simplicity can be achieved with a clear and clean design as mentioned by Karvonen (2000). He reported that users were appreciative of the simplicity in his design, and since the design was pleasing to them; they were also ready to trust it more easily.

The Universal design book by Lidwell,Holden and Butler (2010) further outlines one of the basic guidelines for improving simplicity, which is to remove unnecessary complexity.

5.7.5 Contour Bias

The region of the human brain called the amygdala tends to be activated when presented with sharp-angled objects or objects that possess pointed features. The amygdala is involved in fear processing, and this mechanism in our subconscious probably evolved to help us to detect perceived threats. This fear response related to angular features also influences our affective and aesthetic perception of objects, and this response is similar in both males and females – we all have a deep-rooted contour bias (Lidwell,Holden and Butler, 2010).

In a recent study by McColl and Nejat (2014), they found that users were much appreciated for the exaggeration of body movements as it produced higher emotional intensity ratings compare to a normal one. In light of the previous points, and as we planned to use a smartphone with the video of the face only, we decided to attach it to a shoulder-shaped platform (Figure 5.6). As it was practical and still allowed us to include many of the key communication characteristics. This would give the end user a sense of human appearance in the video call, and therefore potentially increases the acceptance of the video phone applications user. The portrait video will make the focus only on the face not getting distracted by the side view as in the landscape setting.

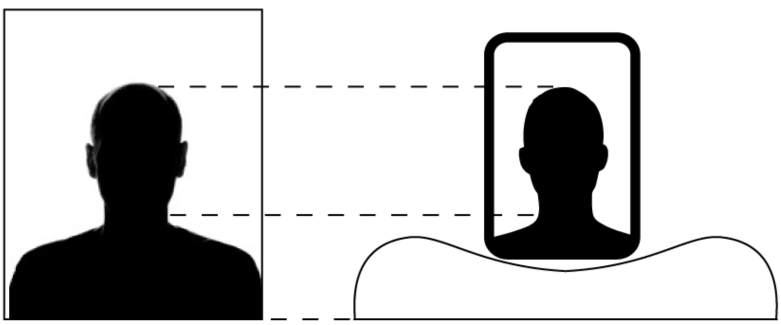


Figure 5.6. Shoulder-shaped proposed platform design

5.8 Design process

The design process for MobiBot included four stages: (a) freehand appearance sketches; (b) DoF placement animation exploration; (c) abstract skeleton prototype; and (d) final solid design and construction.

5.8.1 Freehand appearance sketches

At an early stage of the design process, we developed the appearance and personality of characters of the robot through a freehand sketch for the preliminary design. The aim of the sketch was to explore the appearance possibilities unconstrained by mechanical considerations. Given the decision that the phone should be part of the robot's body, a freehand appearance sketch explored this concept by designing a shoulder shape-like form, considering its relationship to the mobile device. The far end of MobiBot is shaped like a seat, creating a place for the smartphone to be attached. The scale of head as represented in the video call was considered when designing MobiBot, with the aim of creating a familiar relationship with the human body (Figure 5.7).

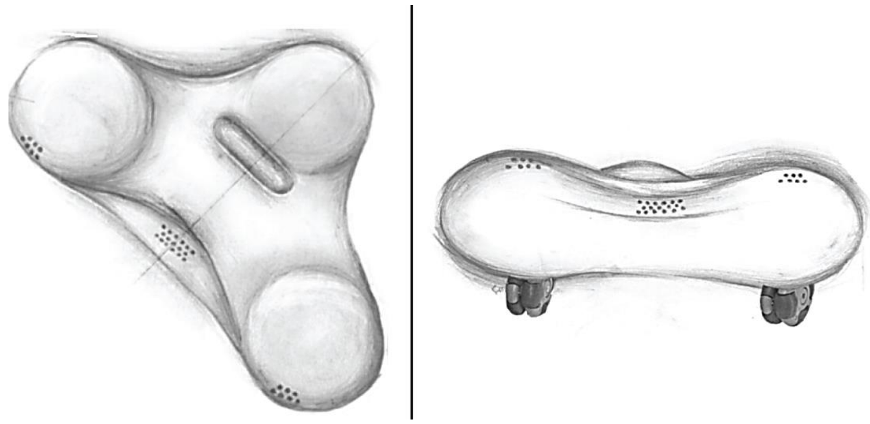


Figure 5.7. Platform Sketches

5.8.2 Detailed 3D and skeleton design

After the general form was determined, the design process shifted to the next, more involved, step of 3D representation. A rough model of the robot was built in a 3D design program, Creo (3D CAD software), replacing detailed elements with geometric approximations of these parts.

Some of the major appearance parameters considered at this stage were the bulk of the base, the width-height proportions of the robot and the relationship between smartphone and body (Figure 5.8). In addition, a skeleton design was developed in order to explore the robot’s expressive movement in physical space in relation to the smartphone.

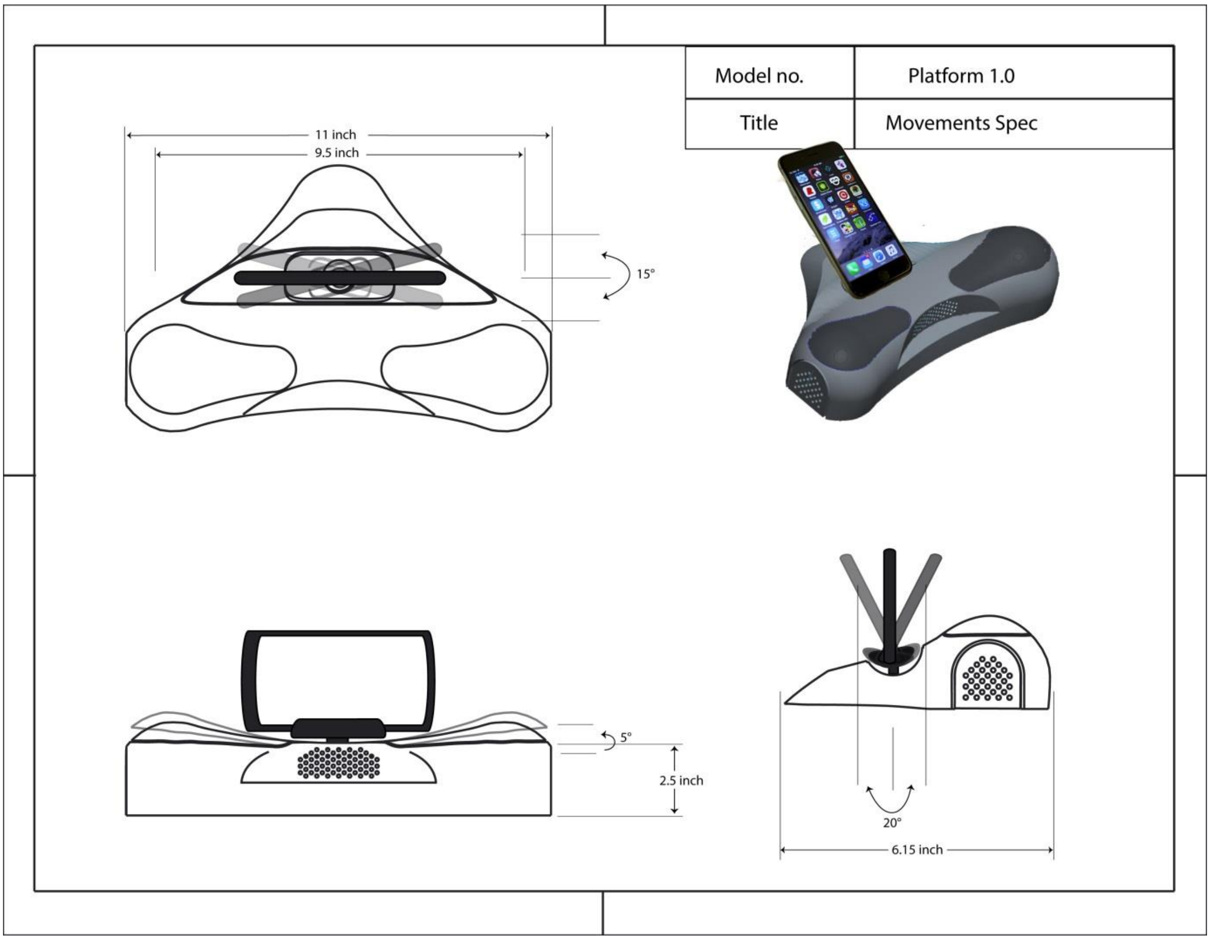


Figure 5.8. Detailed and Skeleton Design

Having set the parameters for MobiBot, smartphone position and relationship of the DoFs in the previous step, we used a 3D printing machine to print the solid body, which we used later with a vacuum-forming machine to build the shell body out of rigid material. For the shoulders, we decided to use a softer material and therefore we used foam sheeting which we reformed, also using the vacuum-forming machine. We felt that the soft material would help us to deliver the shoulder movement for our system (Figure 5.9).

Figure 5.9. MobiBot Shell

Having achieved a 3D model, we had the opportunity to reveal design requirements and problems around placing of the hardware components. For the alignment of the hardware components, we designed an acrylic base using a laser-cutter to define the placement of each part of the MobiBot hardware. The acrylic base model can be seen in Figure 5.9.

5.8.3 Final Solid Design and Construction

The final solid design stage combined the insights in terms of DoF number, placement, and orientation, tested on the skeleton design, and the shape explored in 3D design stage, plus it resolved the issues of the physical constraints and dynamic properties of the motors used (Figure 5.10).

For the shell colour, we decided to use a matte material so it wouldn't reflect any light which might distract the users, and for the same reason, we decided to go with a dark colour which in our case was black. For the smartphone holder, we bought an off-the-shelf one, and modified it to meet our requirements.

Figure 5.10. MobiBot

5.8.4 Movement implementation requirement

At the beginning of this section, I should highlight that a lot of the following decisions were made with the advice and direction of the technical staff involved and the limitation of the electronic hardware used.

To deliver the proposed movements, it was decided to use an Arduino Uno and program with a microphone (Figure 5.11). In order for the Arduino program to detect the user's audio voice levels reliably, the microphone was attached to a high gain amplifier to boost the signal being fed to one of the Arduino Uno analogue to digital converter (ADC) inputs. After extensive iterative testing of various amplifiers and microphones, including a throat mic, the most stable combination was selected. Trying to detect the maximum audio peak with the Arduino Uno was limited by the sampling time of the Arduino Uno processor.

This limited processing power also eliminated using the FFT (Fast Fourier Transform) library, as the time to further process and store the values caused some of the peak audio level values to be missed randomly i.e. words and audio level were missed. Alternative processors were disregarded due to space and redevelopment time.

The compromised solution involved an amplifier and peak hold detector circuit, which was used to condition the audio microphone signal to enable the peak values to be measured more consistently and accurately by the Arduino ADC input, when the user was talking at a normal speed.

Besides the sensors, MobiBot was fitted with geared dc motors which are also controlled using the Arduino Uno processor. The current and voltage from the processor are be amplified using a motor drive control board to control the motors on the MobiBot device. The two small DC motors have gearboxes to increase the torque force that controls the wheels and to orientate the device body, and the four servo motors control various upper movements. There is a servo to control the left to right or pan movement of the mobile holder part, and a servo motor to control the up and down or tilt movement of the mobile holder. The last two servo motors control the simulated left and right shoulder movements up and down.

Other problems experienced during development that had to be overcome were jittery movements from the servo motors, and their associated electrical noise from servo motors and dc motors feeding into the Audio amplifier conditioning circuit.

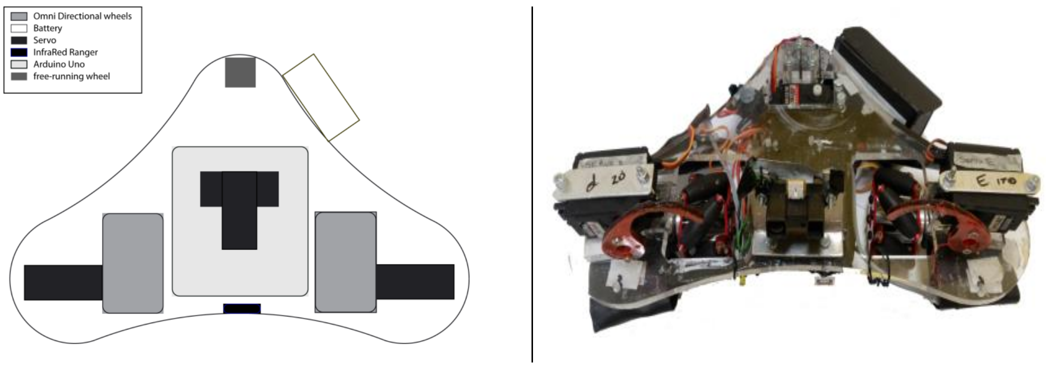


Figure 5.11. MobiBot Hardware placement

5.9 Movement Alpha Version

The measured audio values are used to automatically select a pre-programmed reaction sequence within the Arduino Uno program to control the tilt servo motor or shoulder simulation servo motors. MobiBot can react at any given moment or over a period to the user's audio voice levels. As the main point was to detect the audio level of the user to produce the movement, the user is required to wear the audio headset (electret type) with the microphone fitted close to the mouth. The wearing of the microphone allows for more consistent audio levels to be detected, the audio signals being amplified and transmitted to the Arduino Uno processor in this case. The Arduino is programmed to move the device motors according to the following pattern or sequence which can be seen as the Alpha version of the proposed movements:

1. Firstly, the device positions itself to face the user's midpoint, using one of the small DC motor gearboxes, the left-facing in this case. This is achieved using voltage values detected from the audio sensor feed to the analogue input of the Arduino Uno processor, which turns the motor until the optimum value is detected.
2. Once the device is facing the user, the closely fitted Microphone's audio voice levels are detected and analysed, and if the user's audio levels increase to a predetermined level, the device moves the mobile holder up 10 deg (tilt).
3. After a period (30 sec) of continuous talking at audio levels above the predetermined level by the user, the device moves the mobile holder up and down (tilt), using the servo motor, at a rate of about every 5 detectable audio level changes e.g. words.
4. If no audio voice level is detected from the microphone for a 30-sec period, the device moves the mobile holder back down (tilt) to home via that servo motor.
5. If the person on the mobile phone to whom the device user is talking also starts talking (audio levels go above the predetermined level), the mobile holder on the device moves up an additional 5 deg (tilt). If the user is also

talking it would already be at 10 deg up (tilt) and so will move a further 5 deg to a total of 15 deg tilt up.

6. If the audio level detected from the user goes above an ‘average’ level (a level previously determined by audio levels derived from iterative testing measurements), for that user, the simulated shoulders move up and when the level drops they return to the home (down) position.

This pattern sequence will continue until the program is reset for the next user/ person.

5.10 Audio sensor limitation

While these movements could be promising features for the system specification, some difficulties were encountered some difficulties in achieving them. The initial plan was to implement these human behavioural traits using the vocal stream via audio sensors.

Using this type of sensor would only help to achieve my goal if there were some reliable everyday voice sounds available. People’s voices change naturally over the course of a day, so that even between morning and afternoon there can be enough variation to affect accuracy. Stress, fatigue, and other factors can also cause vocal modifications and consequently reduced accuracy. Moreover, the speech of men and women differ most in terms of vocal pitch. Adult women have high pitch levels and greater pitch variability than do adult men because of the male’s thicker vocal folds (Mark L Knapp and Daly, 2011, Villaume and Cegala, 1988, Beattie and Barnard, 1979). Other gender differences in speech appear due to socialisation processes as opposed to biological factors (Mark L Knapp and Daly, 2011). While computer programs are commonly designed to produce a precise and well-defined response upon receiving the proper (and equally precise) input, the human voice and spoken words are anything but precise. Each human voice is different, and identical words can have different meanings if spoken with different inflections or in different contexts.

Based on this, it was decided to refine movement input method into something that could be implemented without the technology limitation. Further details are provided in the next section.

5.10.1 Beta version

- ***Automatic Vocal-triggered movement***

One of the initial program requirements was to use the user voice audio levels/strength detected on a microphone and amplifier, mounted on MobiBot, to determine the user's direction and then have the device move to face the user, via a directional microphone. This proved difficult for various reasons, mainly the fact that the voice level varies naturally when someone is speaking. This proved unreliable due to background noises and difficulties in detection of audio source direction, coupled with inconsistent audio levels from the user's voice. In addition, the difficulty in using voice as an input lies in the fundamental differences between human voices e.g. pauses and the voice levels detected by the microphone and high gain amplifier circuit. MobiBot was rotated through incremental angles around a fixed audio source, namely the user, using one of the DC motors on one of the wheels, turning for a set time, as no encoders were fitted for simplicity. As it proved difficult to derive meaningful values, it was decided not to use the audio levels detected by the microphone and amplifier to do the person position detection. I selected a more basic approach by fitting the device with an IR (Infra-Red) sensor to detect the proximity to an obstacle or person which worked fairly consistently.

The IR sensor values were measured, stored and compared with values for an object between 0.4 and 0.8 meters. Thus, on the start-up, the device turns using one of its small DC motor gearboxes. While turning, the IR proximity sensor detects the presence of an object (assumed person) at a distance of between 0.4 and 0.8 metres. When it detects an object within this range, the device rotates to roughly the centre of the detected object, which should basically face the user. This assumes that it actually detects a person at this point, as there would be no other obstacles at that distance.

After the user (object) was detected within range and MobiBot had turned to face the user, the next requirement was to detect and react to the user talking for periods of longer than 30-40 words, and also if the 'speaker' on the mobile phone in MobiBot’s cradle was talking before the device started nodding. After much deliberation in deriving the above program to the outlined specification, it was then decided to be changed to meet a lesser specification for a more refined approach. This involved decreasing the number of user words detected to 20 from 40 before the device started nodding, although with less angle variation and at a slower speed. A further additional requirement was that if, after more than 20 words and a high level of audio volume was detected, the device would nod with a faster nod rate than the previous slower speed nod rate, but with same adjusted nod angle used for the lower user volume audio movement.

Further to this, MobiBot also needed to sense and react if the user’s voice was raised above certain predetermined audio levels. Initially, this was done by getting each specific user to speak at normal levels at the power on, or reset, and starting to drive audio calibration levels for a normal voice, for approx. 25 words. The user was notified when this was complete by the red indicator LED on the front being switched off. From these values, the program derived the values for when to react if the user’s voice went above the specific user program calculated normal voice level, plus a percentage of tolerance. The tolerance and values were derived from iterative testing. This was later approximated to a fixed maximum value due to non-consistency of audio levels between users. A green indicator LED was fitted at the front to indicate when each user’s word was detected, and the same previously mention red LED provided a further function now to provide an indicator when the user’s voice went above the peak value set by iteration and trials.

A further initial requirement was to detect changes in the tone in the user's voice when the person’s stress level changed. Using the Arduino FFT (Fast Fourier Transform) library code (see ref http://wiki.openmusiclabs.com/wiki/ArduinoFFT) to determine the frequency at which the max audio level occurred in someone's voice, this was displayed on the attached computer using the available processing code from the above website. However, since the tone varied with different words, this was not found to be a simple way to detect levels of stress. A library of comparison values would probably be needed for each person, and more research would be needed into that field, therefore, this requirement was omitted from the program.

The program also needed to detect when there was a lack of audio words for a period of 30sec of no speech, and then react by returning the tilt servo back to home position.

The shrug movement and the back/reverse movement audio mic volume trigger levels were eventually increased greatly, and set to fixed values, thereby lessening their chance of occurrence. This removed the need for individual user audio calibration levels to be used to adjust trigger reaction levels for the shrug movement and word detection levels between different users, simplifying the program significantly, and had this been done earlier, it could have reduced the need for much of the previous deliberation over the program.

- ***Mimicking, controlled movement***

If no words are spoken within 10sec of program start-up or switch on, the program automatically selects a user interface program, which continually monitors the user keyboard input and relays the characters via a serial input port/USB port of the Arduino. If a specific string character or keystroke is typed and received, it is compared and used to control MobiBot to react in the following way (Figure 5.12):

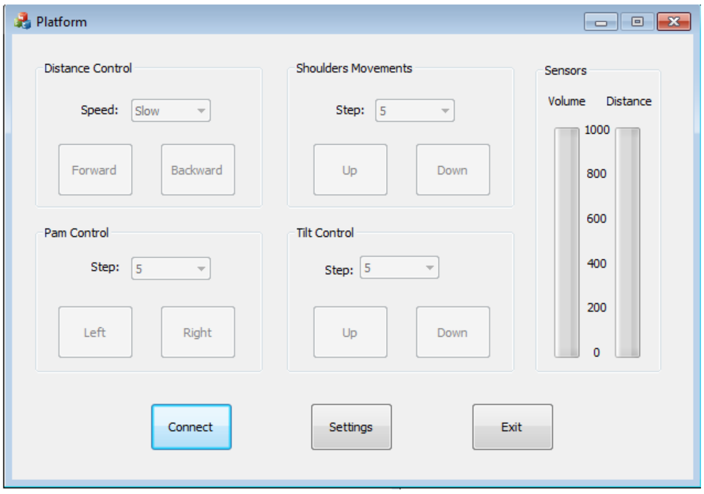


Figure 5.12. MobiBot Controlled movement interface

1. Capital 'A' moves MobiBot forwards with a pre-set speed. Pressing the 'a', 'b' or 'c' key prior to pressing 'A' selects one of three fixed speeds, at which the forward DC motors move. The default speed is the lowest speed 'c'.
2. Capital 'B' moves MobiBot backward. Pressing 'a', 'b' or 'c' also alters the backward speed.
3. Capital 'C' stops MobiBot moving forwards/backward.
4. Capital 'D' moves the simulated shoulders up by a specific fixed amount, then automatically down after a short period. The amount is not changeable by user control.
5. Capital 'G' moves pan servo F motor left by a set value. The value is set by pressing 'g', 'h' or 'i' prior to pressing 'G', with default value 'i'.
6. Capital 'H' moves pan servo F motor right by a set value. The value is set by pressing 'g', 'h' or 'i' prior to pressing 'H', default value 'i'.
7. Capital 'J' tilts servo G motor upwards to a pre-set set point and back down, not changeable by the user.
8. Capital 'E' moves the simulated shoulders down by a specific amount. This amount is set by pressing 'd', 'e' or 'f' prior to pressing 'E', which increments the repetitive step amount value by which the shoulder will move down, on each press of 'E'. The default is 'd' (moves servo motors D and E).
9. Capital 'K' brings tilt back down in incremental amounts to a set min value; this value can alter by pressing 'j', 'k' or 'l' prior to pressing 'K'.
10. Based on the requirements above the differential steering/driving system two motors which would also control using the Arduino Uno processor program. This will be explained in detail in the next section.

5.11 Summary

This chapter has presented the vocal stream movement interaction based system design. The Alpha design of the system was based on the pilot study findings. The technology and budget limitations required an iterative evaluation procedure, mainly based on technician’s feedback and observation. The Beta version was the fruit of the results obtained from the conducted tests. The following chapter discusses the evaluation method of the system and the development of the hypotheses.

Chapter 6 Social Acceptance of a Socially Interactive Telepresence Robotic System

This Chapter elaborates on the conceptual framework development, hypotheses development and system measurement design.

6.1 Introduction

In order for such systems to deliver their intended benefits and for users to socialise through them, users need to accept the systems. Without the user's social acceptance, technology cannot hope to deliver whatever value it may offer. Therefore, user social experience and the social behaviour of these systems need to be carefully investigated, along with appropriate measurements and approaches, to lay the basis for future ways of working.

For the evaluation of the prototype TP application, the second study employs a customised Unified Theory of Acceptance and Usage of Technology Mode (UTAUT). The UTAUT aims to understand and explain users' attitudes and the factors that influence the attitudes towards the MobiBot TP system for video call. This chapter discusses our conceptual framework development and hypothesis development; furthermore, the chapter elaborates on the development of evaluation measurements based on the framework.

Section (6.2) discusses the Unified Theory of Acceptance and Usage of Technology model. Section (6.3) explains the development of the conceptual model for a Social Robot system based on the UTAUT (the Unified Theory of Acceptance and Usage of Technology model), including the TP system evaluation, user social interaction evaluation section (6.4), user cognitive experience in section (6.5) and the role of embodiment in the system under consideration in section (6.6). The research on measurement development is presented in Section (6.7), followed by summarised highlights in Section (6.8).

6.2 Unified Theory of Acceptance and Usage of Technology Model

The original Technology Acceptance Model (TAM) developed by Davis (1989) (Figure 6.1), was based on the work of Ajzen and Fishbein on the Theory of Reasoned Action (TRA) (Ajzen and Fishbein, 1988).

The TAM, at the most basic level, states that the perceived ease of using the technology and the perceived usefulness of the technology are the main factors that influence the user's intent to use a system, which in turn is the main predictor of actual use of the system. Venkatesh *et al.* (2003) offered an extended version of the TAM (Figure 6.2) where a number of external variables were added, which were assumed to influence the behavioural intention to use a system. The Unified Theory of Acceptance and Usage of Technology was the modified version of it. For the evaluation of the prototype MobiBot application, this study employs a customised Unified Theory of Acceptance and Usage of Technology (UTAUT) model proposed by Venkatesh *et al.* (2003), evaluated by (De Ruyter and Aarts (2004) and further improved by Heerink *et al.* (2009).

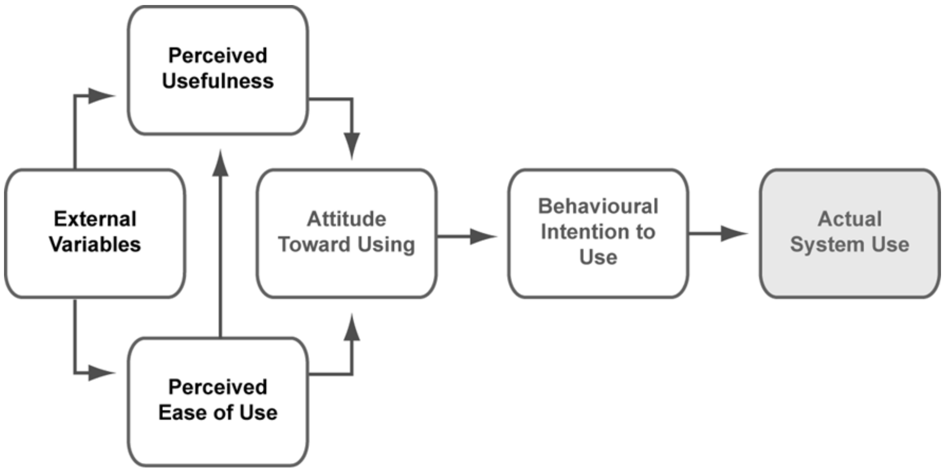


Figure 6.1. Technology Acceptance Model (TAM)

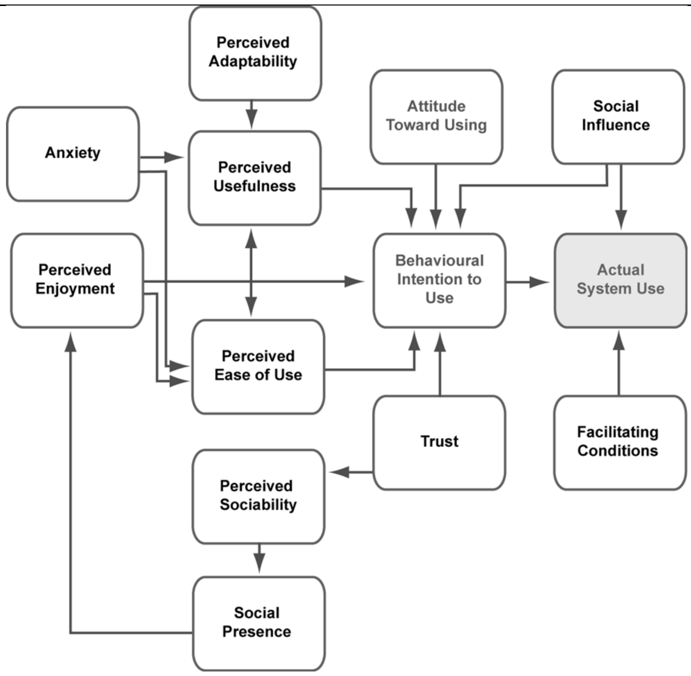


Figure 6.2. The Unified Theory of Acceptance and Usage of Technology

The UTAUT was the first acceptance model to add social acceptance as a relevant aspect of human-robot interaction. The UTAUT aims to understand and explain customers/users attitudes, and the factors that influence their attitudes towards the socially interactive robotic system. It covers the following constructs: the perceived usefulness of the technology which encompasses a broader range of ideas, and was renamed Performance Expectancy. This term outlines the expectations that the user has about the performance of the system. Perceived ease of use was also more broadly defined and was renamed Effort Expectancy. This term describes the expectations the user has of the effort that is needed to use the system. Social Influence (SI) which is mainly about the user's perception of how people think about him using the system and Facilitating Conditions (FC) which is an objective factor in the environment that facilitates using the system. The UTAUT can be constructed with additional factors, and extended in order to ensure appropriateness and comprehensiveness to adapt to specific characteristics of the technologies. The building of new constructs is based on the system's characteristic and the nature of the tasks.

6.3 Development of the conceptual model of the acceptance of MobiBot TP system as a conversation partner

The Unified Theory of Acceptance and Use of Technology (UTAUT) has been widely used by scholars to investigate the attitudes and acceptance of new technologies in different contexts, including social agent technology. Most of these studies focused on elderly people's acceptance of socially assistive robots; only a few experiments were conducted with adults. They investigated the acceptance of robots or their virtual agents as a conversational partner (Table 6.1) (De Ruyter and Aarts, 2004, De Ruyter *et al.*, 2005, Heerink *et al.*, 2006, Ates, 2008, Heerink *et al.*, 2008b, Zaad and Allouch, 2008, Heerink *et al.*, 2009, Heerink *et al.*, 2010, Looije,Cnossen and Neerincx, 2006, Heerink, 2011, Timothy W Bickmore,Caruso and Clough-Gorr, 2005).

Table 6.1. Previous studies that have adopted UTAUT for HRI evaluation

References	UTAUT Construct	Study Interest	Subject Age
de Ruyter and Aarts (2004) and Saini et al. (2005)	Modified version of the Unified Theory of Acceptance and Usage of Technology (UTAUT)	Exploring the concept of social intelligence in the context of designing home dialogue systems for an Ambient Intelligence home	Elderly
Heerink et al. (2006)	Performance Expectancy, Effort Expectancy, Social Influence, Attitude, Self-Efficacy, Anxiety, Intention to Use	Investigating the influence of perceived social abilities of a robot on user's attitude towards and acceptance of the robot	Elderly
Looije et al. (2006) and Bickmore and Schulman (2007)	Modified version of the Unified Theory of Acceptance and Usage of Technology (UTAUT)	Investigating the possibility of incorporating a personal robot as a personal assistant for health assistance	22–29 students
Weiss et al. (2008)	Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions.	Field study of a conversation agent with passing pedestrians, to investigate a proposed methodological concept	18–75
Heerink et al. (2008) and Zaad and Allouch (2008)	Modified version of the Unified Theory of Acceptance and Usage of Technology (UTAUT)	Investigating the acceptance of robotic technology by elders and in particular the influence of a robot's social abilities on acceptance	65–89
Heerink et al. (2009)	Social Presence, Perceived Enjoyment, Intention to Use	Investigating the Influence of Social Presence on Acceptance of an Assistive Social Robot and Screen Agent	65–94
Heerink et al. (2010)	Intention to Use, Social Presence	Exploring the relationship between social presence, conversational expressiveness, and robot acceptance,	Elderly
Heerink (2011)	Anxiety, Attitude, Facilitating Conditions, Intention to Use, Perceived Adaptiveness, Perceived Enjoyment, Perceived Ease of Use, Perceived Sociability, Perceived Usefulness, Social Influence, Social Presence, Trust, Use.	Development and testing of an adaptation and theoretical extension of the Unified Theory of Acceptance and Usage of Technology (UTAUT) related to functional evaluation and social interaction	65–92

The UTAUT model seems a sound basis to start exploring factors that determine users' acceptance of social robots. This is due to its extensive validation and the potential applicability of the model to human-robot interaction as indicated by (De Ruyter *et al.*, 2005), it also covers the social factors as explained earlier. The current study used a modified version of the UTAUT to include further constructs that were deemed essential in determining acceptance of a social robot agent as a conversational partner.

6.4 MobiBot user social experience evaluation

The usual research methodology on socially embodied agents includes a comparison of characters that lack expressive behaviours with characters that express those behaviours. Characteristics which include natural-seeming interactions are usually seen as expressions of the advantages of characters with expressive behaviours over the more neutral characters (Timothy Bickmore and Cassell, 2001), satisfaction with interaction (Timothy Bickmore and Cassell, 2001, Heylen, 2005), closeness to human characteristics (Thórisson, 1997), appropriateness of movements (Thórisson, 1997, Sidner, 2004, Heylen, 2005), joint attention (Imai,Ono and Ishiguro, 2003) and attention getting and cooperation elicitation from strangers (Bruce,Nourbakhsh and Simmons, 2002) and emotional responses (Sylvia Tzvetanova Yung,Tang and Justice, 2009).

In common with the above works, we have observed that test-users reacted positively towards the more expressive and human-like characteristics which were introduced by UTAUT. Perceived Enjoyment, Perceived Sociability and Social Presence are essential factors for both the functional and social acceptance of embodied agent technology (Figure 6.3).

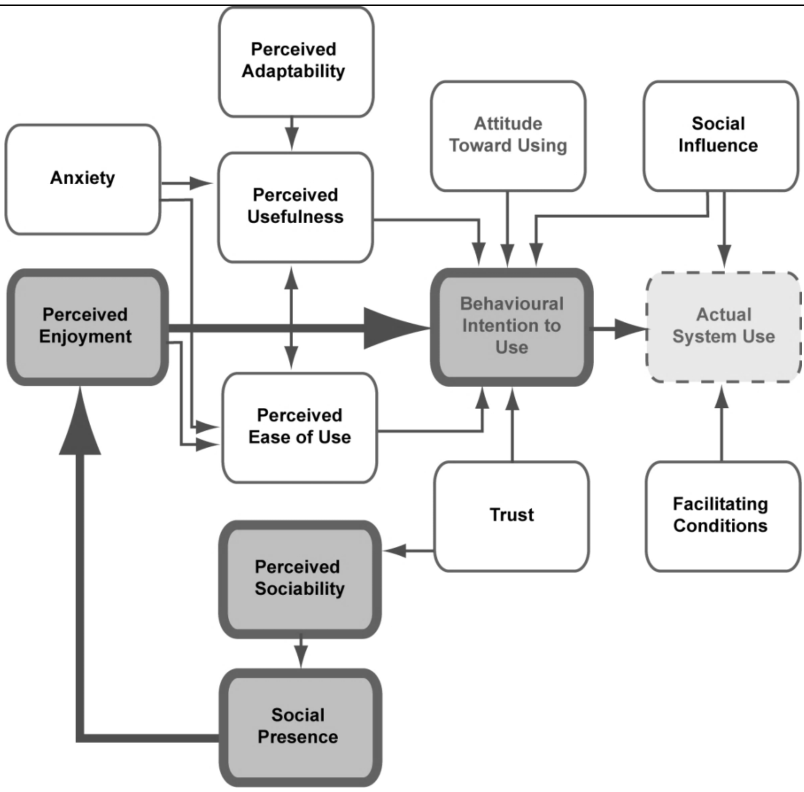


Figure 6.3. the highlighted related part to our study from UTAUT and Social Acceptance

As previously mentioned, UTAUT found that an increase in a robot's social abilities leads to an increased sense of Social Presence, Perceived Sociability and this again leads to an increase in Perceived Enjoyment. As we are mainly interested in the specific influences of increasing presence for the users, we have used this part of the UTAUT constructs as visualised in figure 6.4.

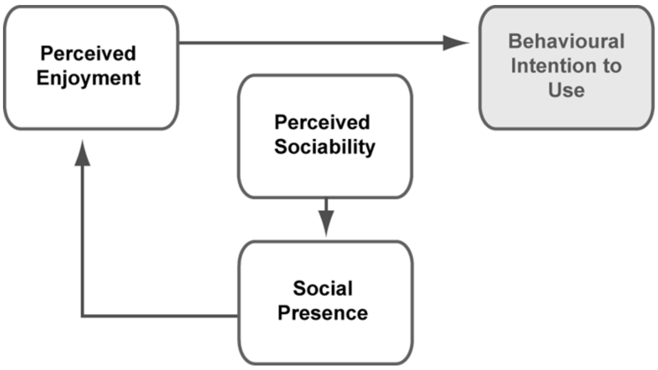


Figure 6.4. Social Acceptance

- ***Perceived Sociability***

The social acceptability of Individuals' actions is decided by the information those individuals gather from their immediate environment plus knowledge they already possess. They carry out the appropriate action, and gather feedback from observing the reaction of onlookers. This activity of interfacing and determining social acceptability happens all the time, and the result changes as time goes on. Using technology and the social acceptability factors that surround it cannot simply be attributed to etiquette or alleviating feelings of awkwardness, but need to be an ongoing evaluation taking into account a number of factors (Rico and Brewster, 2010). These factors include how the activity is performed, what the onlookers appear to be thinking about the activity and what the user wishes to gain from the activity. These factors can be appraised at an early stage of the development of the technology by a variety of methods with the aim of going beyond standard user evaluations and into the appraisal of relevant techniques of interaction.

By this, it meant that gesture and speech are both employed in the task of the production of patterns of action that may serve for others as representation of the meaning. It does not mean that they serve this task in the same way. When speech cannot be used, circumstances may make it possible to come about that gesture can be organized to do all the things that speech can do. Where speech is available, then we find that gesture and speech are employed differentially, in complementary roles, speech serving one set of communicative functions, gesture another.

A facilitating robot with the social abilities that occur between humans would help in making the interaction more familiar for users than using a static robot (Cynthia Breazeal, 2003, Cynthia Breazeal, 2004, Fong,Nourbakhsh and Dautenhahn, 2003, Li *et al.*, 1996). This is because the interaction either allows the user to make use of their existing knowledge (e.g., social models of human-human interaction) or it is sufficiently compelling that the user does not perceive the interaction as requiring effort (e.g., toy robots; (Fong,Nourbakhsh and Dautenhahn, 2003).

Thus, the following hypothesis are proposed:

H1: The Perceived Sociability of the system has a positive effect on Attitudes

- ***Perceived Enjoyment***

As it becomes more engaging or interesting to interact with the social agent, so the Perceived Sociability of the robots increases, which has been found to point to significantly higher levels of Perceived Enjoyment (Heerink *et al.*, 2010, Heerink *et al.*, 2009, Looije,Cnossen and Neerincx, 2006).

It has also been established that user acceptance is influenced by an element of pleasure when interacting with the agent (Davis,Bagozzi and Warshaw, 1992, Chesney, 2006, Sun and Zhang, 2006, Van der Heijden, 2004). This also was reported by in Heerink *et al.* (2008b) where it is reported from a long-term evaluation at eldercare institutions with the iCat that the more the people perceive a robotic system to be enjoyable, the more they intend to use it. And the more people intend to use a robotic system, the more they will actually use it. On the basis of this, the following hypothesis is developed:

H2: The Perceived Sociability of the system affects Perceived Enjoyment

H3: The Perceived Enjoyment of the system positively affects Attitudes

- ***Perceived Social Presence***

Social Presence generally causes an individual to assign social values to other to the same degree as others assign social values to them, as stated by Simmel (1910). He explains that as humans we tend to socialise with other humans for purposes other than the strictly utilitarian. It is a factor of human existence that we seek out the company of others. In a technology-mediated world, a social presence can be defined as those parts of an individual's desire for sociability that can be met through the use of technology. Hence, a socially embodied agent must be able to evoke meaningful social experiences in order to be accepted by users.

Thus, the following hypotheses are proposed:

H4: Perceived Sociability of the system positively affects Perceived Social Presence

6.5 MobiBot user cognitive experience

For a practical and theoretical reason, the cognitive part needs to be included. Firstly, for the new technology, an enhanced sense of presence is central to the use. These technologies either are now changing or are expected soon to change many of the ways we work, play, and live. Secondly, as we have stated before, Social Acceptance of robots, as positioned by UTAUT, has a direct impact on the attitude towards using a particular emerging technology such as robotic and system with human-like characters. This, in turn, influences the behavioural intention to use the system, and then the actual use of the system. In the context of communication technology that are endowed with human-like emotions and behavioural experience, however, this factor would not give a sufficiently comprehensive insight to reveal an individual's experience while using the embodied technology (Trevino and Webster, 1992). The incorporation of an embodied technology capability itself, as discussed earlier in Chapter 2, has specific characteristics that distinguish it from other kinds of systems, therefore it requires additional consideration. In our case, little is known about the individual's engagement experience while interacting with the technology and how time is perceived during the session. This experience is described in the next section.

Research into user engagement experience has been dominated by the idea of the cognitive experience. A state of deep engagement with the software is described and represents an individual state that is specific to the situation (Agarwal,Sambamurthy and Stair, 1997, Agarwal and Karahanna, 2000, Tellegen and Atkinson, 1974). Chandra,Srivastava and Theng (2009) and Zhang,Li and Sun (2006) slightly altered the definition to that of a state of deep engagement, which is also described as a state of involvement in a technology by Chesney (2006).

Agarwal and Karahanna (2000) introduced cognitive absorption as one of the significant predictors to outcomes related to technology acceptance and use (Saadé and Bahli, 2005), which can best be demonstrated by the level of user engagement (Agarwal and Karahanna, 2000). Previous studies have suggested that this type of engagement can result in positive outcomes (Ghani and Deshpande, 1994).

These outcomes may include increased experimental technology use and more positive attitudes towards target behaviours.

The original work by Agarwal and Karahanna (2000) put 5 cognitive absorption dimensions under examination: curiosity, focus immersion, heightened enjoyment, temporal dissociation and control (Figure 6.5). Our study has its focus on the impacts of 2 cognitive absorption dimensions (i.e., focus immersion and temporal dissociation) on an individual's experience. The first of the dimensions, focus immersion, relates to the construct of engagement (Webster and Ho, 1997). We used these dimensions because they are associated with emotions and cognitive characteristics in a computer interface rather than basic human experience which is represented in the rest of the dimensions. The second dimension, temporal dissociation, relates to constructs including transformation of time (Csikszentmihalyi and Mihaly, 1990), and telepresence/time distortion (Novak,Hoffman and Yung, 2000).

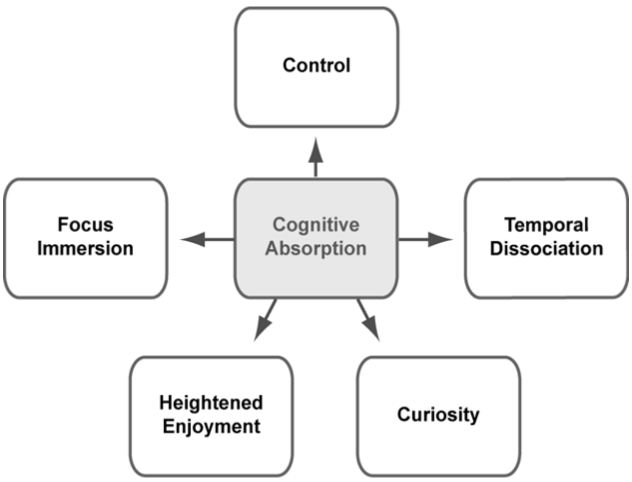


Figure 6.5. Cognitive Absorption

- ***Focus Immersion***

The terms ‘presence’ and ‘immersion’ can be considered either interchangeable, or one can consider the perception of presence with an immersion as an incorporated factor. Presence as a concept can be defined in terms of ‘the extent to which a person’s cognitive and perceptual systems are tricked into believing they are somewhere other than their physical location’ (Emilee Patrick *et al.*, 2000).

This definition gives us to understand that our perception of presence will increase proportionally in line with our increase in perceived immersion, and the commanding tasks. When a user is focusing their attention solely on an experience and ignoring other demands, it can be stated that the user’s level of focus immersion is high. This will then result in a reduced cognitive burden level associated with the performance of that experience (Evaristo and Karahanna, 1998). If the experience is related to technology, the user’s attitude towards that technology will be predicted by their seeming to be experiencing a familiar social situation.

As previously stated, this familiarity with a social experience comes about when technology follows the social rules. When a robot or screen agent is a character with a familiar social presence and sociability, the user’s expectations will influence their immersion experience. Based on the aforementioned, the current research proposes:

H5: Perceived Social Presence of the system positively affects Focus Immersion

- ***Temporal Dissociation***

Temporal dissociation occurs when users lose track of time. Previous work has suggested that there are positive outcomes from this type of cognitive absorption variable (Weniger and Loebbecke, 2010, Ghani and Deshpande, 1994). These users are also likely to hold positive attitudes concerning the system.

In relation to focused immersion, Brown and Cairns (2004) described it as an experience in one moment in time. By this, we believe that temporal dissociation and focused immersion are correlated. Individuals with a high propensity to engage in events with total attention while interacting with a system will also experience a loss of consciousness and time.

This assumption is also supported by the theory of flow developed by Csikszentmihalyi and Mihaly (1990) who describes 'the state in which people are so involved in an activity that nothing else seems to matter.'

Additionally, a socially embodied agent elicits more social responses, and results in more Temporal Dissociation. This contributes to the emergence of temporal dissociation, while interactivity features stimulate a user's feeling of a familiar social experience. and therefore, the current study proposes:

H6: Perceived Social Presence of the system positively affects Temporal Dissociation

6.6 Embodiment of the System

All of the previous parts highlighted the important of the embodiment part of the system as they depend on it. An embodiment of a robot may affect peoples' social perceptions of them (Kose-Bagci *et al.*, 2009, Fong,Nourbakhsh and Dautenhahn, 2003) which in turn affect the cognitive and social experience. This embodiment aspect includes physical appearance and social ability in a robotic interface as explained by Bartneck and Forlizzi (2004) point out in relation to this embodiment: 'the appearance and function of a product, affects the way that people perceive it, interact with it'. The consistency of social abilities and appearance in relation to forming and meeting a person's expectations of the appropriateness of technical capabilities and social cues is an important factor as highlighted by different studies (Minato *et al.*, 2006, Goetz and Kiesler, 2002, Goetz,Kiesler and Powers, 2003).

- ***Physical appearance***

A robot's physical appearance affects a person's attitude towards robots. Mori's "Uncanny Valley" hypothesis is a classical work, formulated in the 1970s (Mori and M, 1970). According to Mori's hypothesis, there is a nonlinear relationship between familiarity or arousal and the humanoid appearance of a robot. If a robot's appearance is too similar to that of a human, the viewer may experience uncanny feelings such as negative attitudes or anxiety. This study was among the pioneering studies looking into the area of personal robots.

It focused on human attitudes towards service robots, with particular emphasis on physical appearance (Dario,Guglielmelli and Laschi, 2001). thus, the study hypothesises:

H7: Physical Appearance of the system positively affects Attitudes

- ***Social abilities***

Social abilities encourage a user's social behaviour, to interact with the robot, and to give them the motivation to continue to interact with the robot, once the initial 'novelty' has worn off – beyond the first few minutes. Our belief is that perceived sociability and social presence will be positively affected. On the basis of this, the following hypotheses is developed:

H8: Social Abilities of the system positively affect Perceived Sociability

H9: Social Abilities of the system positively affect Social Presence

As a result Figure, 6.6 was the transplantation of the previous hypothesis and review.

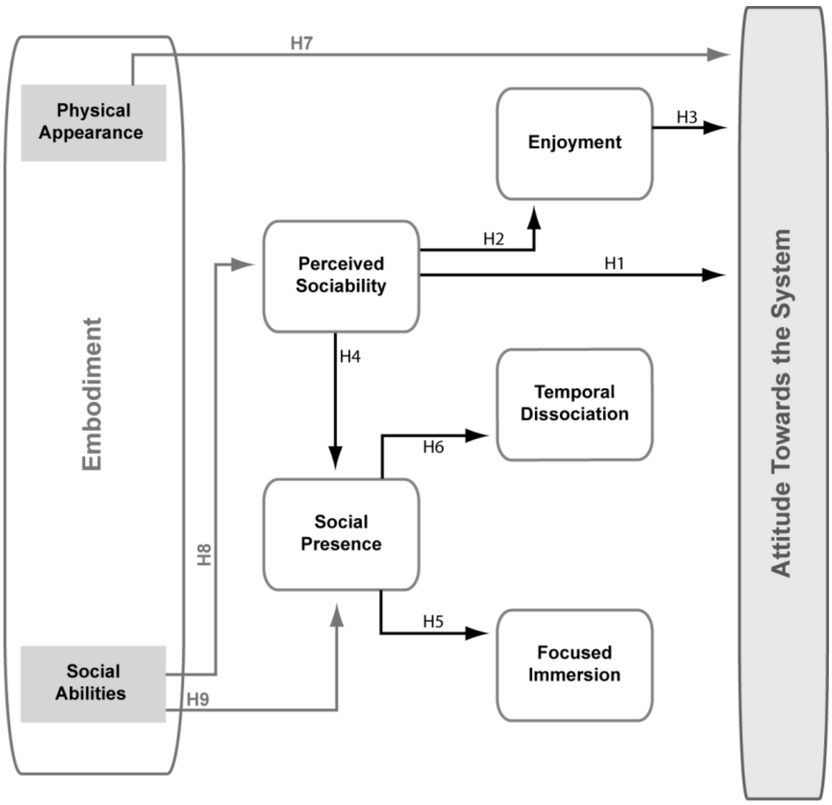


Figure 6.6. Research model based on UTAUT

6.7 Research Measurement Development

Most of the measurement methods adopted in this thesis have been guided by the measurement tools which have previously been widely used in this field.

6.7.1 Measuring Perceived Sociability

Perceived Sociability of can be defined as "*The perceived ability of the robot to perform sociable behaviour."*(Heerink et al., 2009, p. 529). The perceived sociability measurement terms were developed by Heerink *et al.* (2010) and scholars have used them extensively since then as a guide.

These terms will be used in the current study with slight modifications to make them correspond to the system's characteristics (Table 6.2).

Table 6.2. Measuring Perceived Sociability Items

Heerink et al. (2010)	Question	5 Point Likert Scale
Perceived Sociability PS		
1	**I consider the System a pleasant conversational partner**	• Definitely true • Probably true • Neither true nor false • Probably false • Definitely false
2	**I find the System pleasant to interact with**	• Strongly agree • Somewhat agree • Neither agree nor disagree • Somewhat disagree • Strongly disagree
3	**I feel the System understands me**	• Definitely true • Probably true • Neither true nor false • Probably false • Definitely false
4	**I think the System is**	• Very unappealing • Somewhat unappealing • Neither unappealing nor appealing • Somewhat appealing • Very appealing

6.7.2 Measuring Social Presence

Similar to the perceived sociability construct, a social presence scale was developed by Heerink *et al.* (2009). Heerink defines social presence as "*The experience of sensing a social entity when interacting with the robot*" (p.529). The following items used Heerink's scale as a guide, with slight alterations, to evaluate the system characteristics (Table 6.3).

Table 6.3. Measuring Social Presence Items

Heerink et al. (2010)	Question	5 Point Likert Scale
Social Presence SP		
1	**While using the System, the interaction with my partner feels**	• Very personal • Somewhat personal • Neither personal nor impersonal • Somewhat impersonal • Very impersonal
2	**While using the System, the interaction with my partner feels**	• Very unnatural • Somewhat unnatural • Neither unnatural nor natural • Somewhat natural • Very natural
3	**While using the System, the interaction with my partner is close**	• Definitely false • Somewhat false • Neither false nor true • Somewhat true • Definitely true
4	**While using the System, the interaction with my partner feels**	• Very human • Somewhat human • Neither human nor inhuman • Somewhat inhuman • Very inhuman
5	**My experience of interacting with my partner using the System is consistent with a real world experience**	• Strongly agree • Somewhat agree • Neither agree nor disagree • Somewhat disagree • Strongly disagree
6	**The forward movement of the System makes me feel as if I am having a face to face with my partner interaction**	• Definitely true • Probably true • Neither true nor false • Probably false • Definitely false
7	**The nod movement of the System makes me feel as if I am having a face to face interaction with my partner**	• Strongly disagree • Somewhat disagree • Neither disagree nor agree • Somewhat agree • Strongly agree
8	**The forward tilt of the System makes me feel as if I am having a face to face interaction with my partner**	• Definitely false • Probably false • Neither false nor true • Probably true • Definitely true
9	**The backward tilt of the System makes me feel as if I am having a face to face interaction with my partner**	• Strongly agree • Somewhat agree • Neither agree nor disagree • Somewhat disagree • Strongly disagree

6.7.3 Measuring Enjoyment

De Ruyter *et al.* (2005) UTAUT covers the social factors which the perceived enjoyment one of them. It defines as "*Feelings of joy or pleasure associated by the user with the use of the robot*" (p. 696). Perceived enjoyment was measured using four semantic differential scales (Table 6.4)

Table 6.4. Measuring Enjoyment Items

De Ruyter et al. (2005)	Question	5 Point Likert Scale
Enjoyment ENJ		
1	**Interacting with my partner using the System is enjoyable**	• Strongly disagree • Somewhat disagree • Neither disagree nor agree • Somewhat agree • Strongly agree
2	**Interacting with my partner using the System is not boring**	• Definitely false • Probably false • Neither false nor true • Probably true • Definitely true
3	**Interacting with my partner using the System is interesting**	• Strongly agree • Somewhat agree • Neither agree nor disagree • Somewhat disagree • Strongly disagree
4	**Interacting with my partner using the System is entertaining**	• Definitely not • Probably not • Might not or might • Probably yes • Definitely yes

6.7.4 Measuring Perceived Temporal Dissociation

An additional variable that can be relevant to the acceptance of the platform is Temporal Dissociation. The scales used in the study reported here consisted of five items capturing perceived Temporal Dissociation, from Agarwal and Karahanna (2000) with slight alterations made to adapt it to our study (Table 6.5).

Table 6.5. Measuring Temporal Dissociation Items

Agarwal and Karahanna (2000)	Question	5 Point Likert Scale
Temporal Dissociation TD		
1	**Sometimes I lose track of time while I am interacting with my partner using the System**	• Strongly agree • Somewhat agree • Neither agree nor disagree • Somewhat disagree • Strongly disagree
2	**Time flies while I am interacting with my partner using System**	• Definitely false • Probably false • Neither false nor true • Probably true • Definitely true

6.7.5 Measuring Focused Immersion

Focused immersion suggests that all of the attentional resources of an individual are focused on the particular task. To measure this factor, we used an adapted version of the focused immersion in cognitive absorption model (Saadé and Bahli, 2005), adapted for our system (Table 6.6)

Table 6.6. Measuring Focused Immersion Items

Cognitive absorption Saade et al. (2004)	Question	5 Point Likert Scale
Focused Immersion FI		
1	**While interacting with my partner using the System I am able to block out most other distractions**	• Strongly agree • Somewhat agree • Neither agree nor disagree • Somewhat disagree • Strongly disagree
2	**While interacting with my partner using the System, I am absorbed in what I am doing**	• Definitely false • Probably false • Neither false nor true • Probably true • Definitely true
3	**While interacting with my partner using the System, I am immersed in the task I am performing**	• Strongly disagree • Somewhat disagree • Neither disagree nor agree • Somewhat agree • Strongly agree

6.7.6 Measuring Attitudes Towards the System

Attitudes were measured following the TAM methodology (Heerink *et al.*, 2010) (Table 6.7).

Table 6.7. Measuring Attitudes towards the System Items

Attitude towards technology Heerink (2010)	Question	5 Point Likert Scale
Attitude towards technology ATT		
1	**Using the System is a**	• Very good idea • Somewhat good idea • Neither good idea nor poor • somewhat poor • Very poor
2	**The System will make conversation more interesting**	• Strongly agree • Somewhat agree • Neither agree nor disagree • Somewhat disagree • Strongly disagree
3	**The System is**	• Very inspiring • Somewhat inspiring • Neither inspiring nor uninspiring • Somewhat uninspiring • Very uninspiring
4	**I prefer to do not video calling using this System**	• Strongly agree • Somewhat agree • Neither agree nor disagree • Somewhat disagree • Strongly disagree
5	**I would like to experience this again**	• Definitely true • Probably true • Neither true nor false • Probably false • Definitely false
6	**This experience is not consistent with my expectation from the System**	• Strongly disagree • Somewhat disagree • Neither disagree nor agree • Somewhat agree • Strongly agree
7	**I am satisfied with the experience**	• Definitely yes • Probably yes • Might or might not • Probably not • Definitely not

6.7.7 Measuring Embodiment

The embodiment describes the relationship between a system and its environment and can be measured by investigating the different channels which have an impact on social expectations (Astrid Weiss *et al.*, 2009). As previously explained, this measure concerns the robot's physical appearance and social abilities. The current study has adopted elements from the Astrid Weiss *et al.* (2009) embodiment questionnaire, to cover both the physical appearance elements and the social abilities of the proposed system (Table 6.8).

Table 6.8. Measuring Embodiment Items

Embodiment (EMB) Weiss et al. (2009)	Question	5 Point Likert Scale
Physical Appearance PA		
1	**The design of the System**	• Very appealing • Somewhat appealing • Neither appealing nor unappealing • Somewhat unappealing • Very unappealing
2	**The size of the System**	• Strongly like • Somewhat like • Neither like nor dislike • Somewhat dislike • Strongly dislike
3	**The height of the System**	• Strongly dislike • Somewhat dislike • Neither dislike nor dislike • Somewhat like • Strongly like
4	**The colour of the System**	• Strongly dislike • Somewhat dislike • Neither dislike nor dislike • Somewhat like • Strongly like
5	**The System looked similar to a human**	• Strongly agree • Somewhat agree • Neither agree nor disagree • Somewhat disagree • Strongly disagree
Social Abilities SA		
1	**The System has humanlike behaviour features**	• Strongly agree • Somewhat agree • Neither agree nor disagree • Somewhat disagree • Strongly disagree
2	**The System's forward movement features**	• Very appealing • Somewhat appealing • Neither appealing nor unappealing • Somewhat unappealing • Very unappealing
3	**The System's forward tilt movement features**	• Very appealing • Somewhat appealing • Neither appealing nor unappealing • Somewhat unappealing • Very unappealing
4	**The System's backward tilt movement features**	• Very unappealing • Somewhat unappealing • Neither appealing nor unappealing • Somewhat appealing • Very appealing

Embodiment (EMB) Weiss et al. (2009)	Question	5 Point Likert Scale
5	**The System's ability to nod**	• Very unappealing • Somewhat unappealing • Neither appealing nor unappealing • Somewhat appealing • Very appealing
6	**The System's ability to shrug**	• Very appealing • Somewhat appealing • Neither appealing nor unappealing • Somewhat unappealing • Very unappealing

6.8 Summary

This chapter demonstrates the theoretical framework used in this thesis to evaluate a MobiBot TP system, including the hypothesis development. Perceived total immersion and embodiment values were integrated into the UTAUT, in addition to the perceived sociability and social presence, as these were the main factors affecting attitudes towards the MobiBot TP system. Finally, explanations of the study measurement instrument design for the main study and both the comparative studies have been presented.

Chapter 7 MobiBot System Evaluation

This chapter presents the system evaluation procedure and evaluation results. A user trial was conducted in order to evaluate the system acceptance; user trials using representative users were conducted. A comparative study of the interaction method was accomplished in addition to a future usage study considering the overall system features. Qualitative and quantitative analyses of the obtained data are presented followed by a discussion of the results.

7.1 Introduction

This chapter elaborates on the system evaluation which includes data collection, preparation, and analysis in order to answer the research question and test the proposed hypotheses. Firstly, the objectives used to evaluate the system are discussed in Section (7.2). In Section (7.3), the study methods, data preparation and sample characteristics are discussed. Section (7.4) presents the technical equipment used in this study. A description of the tools used in this study presented in section (7.5), followed by the psychometrics properties for them in section (7.6). Pilot testing was presented in section (7.7). Section (7.8) present data analysis methods, followed by its result in section (7.9). The comparative study was presented in section (7.10). Users' preference to use the system in the future was evaluated against part of the customised Unified Theory of Acceptance and Usage of Technology Model (UTAUT) in Section (7.11), with the future use in section (7.12). Section (7.13) tests the study hypothesis. Results for these objectives were presented in section (7.14), (7.15) and (7.16). Finally, the results of the analysis are discussed in Section (7.17), followed by summarised highlights in Section 7.18.

7.2 Why examine the mimicked movement?

Using user's input by actively specifying the movement is one of the direct translation and relatively easy to implement as a new features in TP robotics system. Though using this type of inputs might improve or add more personal and natural interaction, it has its downside. First, the driver of the robot must consciously tend to control this, which unlike natural gesture and becomes more burdensome in a scenario where the driver is also holding a conversation and also driving the robot as Adalgeirsson and Breazeal (2010) explained "*it is difficult to manage the boundary between operating the robot and doing something locally with your arms that is unrelated to the interaction*" (p.40).

Therefore, the aim of the testing was to determine both the effectiveness of the method and the users' acceptance and preferences for the interaction method, in this specific scenario. In other words, the test was conducted in order to study the role of an interaction method in influencing the independent variables, from the users' perspective and point of view.

- **Video call with MobiBot TP Automatic Vocal-triggered movements Vs. Mimicking movement**

In order to understand how users perceive conventional TP systems against MobiBot TP systems, the current study conducted a comparison between both methods. The same measurement tool was used to evaluate both systems. The seven factors were considered in evaluating the users' attitudes towards both type of movements systems, perceived Sociability, Social presence, Enjoyment, Temporal Dissociation, Focused Immersion and embodiment. The questionnaire items that were utilised were drawn from the main study questionnaire. For MobiBot automatic vocal-triggered movement questionnaire, 40 items in total were used to evaluate it, distributed as follows: perceived sociability 4, social presence 9, enjoyment 4, Temporal Dissociation 2, focused immersion 3, attitudes towards the system 7 and Embodiment 11 which was also the same for the mimicked movement (Appendix B).

7.3 MobiBot TP system evaluation: Data preparation and sample characteristics

The Socially Interactive Telepresence Robotic System (MobiBot) for video calls was evaluated from the representative user's perspective, which was divided into three main objectives. The first objective was to evaluate the vocal-triggered movement interaction method for future use. The second objective was to evaluate the user's attitudes towards the system based on the Unified Theory of Acceptance and Usage of Technology Model. Finally, the third experiment was intended to evaluate the MobiBot automatic vocal-triggered movement system against the MobiBot mimicking movement system. A within-subject design was used, with seventy-six representative users participating in the study, which was carried out in the Private room in the University of Strathclyde.

7.3.1 Methods

- ***Participants***

Seventy-six representative users participated in the study. The ages of the subjects varied between 18 and 40. They were all from the University of Strathclyde, both undergrad and postgrad students, and their subjects represented a variety of university majors. The inclusion criteria for this study were subjects should be able to speak English fluently. Whereas the exclusion criteria was hearing or visual impairments and non-fluent English speakers.

- ***Recruitment***

i. Sampling framework

A nonprobability convenience sampling strategy was employed for data collection, as subjects were selected because of their convenient accessibility and proximity to the researcher.

ii. Sampling techniques

After supervisor and department confirmation, an advertisement was distributed electronically using the advice centre newsletter email. The email included details of the purpose of the study and information regarding participation and what this would entail. Besides this email, hard copies of the advertisement were displayed around the university campus to ensure as large a number of volunteers as possible.

In order to overcome the bias resulting from sampling and coverage, the advertisement was distributed in such a way as to cover different university levels and different subject areas.

iii. Sample size

This experiment’s main purpose was to explore the potential for replicating the part of nonverbal behaviour that supports interaction and feedback functions between people, which could be useful in the development of a model for human social behaviours to implement in a Telepresence platform. Therefore, there was no defined number of participants to be recruited, rather the researchers aimed to identify and recruit as many as were willing to participate. In total, we managed to get Seventy-six subjects.

7.3.2 Study Design

One group in two conditions quasi-experimental design was used in this study, as it utilises a similar structure to experimental design, however, it lacks key ingredients (Shadish,Cook and Campbell, 2002).

1. It does not have the time and logistical constraints associated with true experimental designs.
2. Since most the variables are controlled, it reduces the difficulty and ethical concerns
3. Because of the controlled environment, the time and recourses may be reduced.
4. Researcher can tailor the study to fit their needs with while still maintaining the validity of the design.
5. It lacks the randomality that we find in the true experiment as well as the blinding.

Therefore, the experiment involves two different conditions for producing the movement, Vocal-triggered automatic and Mimicked movements.

i. *The mimicking movement condition*

Participants interacted through a video call where part of the body and head movements replicated manually by the researcher using MobiBot. The movement can be described as a big and slow movement in this condition. The researcher connects MobiBot via USB cable to the laptop to generate the movement.

ii. The vocal-triggered automatic movement condition

Participants interacted through a video call where part of the body and head movements replicated automatically using Vocal stream. Participants wear a microphone to ensure the transmission of the sound to generate the movement by MobiBot. Small and quick movement were used in this condition. This was based on Rudolf Laban movement theorist (1879-1958) (Maletic, 1987), where he identify three movement elements time, space and force. He characterizes them through describing their opposite extremes: fast-slow for time; near-far for space; weak strong for weight; and bound-released for flow. By varying these elements the quality of movement will changes accordingly.

This as consequences will not only change how the movement looks from the outside, but also how it feels for the person performing it which will affect users sense of presence.

In our design, the initial angle of the mobile phone holder (Servo G) at rest of the mobile phone holder was approx. 115deg. from the horizontal i.e. leaning back away from the user.

Movement of Servo G alters the mobile phone holder angle of elevation from the user. Movements for servo G varied as described here,

- The 30deg programed movement of servo G corresponds to 16mm up/down in 300ms in 10ms per deg (116-121.5 = 5.5deg).
- The 20deg programed movement of servo G corresponds to 20deg servo up/down corresponds to 10mm up/down in 150ms gradual 1 deg (116-119.5 = 3.5 deg).
- The 10deg programed movement of servo G corresponds to 5mm up/down in 150ms (116-117 = 1deg).

This design involved one set of the questionnaire being repeated, with the measurement taken twice in different conditions on one group of subjects. After performing two tasks in the two conditions, subjects answered the questionnaire. In this study, the independent variable was the questionnaire construct, where the dependent variable was the movement type.

7.3.3 Task

Small talk has been defined as a neutral, non-task-oriented style of conversation about things which are not important, where no specific goals need to be achieved, often between people who do not know each other well. The decision to use small talk was based on the fact that it helps in building a rapport among the interlocutors, provides time for them to "*size each other up*", establishes an interactional style, and to allows them to establish their reputations (Dunbar, 1998).

Small talk can be about neutral topics (e.g., weather, aspects of the interlocutor's physical environment) or in which personal experiences, preferences, and opinions are shared (Laver, 1975).

Schneider (1988) suggests three categories from which topics that might occur during small talk; immediate situation, external situation and communication situation topics.

Following Schneider and considering our experimental setting at a university campus with student participants, we chose to classify topics as follows:

- ***External situation***

The students talk about studies or the university in general (as a super situation for recordings at a university), friends or other people they know, or public topics such as music or movies.

- ***Communication situation***

Interlocutors focus on topics concerning themselves, such as their places of origin, hobbies, going out at night, personal habits or even their health.

We chose to use the small talk task because it has the following characteristics which are beneficial to our evaluation:

1- It enacts social cohesiveness
2- It sparks an active conversation between both sides or avoids undesirable silence
3- It reduces inherent threat values of social contact
4- It helps to structure social interaction

7.3.4 Procedure

- ***Recruitment procedure***

The consent of the University of Strathclyde was obtained as a first step and in particular the approval of the department of Design, Manufacture and Engineering Management (DMEM). This approval was obtained through the academic supervisor. The department of Design, Manufacture and Engineering Management is the authoritative body that provides the final consent of study approval. The approval letter included all the details concerning the background of the study, aims and objectives, and exactly what was the needed from the subjects. The original copy of the approval was kept filed with the academic supervisor and the researcher.

The researcher arranged an appointment with each subject to describe the study and what was required from them, before signing the consent form in order to start the experiment.

- ***Study Procedure***

Procedure for participants

Prior to starting the experiment, participants were provided with a hard copy of the information sheet and a copy of the consent form. This was to remind them about the rationale for the study, and included aims and objectives, the nature and maximum duration of the session, which their feedback was going to be collected using a digital questionnaire, and how the data would be shared and stored. Then, two disclaimer signatures were obtained: one indicating acceptance of participation in the study and the other consenting to the video recording of the session.

The purpose of the research was clearly stated in the form, and it was confirmed to them that there was no risk to those who were involved in the study.

As the experiment was to be recorded using a video camera, the researcher confirmed to the subjects that any data they provided would be kept confidential. No identifiable data would be included in any report arising from the study. Completion and return of the online questionnaire constituted consent. This was because we were seeking to retain anonymity of data as far as possible, and completion of a consent form would identify the participant. Furthermore, it was made clear to them that they had the right withdraw from the study at any point, and that withdrawal would have no repercussions for them.

Once the researcher had ensured the completion of these actions, the subjects were asked to move to a laboratory room on the University campus. As the purpose of this experimental test was to gather real-time data under two separate conditions - a video call (Skype), and using a telepresence robot (MobiBot), the researcher explained that the subjects would be paired with the researcher, and they would be located in different rooms. As the task was to complete a small personal talk, participants were given a set of conversation topics to choose from, or if they preferred, they could choose a topic of their own. Then, subjects were asked to seat themselves wherever they liked, facing the device (MobiBot) with less than 600mm between themselves and the device. The subjects were given a short demonstration of the use of MobiBot, and a package of questionnaires to answer at the end of each condition. In the vocal-triggered automatic movement condition, the subject was also asked to speak into a microphone in order to allow for the device to collect the vocal stream to translate into movement. Whereas, in the controlled condition, the movements were produced manually using a program installed in the researcher's laptop.

In general, subjects were instructed:

"*As you will be paired with me in different rooms, will not be able to see each other physically, but will see one another's faces through the video call using the phone camera placed on top of MobiBot in front of you. Using MobiBot, both of us will work collaboratively together and will be able to hear and speak using video call. After choosing your preferred topic, we will complete an interpersonal small talk for less than 20 min. For each of two separate conditions, the talk topic will be different. You can respond with any answer that you feel comfortable with, and you can interrupt me if you want to make any comments or share anything related to the topic we are discussing. I will stop you when the time is up and you will be asked to complete an online questionnaire. Your answers and your details will remain* *anonymous throughout the interaction, and also when the video recordings are analysed."*

The purpose of the experimental test was to gather real-time data in two isolated shared visual contexts; however, the procedure was slightly different in the two experimental conditions. In the vocal-triggered automatic movement condition only, the researcher assisted the subject with placement of the microphone, and ensured that the subject was comfortable before asking to begin the test.

To avoid the learning effect, half of the subjects were started in the first condition, and the other half were started in the second condition.

Procedure for the researcher

During the tests, the researcher adopted and maintained a consistent speaking across all the participants. The style adopted was warm, accepting, and showing interest without excessive friendliness or appearing to be ingratiating. In order to achieve and maintain the correct style, the researcher limited the number and content of the comments made during the test to:

- "Could you tell me anything else?"
- "Are there any other things?"
- "Could you expand on something you've already mentioned?"
- "Would you like to go into detail about anything?"
- "Is there anything else?"
- "Have you left anything out that might be important?"

The researcher tried not to direct the talk, but instead listened and let the conversation take its natural course, to encourage the dialogue between researcher and participant.

For both conditions, the researcher monitored the environment, to ensure that everything was working as planned.

Procedure for the location

The room where participants completed the experimental tasks was a private, locked participant room, ensuring privacy and a quiet environment. The room was equipped with MobiBot and camera to record the interaction as previously explained. The video camera was placed less than 2m distant from the participants. It was also placed in a way to ensure coverage of the participant as well as MobiBot (Figure 7.1).

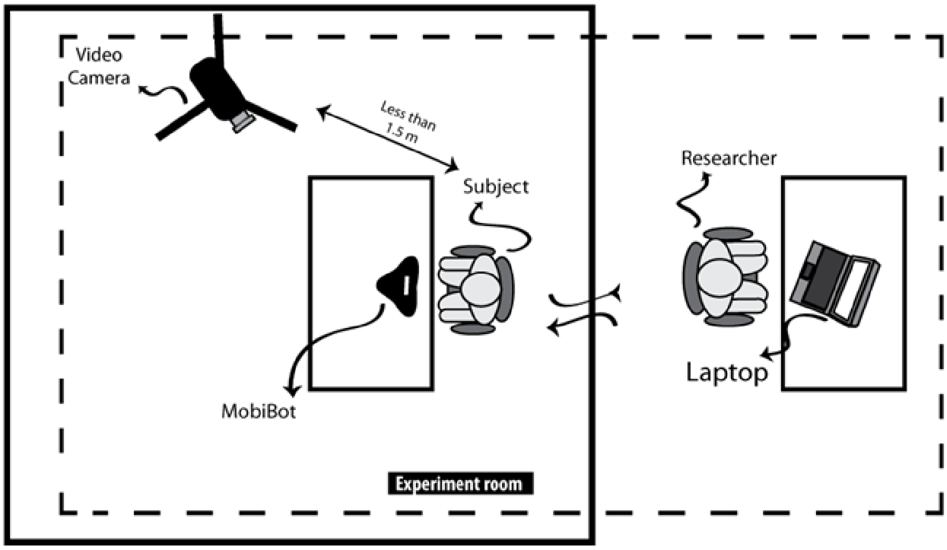


Figure 7.1. Experiment Design

7.4 Technical equipment

- MobiBot: We developed a robotic platform which is able to replicate some of the speech related gestures, by only relying on the vocal stream. This was described in Chapter (5).
- A laptop was used by the researcher for video calls for both conditions.
- The Skype application was used for the video call.
- A microphone was used by the subject in the vocal-triggered automatic condition

7.4.1 Tools

The survey primarily aimed to collect data through an online questionnaire, using Qualtics Online Survey. Consideration was given to the speed and efficiency of data collection and the ability to maintain maximum control over data, thus providing more accurate data management (Weber et al., 2005).

7.4.2 Data Collection

Questionnaire: Users' responses were collected using our digital questionnaire, which included the evaluation of 7 measurements. Those measurements were; embodiment, perceived sociability, social presence, enjoyment, Temporal Dissociation, Focus Immersion and attitude toward the technology, as described in Chapter (6)

7.5 Development of the questionnaire

Most of the measurement methods adopted in this thesis have been guided by the measurement tools which have previously been widely used in this field, as discussed in the previous chapter.

Subjects were requested to give a rating of each statement using a 5-point Likert scale. The design of the questionnaire took into consideration a number of issues in order to give a measurement of acquiescence bias. 2 negatively worded, 23 positively worded and 11 objective statements were combined in order to minimise the risk of acquiescent response bias. The scales of response for the statements utilised the extreme forms of each adjective for the first and fifth points, and two antonyms on the second and fourth points, with the centre point being a neutral form. For instance – scale of responses could range from Strongly Agree = 4, Agree = 3, Neutral = 2, Disagree = 1 and Strongly Disagree = 0.

The alignment of the scale points was also reversed to put the negative end to the left after it had been to the right on the previous section. Randomisation of the questions was also used irrespective of the grouping for analysis, which meant that questions from the same analysis group were not put in sequence, but the arrangement was made in order to make it simple for the user to complete the questionnaire.

7.6 Psychometric properties

Any newly developed tool should show basic psychometric attributes in order to be statistically acceptable. Addressing validity, reliability and components of principle analysis will give an idea about the weakness or the strength of the measure (David A Cook and Beckman, 2006).

7.6.1 Reliability

Reliability is a measure of the questionnaire's stability, internal consistency or repeatability (Jack and Clarke, 1998). A common measure of reliability is via the use of Cronbach's alpha. This is a statistic that correlates items to work out whether those items are actually measuring the same area (Bowling, 1991, Bryman and Cramer, 2005, Jack and Clarke, 1998). In the case of strong internal consistency, a Cronbach's alpha of over α 0.70 should be achieved if the questionnaire is under development. A more established questionnaire should achieve α 0.80 (Bowling, 1991, Bryman and Cramer, 2005). The rule of interpreting the α value that is most often adopted is shown in (Table 7.1).

Table 7.1. Cronbach's Alpha value interpretation - George and Mallery (2003)

Excellent	Good	Acceptable	Questionable	Poor	Unacceptable
α value >.9	α value >.8	α value >.7	α value >.6	α value >.5	α value<.5

Reliability analysis was carried out for all the questionnaire constructs; embodiment, perceived sociability, social presence, enjoyment, temporal dissociation, focus immersion and attitude toward the system using data from both conditions.

- ***Validity***

Validity is a term to describe the usefulness, appropriateness, correctness and meaningfulness of inferences made by a researcher (Fraenkel,Wallen and Hyun, 1993).

This includes face and content validity.

i. Face validity

Face validity is important in that it can assist in ensuring good response rates, although it is thought to be one of the weakest types of validity. This type of validity was handled by checking out the overall questionnaire appearance, and also putting the questions into a clear and logical order. The wording was checked for ambiguity, and also appropriateness, sequence, clarity and suitability (Brink,Van der Walt and Van Rensburg, 2006).

ii. Content validity

Content validity means that the content should seem logical and comprehensive in order to cover the several domains which are included in a questionnaire. In this case, the questionnaire was assessed by a technician and two lecturers from the university. Most of the questionnaire had been adapted from questionnaires previously used in the field, but the items of each construct were still refined. In actual fact, the number of items in each construct needed to be reduced because it was intended that a holistic measurement of the interface characteristics or components (independent variables) would be completed. Each component of the system was assessed independently for its impact on the dependent variable (attitudes). The questionnaire was further evaluated by asking three Ph.D. student colleagues in the university to state whether they felt the questions were easy to understand and clear, and if they felt the sequence was logical (Brink,Van der Walt and Van Rensburg, 2006).

Testing the questionnaire before starting the process of data collection is a fundamental step (Boynton and Greenhalgh, 2004, Oppenheim, 2000). The section below describes the procedure followed for testing the questionnaire in this study.

7.7 Pilot testing

A pilot study allows the determination of the process by which the feasibility of the main study stages can be assessed, and supports the management of users and data optimisation issues. The resources can also be specified, ranging from time to budget, to the technical aspects and challenges, which can be pinpointed prior to the main study (Thabane *et al.*, 2010).

A pilot study was undertaken in the current research as an initial stage in the system development. The aim of this pilot study was to examine the feasibility of the study design in the proposed research environment. The study objectives were:

1. Test the research environment
2. Test the clarity of the questionnaire
3. Observe the user experience
4. Identify potential difficulties in collecting data

Representative subjects were involved in the pilot study, as recommended by Thabane *et al.* (2010). A total of five users were invited to participate and provide an initial evaluation of the system, plus demographics and video call frequent questions. The subjects also completed the two online questionnaires upon completion of the two conditions experiment.

The results collected from the questionnaires offered an initial appraisal of the system. For example, the users made positive statements regarding using MobiBot - which they were able to gain better information from the social movement of the system compared to a normal video call method. The results of the pilot study guided the required system design improvements. Observing the users' experience indicated that the interaction with MobiBot could be a more enjoyable and effective way of interaction.

In regard to the study design, there were no changes made, as the researcher used the first experiment as a guide for this experiment.

7.8 Data analysis

The quantitative and qualitative data were analysed separately, and interpretations which led to answering the research questions were integrated. The following section explains the approaches to data analysis and what the analysis revealed in each phase of the current research.

7.8.1 Quantitative data analysis tool

The SPSS software package (version 19 for Windows) was used to analyse the questionnaire data.

7.8.2 Treatment of missing data

'Mean imputation' is the method that was adopted in the current thesis to deal with missing data. This method is frequently used and is a way of using a mean value to replace any empty data cells.

A scan of the dataset for missing data was carried out. In the case of constructs that were reliant upon subscale calculations, missing data was treated via imputed means. This approach is a conservative method with avoids changing the distribution, and does not rely on the experimenter approximating data. However, a drawback of the method is that it reduces overall variance (Tabachnick,Fidell and Osterlind, 2001). This only becomes a problem if a large amount of data is missing, which fortunately was not an issue with this dataset. Imputed means were calculated for each condition to replace missing data points, rather than relying on the overall mean.

7.8.3 Questionnaire

- ***Descriptive analysis***

In order to decide which type of statistical analysis should be used with the data obtained, the researcher needs to understand the main features observed in the data. Therefore, the researcher conducted descriptive analysis as a first stage in the overall analysis. Means and standard deviations were calculated for the questionnaire which included embodiment, perceived sociability, social presence, enjoyment, temporal dissociation, focus immersion and attitude toward the system. The characteristics of the study participants were also recorded, and comprised gender, age group, and frequency of making or receiving video calls using services.

A number of parametric tests rely on the assumption of normality (normally distributed data, 0 mean, 1 standard deviation and symmetrical bell-shaped curve), so it was critical that a normality test was carried out on the data. A histogram can give a general impression of normality. Normal Q-Q plots of value and expected value can also be plotted on a graph; a straight line indicates normality, deviation from straight is non-normal.

- ***Inferential analysis***

Analysis was performed in order to address the objectives of the research, and to determine between and within group differences; the types of analysis undertaken are described in the next section. For the purposes of these analyses, a 0.05 significance level was used, which meant that the null hypothesis was rejected if the p-value was lower than 0.05.

i. The univariate analysis

Preliminary analysis using the Pearson correlation was performed before embarking on multivariate analysis. The ‘Linear regression’ coefficient is a bivariate correlation, and this must lie between +1 and -1. The size of an effect is measured using the correlation coefficient, and a result of +0.5 indicates a large effect, +0.3 a medium effect and + 0.1 a small effect. A coefficient result of -1 shows a perfectly negative relationship, +1 a perfectly positive relationship, and 0 no linear relationship between the factors (Field, 2014).

ii. The multivariate analysis

The selection of linear regression analysis was made because: i) the outcome variable responses were ordinal and ii) it was important to be able to assess any potential interaction effects between the explanatory (independent) variables.

Linear logistic regression makes it possible to identify potential interaction effects existing between independent variables, and in this case, provides better control and an improved understanding of the effect that these factors have on the social factors under examination. Linear regression can also predict that association between one outcome variable and multiple independent variables (Field, 2013).

All the factors highlighted by the univariate analysis for each response were simplified, using backward elimination, which removed any terms where the bivariate effect could be explained due to their relationship to another factor. The remaining terms could, therefore, be said to have greater effects on the response than the other factors.

A paired t-test was chosen for its ability to compare two means, representing two different but related conditions, derived from a within-subjects test group.

The purpose of the paired t-test was to find out whether there was any statistically significant evidence that the differences in means between paired observations on a particular outcome was different from zero (Hsu and Lachenbruch, 2008).

7.8.4 Quantitative data analysis

A coding by themes was used as a method to analysis the qualitative data. By combining them under specific categories and then highlight the similar passages of text with a code label. This will make it easier to be retrieved at a later stage for further comparison and analysis.

7.9 Results

7.9.1 Descriptive analysis

The same group of users were used to evaluate the three objectives of this study. Table (7.2) presents numbers and percentages of participant responses. The ages of the participants ranged from 18 to 44 (28.05 ± 1.22) were about 37% percent of them were aged between 21 and 25, and 28% aged between 31 and 35.

Table 7.2. Age group of the users within the study

Characteristic	Participants	Percentage %
Age (years)		
18-20	3	3.9
21-25	28	36.8
26 to 30	21	27.6
31 to 35	13	17.1
36 to 40	8	10.5
41 to 45	3	3.9
Gender		
Male	54	71%
Female	22	29%
Total	**76**	

In regard to frequency of making or receiving video calls using services, (42%) stated that they "*sometimes*" made video calls, while (36%) did this "*about half of the time*" and 16% "*always*" used video calls. So, most of the users had video call experience. On the other hand, 6% of the participants reported that they "*never*" used video call applications (Figure 7.2).

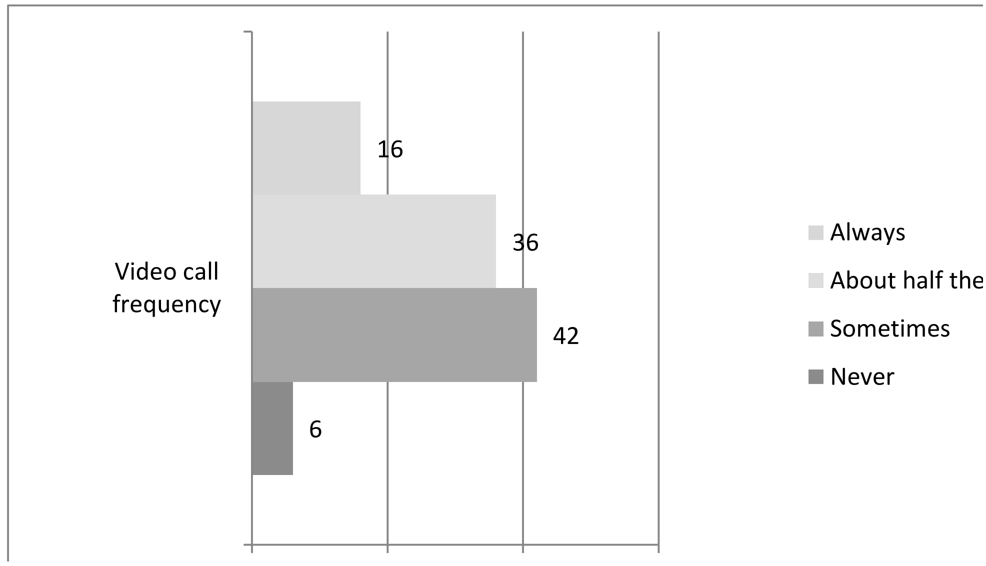


Figure 7.2. Users' frequency of making video calls

A variety of devices were utilised by the users to make video calls. Apart from the conventional methods (desktop computer), 38% of the users used smartphone for video calls, 36% of them used laptops and only 8% used tablets (Figure 7.3).

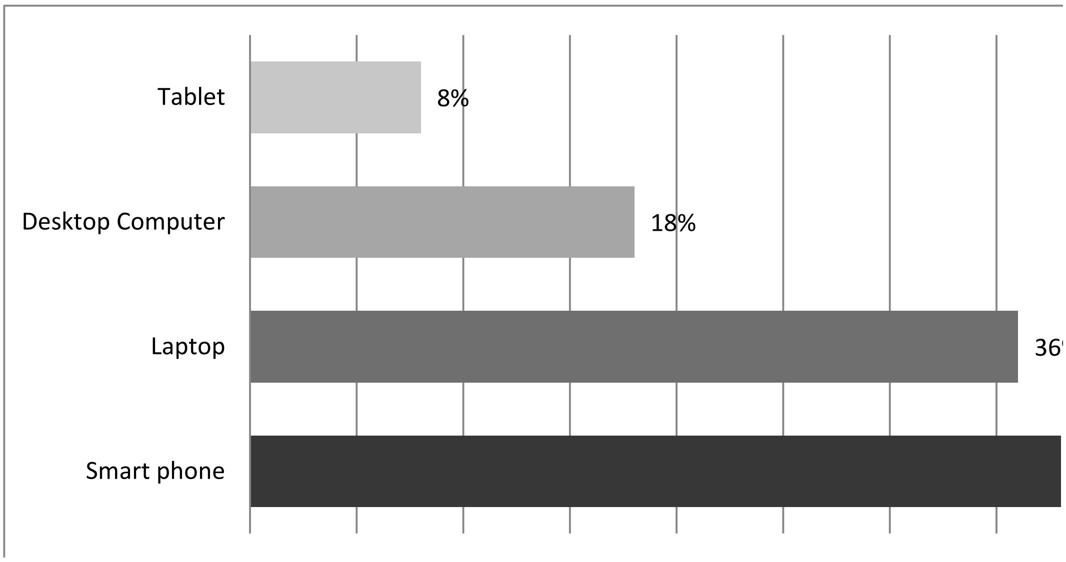


Figure 7.3. The percentage of the devices adopted for video call

It can be seen from the figures above (7.3) that, the subjects were familiar with using smartphones for video calls which count as an important to in the recruit to avoid the novelty effect.

7.10 Comparative study: MobiBot Vocal-triggered automatic vs. MobiBot Mimicking movements

A comparison of the different methods of interaction on video calls using the MobiBot TP system was carried out in order to evaluate their effect. The comparison was between mimicking movement by an operator and automated movement triggered by the user.

Users were given the same tasks, but with different choices of topic for each condition (Vocal-triggered automatic and mimicking movements). The task allowed the user to interact with the researcher using the MobiBot, which was considered adequate to enable the user to make decisions about each interaction method. Figure (7.4) shows the user interacting with the MobiBot TP system in the Mimicking movement and with Vocal-triggered automatic movement conditions.

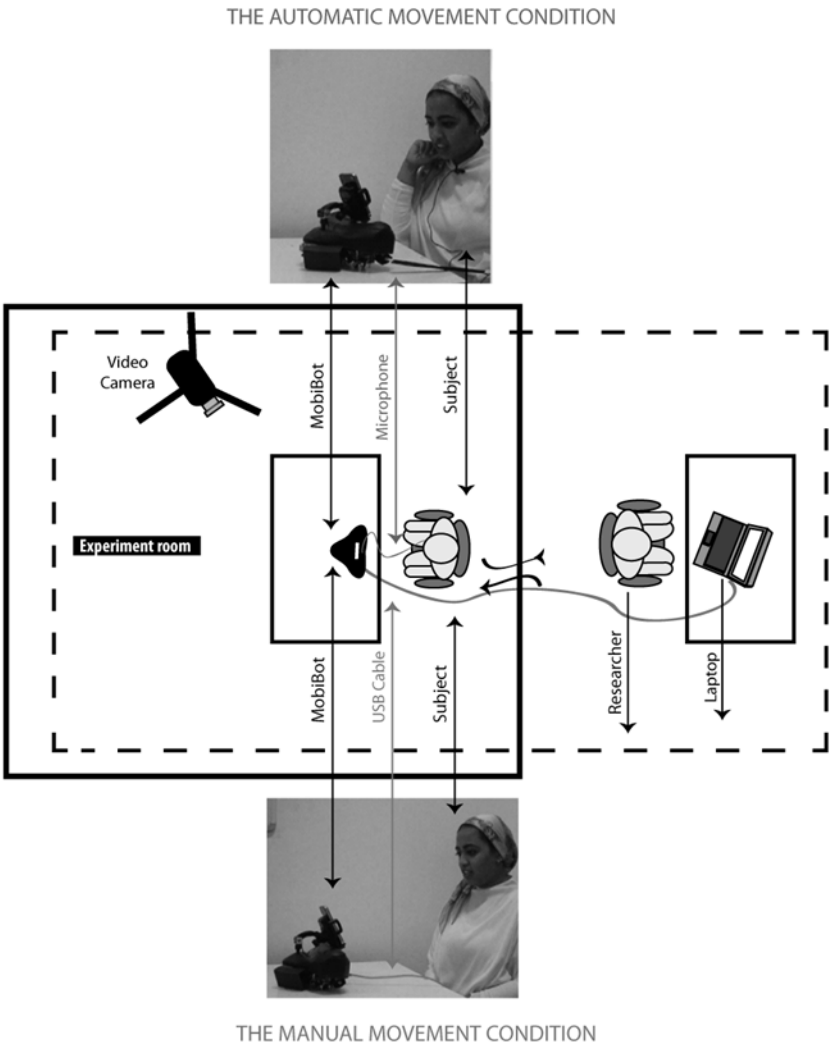


Figure 7.4. Participant interacting with the system in the mimicking and the vocal-triggered automatic movement conditions

7.11 Evaluate the Vocal-triggered automatic movement for MobiBot

To assess the further evaluation of the future usage of the Vocal-triggered automatic movement MobiBot system a comparison using their answers toward their preference to use the system for video call in future against some questions from the UTAUT model. Thirteen questions were used in this comparison; 2 to measure Perceived Sociability, 5 to measure Social Presence, 4 to measure Enjoyment and 2 to measure attitude toward the system.

7.11.1 Questionnaire design and evaluation

- ***Questionnaire design***

After taking part in the trials, users were asked to provide feedback using a digital questionnaire (Chapter 6) which included an evaluation of 7 measurements. The measurements were Embodiment, Perceived Sociability, Social Presence, Enjoyment, Temporal Dissociation, Focus Immersion and Attitude towards the system. For comparison purposes, the same questionnaire was used for each interaction method.

- ***Questionnaire evaluation: validity and reliability***

The validity of the questionnaire was tested. The internal consistency of all the items showed good to excellent reliability (over 0.7 αh values), which indicates a good level of internal consistency. However, the temporal dissociation item showed <0.6 in both movement methods (Table 7.3). The temporal dissociation factor was not included in the analysis as the Cronbach's alpha was less than 0.6 for both conditions, and deleting any of the temporal dissociation items did not cause any changes to the value of Cronbach's alpha.

Table 7.3. Cronbach's Alpha value for both interaction methods

Movement method	Vocal-triggered automatic Movement (Cronbach's Alpha)	Mimicking Movement (Cronbach's Alpha)
Embodiment	.786	.810
Perceived Sociability	.791	.839
Social Presence	.861	.911
Enjoyment	.725	.741
Temporal Dissociation	**.580**	**.626**
Focus Immersion	.804	.750
Attitude towards the system	.843	.826

7.12 Results and analysis

7.12.1 Normality checking

A visual inspection of the vocal-triggered automatic movement (Figure, 7.5) and mimicked movement (Figure, .6) histograms showed that the data for embodiment, social presence, perceived sociability, enjoyment, focus immersion, temporal dissociation and attitude towards the technology were normally distributed.

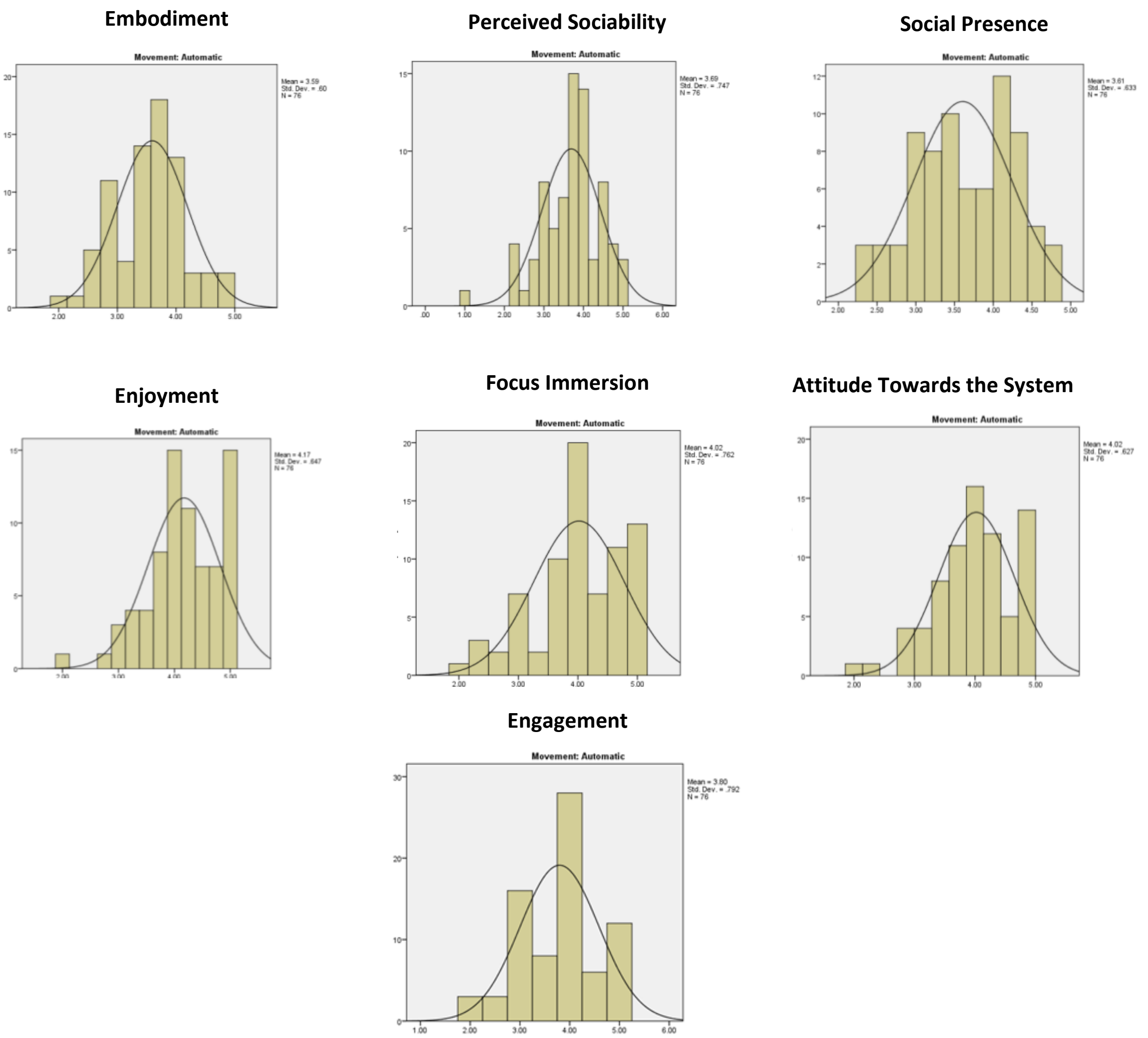


Figure 7.5. Normality histogram distributions for Vocal-triggered automatic movement factors

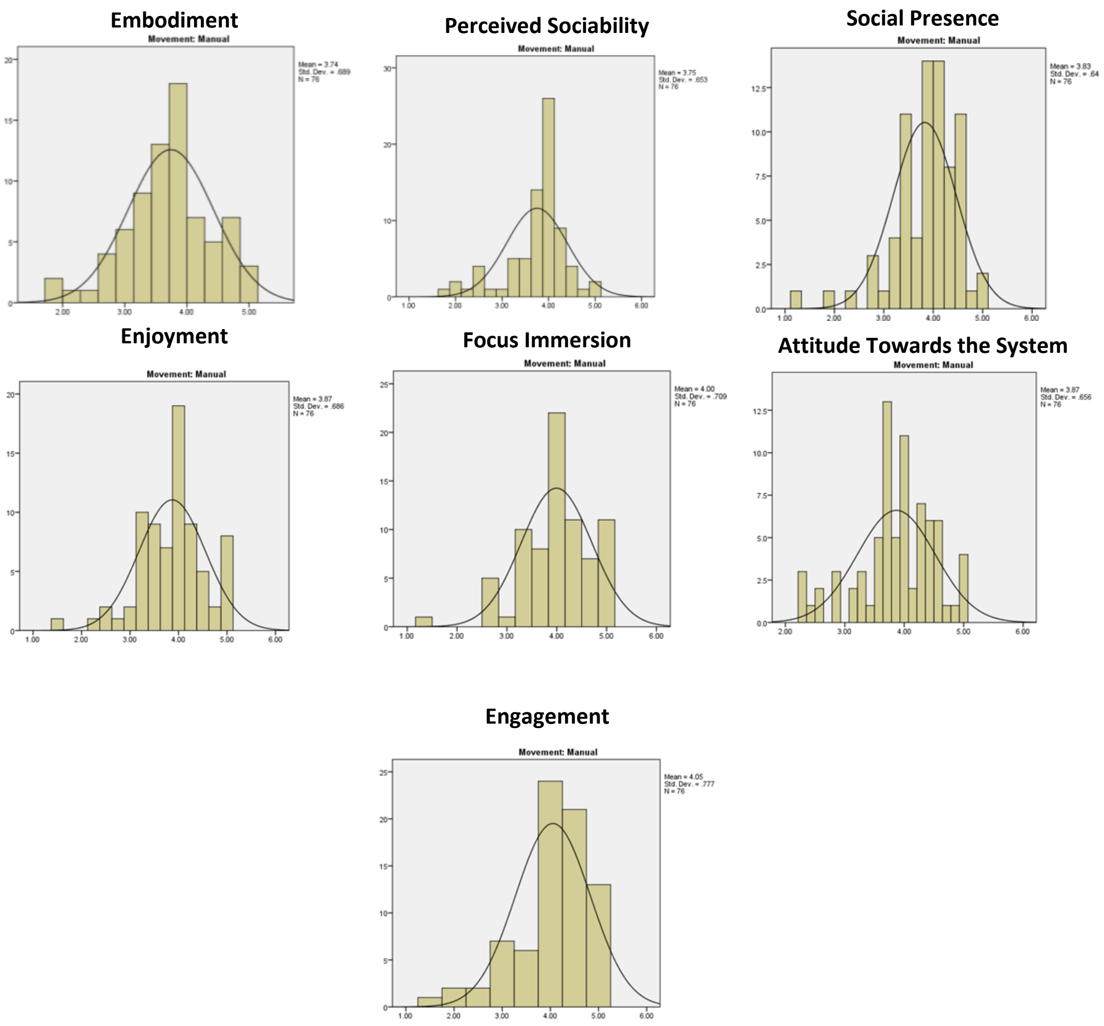


Figure 7.6. Normality histogram distributions for mimicking movement factors

7.12.2 Comparative analysis

The mean scores were measured for all factors in both interaction methods (vocal-triggered automatic and mimicking movement) as represented in (Table 7.4). A paired-sample t-test was conducted to assess the impact of the type of movement on the participants' attitude. The result showed a very small difference between the means of the averages (0.2) in favour of the mimicked movement interaction method. As shown, overall there was no significant difference between the users' preferences for either interaction method.

However, users seemed to favour MobiBot's mimicked movement, as it provided more social presence, with p-value = 0.033, and more temporal dissociation, with p-value = 0.046 with (3.83 ± 0.64) and (4.05+.78), *t*(150) = -2.153 and -2.015. However, they may have considered the vocal-triggered automatic movement to be more enjoyable, as it was statistically significant with p-value = 0.005 and (4.17 ± 0.65), *t*(150) = 2.828.

Table 7.4. Comparing the means of the factors in both conditions

Factor	Movement	Mean ± SD	Sig. (2-tailed)
EMB	Vocal-triggered automatic Movement	3.6± .6	.164
	Mimicking movement	3.7±.69	
PS	Vocal-triggered automatic Movement	3.7± .75	.604
	Mimicking movement	3.7±.65	
SP	Vocal-triggered automatic Movement	3.6± .63	**.033**
	Mimicking movement	3.8±.64	
ENJ	**Vocal-triggered automatic Movement**	4.2± .65	**.005**
	Mimicking movement	3.9±.69	
ENG	Vocal-triggered automatic Movement	3.8± .79	**.046**
	Mimicking movement Movement	4.0±.78	
INV	Vocal-triggered automatic	4.0± .76	.854
	Mimicking movement	4.0±.79	
ATT	Vocal-triggered automatic Movement	4.0± .63	.141
	Mimicking movement	3.9±.66	

7.13 Evaluation of the vocal-triggered automatic movement for MobiBot

This study adopted within-subject design. The same data from the previous study were used to assess the further evaluation of the future usage of the vocal-triggered automatic movement MobiBot system, by comparing the data from the extended UTAUT model against users' answers about their preference to use the system for video call in future.

7.13.1 Questionnaire design and evaluation

- ***Questionnaire design***

The same questionnaire as in the previous study was used to measure the users' perception of the use of MobiBot. Perceived sociability, social presence, enjoyment and attitude towards technology were tested in this section.

The responses to the questions from the previous interaction method assessment were included in the analysis for this questionnaire (Table 7.5). The decision to include some of the items was based on the construct related to the specific objective, which was a total of 13 items for vocal-triggered automatic movement as an interaction method: Attitude towards technology (ATT) contributed 2 items, Perceived Sociability (PS) 2 items, Enjoyment (ENJ) 4 items and Social Presence (SP) 5 items.

The user's answers to the open question about their preference to use the vocal-triggered automatic movement system in the future were also included.

Table 7.5. Factors statements

Factor	Item
Perceived Sociability	I find the System pleasant to interact with
	I consider the System a pleasant conversational partner
Social Presence	My experience of interacting with my partner using the System is consistent with a real world experience
	The forward movement of the System makes me feel as if I am having a face to face with my partner interaction
	The nod movement of the System makes me feel as if I am having a face to face interaction with my partner
	The forward tilt of the System makes me feel as if I am having a face to face interaction with my partner
	The backward tilt of the System makes me feel as if I am having a face to face interaction with my partner
Enjoyment	Interacting with my partner using the System is enjoyable
	Interacting with my partner using the System is not boring
	Interacting with my partner using the System is interesting
	Interacting with my partner using the System is entertaining
Attitude towards technology	I would like to experience this System again
	I am satisfied with my experience of the System

- ***Questionnaire evaluation: validity and reliability***

The reliability and validity procedures that were followed were similar to those already presented in Section 1.9.1.

Measurement of the reliability of the questionnaire involved assessing value Cronbach's Alpha.

As shown in the table, all of the constructs revealed a good to excellent reliability when tested for reliability. The highest p-value (0.8), demonstrated excellent reliability was for attitude towards technology, and the lowest p-value (0.7), demonstrating good consistency, was for perceived sociability (Table 7.6).

Table 7.6. Cronbach's Alpha value for the factors

Construct	No. of items	Cronbach's Alpha
Perceived Sociability (PS)	2	.7
Social Presence (SP)	5	.8
Enjoyment (ENJ)	4	.7
Attitude towards technology (ATT)	2	.8
Total	**14**	

7.14 Results and analysis

7.14.1 Normality checking

A visual test for normality based on histograms of the data for vocal-triggered automatic movement demonstrated normally distributed data for perceived sociability, social presence, enjoyment and attitude towards the technology (Figure 7.7).

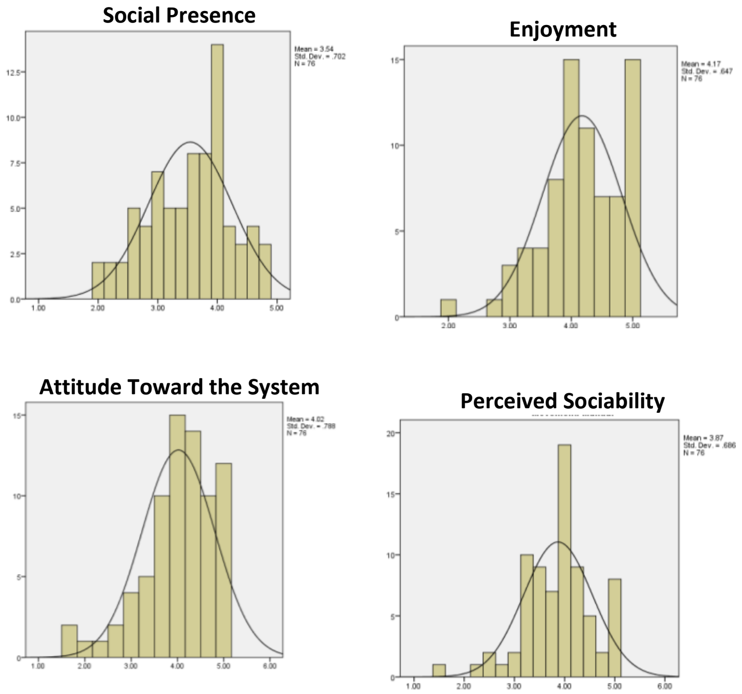


Figure 7.7. Normality histogram distributions

7.14.2 Comparative analysis

A Pearson's *r* data revealed a positive correlation *r* = ATT (.507), ENJ (.485), SP (.373) and PS (.440) (Table 7.7). Users who preferred to use the system in the future reported a higher level of perceived sociability, social presence, enjoyment and attitude toward the technology.

Table 7.7. Pearson's correlation for the factors

	Item	ATT	ENJ	SP	PS
Prefer to do video calling using the System	Pearson Correlation	.507**	.485**	.373**	.440**
	Sig. (2-tailed)	.000	.000	.001	.000

A linear regression analysis was calculated to predict the participants' preference to use the system in the future, based on their perceived sociability, social presence, enjoyment and attitude towards the technology.

A significant regression equation was found for perceived sociability and enjoyment (F (4, 147) = 14.915, *p* = .004 and .014). Participant predicted future usage is equal to .332 (perceived sociability) and .234 (enjoyment). Participant mean future usage increased to .332 for perceived sociability and .234 for enjoyment. The overall model fit was R square =.289 (Table 7.8).

Table 7.8. Regression analysis for the factors

Factor	β	B	R^2	p-value
ATT	-.028	-.020	.289	.826
ENJ	.362	.234		.014
SP	.076	.052		.574
PS	.422	.332		.004

7.15 MobiBot system evaluation: Unified Theory of Acceptance and Usage of Technology Model

The final study objective was to evaluate the users' attitudes towards the vocal-triggered automatic movement, since it was the proposed system based on the Unified Theory of Acceptance and Usage of Technology Model. Based on this objective, this study examined two sets of data. The first set of data examined the impact of the constructs on the cognitive absorption of using MobiBot telepresence robot for socially interactive video calls. The other set of data examined the impact of the cognitive absorption value and the other constructs on the attitudes towards MobiBot telepresence robot for socially interactive video calls.

7.15.1 Questionnaire design and evaluation

- ***Questionnaire design***

The question design followed the same procedure mentioned in Section (4.2) but only included vocal-triggered automatic movement.

- ***Questionnaire evaluation: validity and reliability***

The validity and reliability of the vocal-triggered automatic questionnaire mentioned in Section 7.9.1

7.16 Results and analysis

7.16.1 Normality checking

A test for data normality was conducted, as the validity of some parametric tests relies on the assumption of normality. Visual inspection of the histogram showed that the data for vocal-triggered automatic movement were normally distributed for embodiment, perceived sociability, social presence, enjoyment, temporal dissociation, focus immersion and attitude towards the technology.

7.16.2 Hypothesis testing

To establish how the scores on each construct were interrelated, we performed a correlation analysis, using the scores on UTAUT-derived constructs (Social Presence SP, Perceived Sociability PS, Enjoyment ENJ and Attitude Towards the system ATT), the construct of Cognitive Absorption; (Temporal Dissociation TD and Focus Immersion FI), and System Embodiment (Physical Appearance PA and Social Abilities SA) . As Table 7.9 shows, the score on the (non-UTAUT) construct of Cognitive Absorption (CA) and the System Embodiment (SE) did show a positive correlation with every item of the UTAUT constructs in vocal-triggered automatic movement.

Table 7.9. Model testing: Correlation analysis

Movement Type	Factor	SA	PS	SP	ENJ	INV	ATT	PA
Vocal-triggered automatic	SA	1						
	PS	.547**	1					
	SP	.554**	.700**	1				
	ENJ	**.360****	.679**	.544**	1			
	FI	**.278***	**.343****	**.479****	**.406****	1		
	ATT	.509**	.745**	.697**	.712**	.501**	1	
	PA	**.375****	.563**	.502**	.180	**.393****	.653**	1

Of course, correlations only show that variables are related, not that there is a determining influence in a particular direction. They also do not state which influences are dominating determinants on particular constructs. Therefore, a regression analysis would be appropriate.

In the first section of Table (7.10), it can be seen that social presence was used as a dependent variable, and the perceived sociability and social abilities were used as independent variables (predictors) with the aim of testing hypotheses H4, and H9. The results show a significant regression equation for perceived sociability and social abilities ($F(3, 72) = 28.598$, $p = .000$ perceived sociability and .018 social abilities). Participants' predicted social presence is equal to .678 + .244 (social abilities). Participants' mean social presence is increased by .244 of social abilities. Social presence is equal to .678 + .421 (perceived sociability). Participants' mean social presence is increased by .421 of perceived sociability. In general, perceived sociability can predict almost 50% of the social presence and social abilities can predict 23% of on the social presence. The overall model fit was R square = 0. 544.

In the second section of the Table (7.10), it can be seen that the attitudes towards the system were used as a dependent variable and perceived sociability as an independent variable (predictor), in addition to enjoyment and physical appearance, with the aim of testing hypotheses H1, H3 and H7. The results show a significant regression equation for perceived sociability, enjoyment and physical appearance (F (7, 68) = 26.697, p = .044 perceived sociability, .003 enjoyment and .005 physical appearance). Participants' predicted attitude towards the system is equal to .077 + .190 (perceived sociability). Participant's mean attitude towards the system is increased by .190 of perceived sociability. Perceived sociability can predict almost 22% of the attitude towards the system. In addition, participants' predicted attitude towards the system is equal to .077 + .269 (enjoyment). Participants' mean attitude towards the system is increased by .269 of enjoyment. Enjoyment can predict almost 28% of the attitude towards the system. For physical appearance, participants' predicted attitude towards the system is equal to .077 + .226 (physical appearance). Participants' mean attitude towards the system is increased to .226 of physical appearance. Physical appearance can predict almost 23% of the attitude towards the system. The overall model fit was R square = 0. 544.

In the last section of the table (7.10) testing (H2), by using enjoyment as a dependent variable and perceived sociability as an independent variable (predictor), shows that a significant regression equation was found for perceived sociability (F (2, 73) = 32.372, p = .000). Participants' predicted enjoyment is equal to 1.809 + .505 (perceived sociability). Participants' mean enjoyment is increased by .505 of perceived sociability. The level of the perceived sociability can predict almost 58% of the enjoyment. The overall model fit was R square =.470

Testing H5 and H6, by using temporal dissociation and focus immersion as dependent variables and social presence as an independent variable (predictor), shows that a significant regression equation was found for social presence (F (4, 71) = 8.792, p = .000) with temporal dissociation and (F (4, 71) = 6.994, p = .000) with focus immersion. Participants' predicted temporal dissociation is equal to .992 + .777

(social presence). Participants' mean social presence is increased by .777 of temporal dissociation.

In term of the focus immersion, participants' predicted focus immersion is equal to .968 + .373 (social presence). Participants' mean social presence is increased by .373 of focus immersion. In general, the level of temporal dissociation can predict 62% of the social presence and the level of focus immersion can predict 37% of the social presence. The overall model fit for temporal dissociation was R square = .331 and R square = .283 for focus immersion.

In the last row testing (H8), by using perceived sociability as a dependent variable and social abilities as an independent variable (predictor), shows that a significant regression equation was found for perceived sociability (F (1, 74) = 31.616, p = .000). Participants' predicted perceived sociability is equal to 1.242 + .681 (physical appearance). Participants' mean perceived sociability is increased by .681 of physical appearance. The level of the perceived sociability can predict almost 55% of the enjoyment. The overall model fit was R square =.299.

Table 7.10. Model testing: regression analysis

Regression path		B	β	R^2	p-value	Result
Social Presence predictor						
H4	Perceived Sociability → Social Presence	.421	.497	.544	.000	Supported
H9	Social Abilities → Social Presence	.244	.232		.018	Supported
Attitudes towards the system predictors						
H1	Perceived Sociability → Attitudes	.190	.227	.733	.044	Supported
H3	Enjoyment → Attitudes	.269	.276		.003	supported
H7	Physical Appearance → Attitudes	.226	.231		.005	Supported
Other construct predictors						
H2	Perceived Sociability → Enjoyment	.505	.584	.470	.000	Supported
H6	Social Presence → Temporal Dissociation	.777	.620	.331	.000	Supported
H5	Social Presence → Focus Immersion	.450	.373	.283	.000	Supported
H8	Social Abilities → Perceived Sociability	.681	.547	.299	.000	Supported

The study conducted linear regression analysis to test the research hypotheses. The Figure (7.8) shows the final model proposed with structural path coefficients (R). All the hypotheses of the conceptual model were statistically supported ($p < 0.01$).

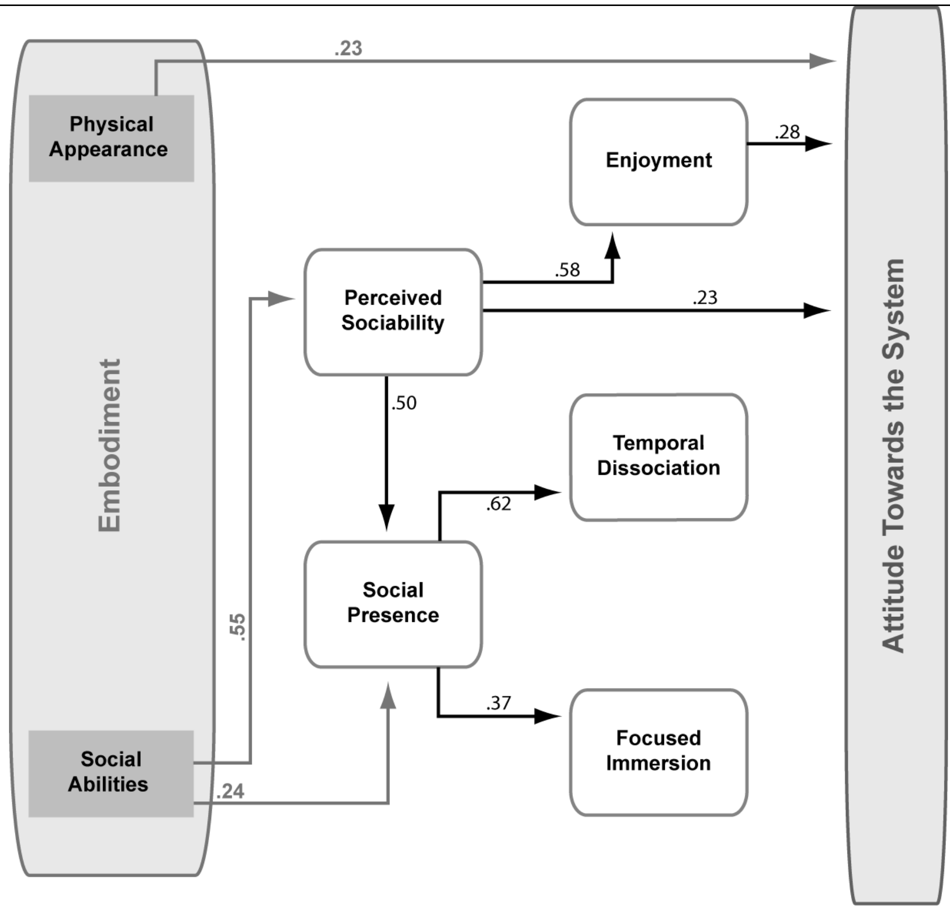


Figure 7.8. The result of the extended UTAUT model shows the path and R coefficient value

7.17 Discussion

7.17.1 Comparative study: MobiBot Vocal-triggered automatic vs. MobiBot Mimicking movements

The results of the interaction method comparative analysis indicated that the mimicked movement interaction method for MobiBot system for video calls was preferred by the users over the vocal-triggered automatic interaction movement method, with respect to the temporal dissociation and social presence. Temporal dissociation, as stated by Sidner *et al* (2005) and Sidner,Lee and Kidd (2005) is supported by three kinds of social behaviours, and most of the users showed a high level of temporal dissociation. This was also obvious in the analysis of the video recording of the user trials by high level of immediacy (physical proximity), expressiveness (nods), altercentrism (gaze), interaction management (word counts), and positive affect (smiling and laughing).

In total, these are the six scales that are necessary for measurement of temporal dissociation as described by Laura K Guerrero (2005) (Chapter 4). In addition, users reported their preference for the mimicking movement, which was higher than the vocal-triggered automatic movement with 64%. This also reflected their comments on the system; "*The experience was more understandable in the sense of conveying emotions and enhancing the conversation*". Temporal dissociation is understood to have a direct effect on social presence, which is consistent with the proposition and the expectations of this study, based primarily on study observations (Chapter 4). The exhibition of naturalistic behaviour and appropriate emotions by a robot can appear to humans as convincing as those of human expectations (Bartneck, 2000), and a significant result of social presence resulted from this in association with mimicking movement. This result also can be inferred from participants' comments "*More normal and attuned to regular human conversation, the experience was more understandable in the sense of conveying emotions*". This comment can be related to the type of movement, as users made positive statements about the movements in the mimicking condition compare to the vocal-triggered automatic movements. They were clearly in favour of the system which caused fewer movements, which was the manually controlled interaction method; "*The mimicking movement experience seemed more purposeful and the movements seemed to obviously correspond with distinct aspects of the conversation*". In addition, the stability and the smoothness of the movement made the conversation seem more peaceful and natural.

In respect to the vocal-triggered automatic interaction movement method for MobiBot system for video calls, this was preferred over the manual mimicked movement interaction method in respect of enjoyment. This was also obvious in the analysis of the video recordings of the user trials, as users were surprised by the ability of the system to move sideways and forwards in order to locate the user's position, and to face them directly, which was missing in the manually controlled interaction movement method.

The comments; "*The device being able to locate where you are and face you, It just felt very interesting and cool*" and "*liked the vocal-triggered automatic experience more than the manual one due to the fact that the robot located me at first and positioned itself perfectly to centre my face in the middle. I found that really cool and it really impressed and put me in a better mood to talk and have a conversation*". However, temporal dissociation and social presence was not significant in this interaction method, due to the sharp and rapid nod movement for the normal tone of long speech, which occurred every five words in the conversation in this interaction method. People reported that this type of nod movement was a bit too often, quick and distracting; "*There was a lot of jittery movement which didn't necessarily correspond to what either person was saying*". This, in term, made it difficult to focus on the conversation; "*It was a little bit distracting on my focus in the conversation. When I speak with this device, I feel like I speak with ignorant people*"

In terms of the embodiment, the colour of the device and the shape were not favoured by the users, as the reported that a more human look would help in perceiving the system positively. In perceived sociability and focus immersion, users expressed no preference for one interaction method over another. It is very likely that was due to motor noise technical issues during production of movement in both conditions, which was reported in the evaluation section. This in total affected responses and attitudes towards the system.

Users reported an interesting point about movement in general, and how it followed the conversation, especially if there was a lag or connection problem, as; "*Encouraging me to talk even with video lag and connection problems*". In addition, they explained that this type of movement worked as a physical emoji, and helped people who have difficulty in expressing themselves "*I liked the whole system. Compared to Skype which can sometimes get awkward as I'm not a talkative person, I did not get a sense of awkwardness using this system*".

7.17.2 The extended UTAUT Model hypotheses testing

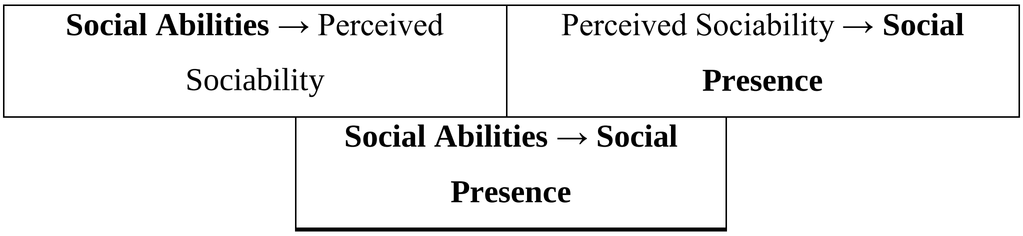


Figure 7.9. The relationship between Social abilities, perceived sociability and social presence

The findings of this study were that social abilities affect the perceived sociability of the robot. This finding was supported by a study carried out by Looije,Neerincx and Cnossen (2010) who used iCat, and found that the robot’s perceived social abilities were appreciated by the users in the same way as they would be in conversation with a human partner. This would lead us to the conclusion that perceived sociability has the strongest influence on the social presence. According to Lee *et al.* (2006), social responses to robots are mediated by feelings of social presence during interactions with the robot, and our result is in line with this study. Technology in the form of an embodied character, that is interacting in a social manner and using non-verbal human behavious and natural language can often be perceived by users during an interaction as a natural social partner. This means that a high feeling of social presence can result in users perceiving the robot as a social entity. Tapus and Mataric (2008) found that in terms of movement type, users’ perceived sociability rating was significantly higher for an extrovert robot than for an introvert one, which again highlights the importance of the social movements in promoting interactivity and reciprocity among communicators which arise as a result of enhanced perceived social presence. We may, therefore, conclude that the sense of presence that people feel when faced with a robot can be manipulated by making changes to the robot’s social abilities (and indeed this does make people change their perception of perceived sociability accordingly).

Social Presence → Focus Immersion

Social Presence → Temporal Dissociation

Figure 7.10. The relationship between Social presence, focus immersion and temporal dissociation

Social presence, of all the factors, has the strongest influence and most significant effect on temporal dissociation (B=.78, P<.000). The more natural real time and real world users perceive an interaction to be, the more the features of the system engage and absorb the users. This also be observed in relation to focus immersion - providing users with social movement in support of the video call was seen to increase the social presence (as explained earlier) which resulted in an improvement to their involvement.

Perceived Sociability → Enjoyment

Enjoyment → **Attitudes**

Perceived Sociability → Attitudes

Figure 7.11. The relationship between perceived sociability, enjoyment and attitudes

There is a meaningful relationship between perceived sociability and perceived enjoyment, and it was observed that a higher perceived sociability will lead to a higher score on enjoyment. This finding was supported in the study by Timothy Bickmore and Schulman (2006) who used a relational agent which developed social bonding. It was found that social abilities (which earlier, we described as influencing perceived sociability) lead to enjoyment being experienced by the users. In other words, when people experience perceived sociability, that is, warm and personal interactions with others, they obtain fun and pleasure through it. Additionally, this sense of enjoyment has a positive impact on attitude towards the system, as found here. In other words: the more the interaction with a robot feels like fun, the more it tends to have a positive impact on attitude towards the system.

Nah,Eschenbrenner and DeWester (2011) discovered that enjoyment was a key antecedent of user attitude, and it was positively affected in such systems (telepresence).

Perceived sociability has been shown to have a significant influence on attitude towards the system. In other words, the perceptions of a robot as a social entity will influence the way that the entity is observed and accepted, which view is supported by different studies (Timothy W Bickmore,Caruso and Clough-Gorr, 2005, Lee and Nass, 2003, Heerink *et al.*, 2008a, Mitsunaga *et al.*, 2008). People are naturally experts in human-human social interaction, and can apply their existing knowledge of human social models to their interaction with robots (Cynthia Breazeal, 2003, Kidd and Breazeal, 2004, Fong,Nourbakhsh and Dautenhahn, 2003). Thus, is a robot has sufficient sociability, interacting with it could seem easier to a person (Heerink *et al.*, 2009), either because it allows them to use existing knowledge (e.g., social models of human-human interaction), or it is sufficiently compelling that they do not perceive the interaction as needing effort (e.g., toy robots) (Fong,Nourbakhsh and Dautenhahn, 2003). Consequently, it could be concluded that the social movements in a TP system provide a very high level of social interactivity with users, which has a direct impact on the creation of enjoyable and pleasing experiences, and as a consequence, promotes positive attitudes towards the technology.

Physical Appearance → Attitudes

Figure 7.12. The relationship between physical appearance and attitudes

This study also found a strong relationship between physical appearance and attitudes with ($B=.68$, $P<.000$). This result is in accordance with the Bartneck (2003) study in social robotics. This study found that participants would put more effort into a task with a robotic character than with a screen character.

To this end, the socially supportive movement method for MobiBot system in the context of video calls is a useful way to improve how a user perceives it as a natural social partner during a conversation that is highly interactive and provides more real-world and more natural real-time interaction. In addition, movement has a direct influence on creating an enjoyable and pleasing experience and as a consequence, promotes positive attitudes towards the technology.

7.17.3 Evaluate the Vocal-triggered automatic movement for MobiBot

A further analysis through an evaluation of users' preferences' toward using the system in the future revealed that users with a noticeable increase in the level of perceived sociability and enjoyment were the main predictor for future preferences of using the system. This finding was supported by Heerink *et al.* (2008a) and Leite (2008), on the design of interactive robots. This study confirms enjoyment as having an indirect effect on intention to use, and that users might actually feel the same enjoyment as if they were playing a game or having a pleasant conversation with a person, and this might encourage them to use the technology.

7.18 Summary

The purpose of this chapter was to examine the role of an interaction method in influencing both the effectiveness of the method and the users' acceptance and preferences for the conventional TP systems against MobiBot TP systems from the users' perspective and point of view. Three studies were conducted in order to evaluate MobiBot TP for video calls system. The studies comprised the evaluation of the interaction method using vocal-triggered automatic against mimicked movement, the attitude towards MobiBot TP for video calls system was evaluated based on an extended UTAUT model and an evaluation for MobiBot TP for video calls system using attitude towards the system, enjoyment, social presence and perceived sociability against preference to use it in the future.

The results of the interaction method comparative analysis revealed that in this particular setting, the users preferred manual mimicking to vocal-triggered automatic as temporal dissociation and social presence interaction methods. Further, the users preferred vocal-triggered automatic to mimicking as an enjoyable interaction method.

Whereas, in term of the perceived sociability and focus immersion no changes detected for one interaction over another.

In respect to the acceptance analysis for the vocal-triggered movement interaction based on an extended UTAUT model, it can be concluded that perceived sociability, social presence, social abilities, enjoyment, temporal dissociation and focus immersion values have a significant positive effect on user attitudes towards the technology.

Finally, the evaluation for MobiBot TP for video calls system using attitude towards the system, enjoyment, social presence and perceived sociability against preference to use it in the future was shown that perceived sociability and enjoyment are significant, which in turn revealed a positive preference to use the system in the future.

The quantitative and qualitative finding from the study demonstrated that incorporating movement, in general, was an idea that showed promise. The major conclusion of this study was that the socially active systems can enhance user performance showing that the movement actually aided in the user's sense of presence experience, cognitive experience which lead to improvement in their attitudes towards the system with also support from the system embodiment. Users in our study direct their attention to the robot more often in interactions and they find interactions more appropriate when movements are present. That's because these movements play a crucial role in human interaction, such resources are also likely to play an important role in human-robot interaction.

Therefore, we believe in implement such nonverbal behaviour resources such as face and head movement in a robotic system, we cannot simply program the robot to move at random but rather we need to consider the ways these actions may be timed to specific points in the talk. However, it needed to be carefully thought out, and any future designs thoroughly user tested and assessed. Users provided improvement suggestions with regards to the types of movement used in the system, and suggested a mix of both vocal-triggered automatic and manually mimicked movements.

Chapter 8 GUIDELINES

This chapter highlights the main guidelines resulted from research findings

8.1 Introduction

Results from the thesis provide guidelines for 'personifying' telepresence conversations and development of such systems and discuss the link between the results and the guidelines.

8.2 Guidelines

- **Simplicity:** The finding from the evaluation study indicated that users were satisfied with the shape. Therefore, simple design should be considered when designing a product which only needs to provide enough information, and to minimise ambiguity at the same time. In our design, we used a simple shape that represents the upper part of the body, which was supported with movements (Figure 8.1).

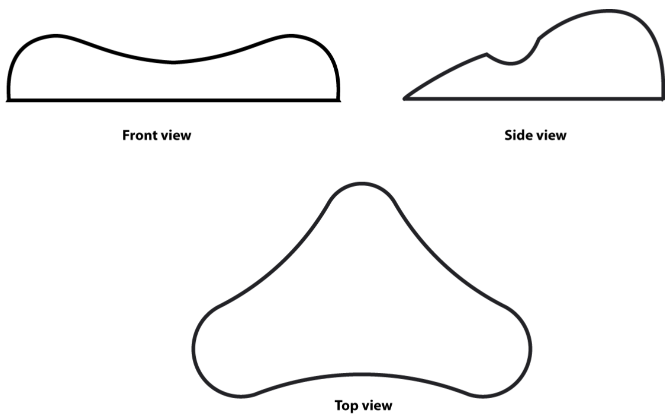


Figure 8.1. Physical design outlines

- **Familiarity:** In our scenarios, we implemented human-like behaviours and movements, which were chosen with reference to typical human-human interactions in a user's every day experience (Figure 8.2). These behaviours were highly dependent upon the robot's embodiment, because users tend to expect a robot to react and behave in a particular way, based on its physical appearance, for example; when using an upper-body type of shape. In addition, a robot should not behave in an intrusive manner as this will distract the user from communicating with the remote user, which could cause frustration resulting from repetitive and unwanted behaviours.

Findings from the evaluation study indicate that TP system provided a very high level of social interactivity with users as they can apply their existing knowledge of human social models to their interaction with robots.

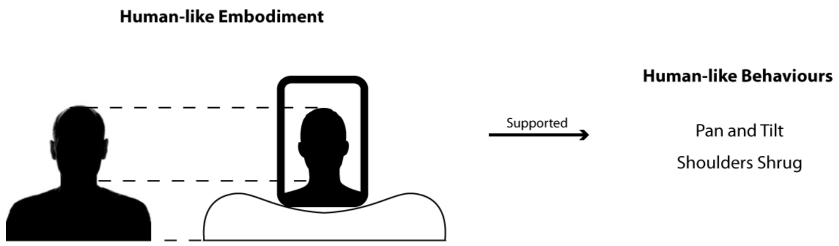


Figure 8.2. Human-like behaviours with reference to human body

- **Personalisation:** Our system was equipped with personalised behavioural planning, based on users' reactions. Every user interacts uniquely with any system, which means that systems need the ability to adapt their behaviours to suit those of each individual user. Human-Robot Interaction benefits from this type of personalisation because the robot will provide appropriate response behaviours, and be more successful in responding to the actions of the user. Users reported that movement in general followed their conversation, as our system translated the vocal stream into movement in order to achieve appropriate response behaviour.

- **Variety:** It is important that the robot is equipped with sufficient variety of actions on its behavioural list in order to avoid repetition, as too much repetition may make it tedious for the users. That was based on one of the evolution of the system part explained that the varieties of movement worked as a physical emoji, and helped people who have difficulty in expressing themselves. Our system was equipped with four main movement types, each one different from the rest in their speed (fast and slow) and their magnitude (small and large).

- **Modality:** The robot needs the ability to operate autonomously, so that it is possible to minimise the amount of superfluous interaction between the robot and the users.

Any autonomous mode of operation needs some basic guidelines incorporating to facilitate the interaction between the user and the remote user, and also to make the Human-Robot Interaction as rewarding as possible too. The autonomous mode should incorporate behaviours that are realistic and believable, to keep and hold the interest of the users during the communication session and Human-Robot Interaction. The autonomous mode should be supplemented by the addition of a user input mode, which should utilise familiar input devices to control the robot's movements. This mode and method of input may make the user feel more comfortable and enhance their experience in communicating with a distant user through the robot. In addition, a user input mode will give the user more control over the robot's behaviours, allowing them to personalise the interaction to their own and their distant partner's preferences, and also may maintain the user's level of interest for longer.

Our system provided users with two movement types, including a vocal trigger and a user input mode which allowed the user to imitate natural movements.

- **Capacity:** Energy efficiency is an important consideration, because social mediators tend to be small robots, with a limited battery life. It is important therefore, that when designing and implementing behaviours, special consideration is given to minimising energy usage. We used lithium ion batteries in our system as they are known for their high capacity.

- **Expressivity**: With the stated aim of making interaction with a robot more natural, we developed a novel interaction format to translate the vocal stream into movement capable of mimicking a user's body language, within a physical space, with nothing attached to the user. Since nothing is in contact with the subject, and there are no tracking markers attached to the user or the device, this method permits the user to freely communicate and expressively interact with the system without predefined space or any restrictions.

In our design, we used a cable microphone, which could easily be replaced by a wireless one (Figure 8.3).

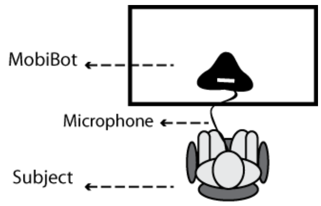


Figure 8.3. Tracking markers free system

Chapter 9 CONCLUSION

This chapter highlights the research contributions, notes the research limitations and outlines future work

9.1 Introduction

During this Ph.D. research, we designed, developed and evaluated platforms to emulate, and therefore enhance natural interaction in remote communication between distant participants. During the process, we identified the importance of face-to-face conversational signals, as they can be powerfully effective with minor actions such as smiles and head nods, causing a speaker to quickly produce more speech or topics of conversation. If these signals are absent, perhaps due to a connection problem, this can have a negative impact on the speaker, which may cause the speaker to repeat the speech more loudly, or even to cease speaking. In general, we found that different levels of media quality cause different physiological responses in users, and these differences can be detected through common physiological measurement techniques. Therefore, we have developed a unique way to deliver part of these actions, using currently available technologies. Section (9.2) highlights the thesis research questions, followed by an overview of the thesis progress in section (9.3). Research limitations and potential future work in section (9.4) and (9.5). Followed by a succinct overview of the thesis contributions in section (9.6) and A conclusion to this dissertation is drawn and finally in section (9.7).

9.2 Research questions

If face to face interaction is considered the gold standard for communication, video conferencing is the current gold standard for remote communication, as it takes place in real time, and incorporates two very important cues; tone of voice and facial expressions. A telepresence robot takes this communication one step further, by also incorporating movement. Therefore, nowadays human-robot interaction research involves empowering a robot with the social functionality to engage human participants. This requires a system to be built with a degree of capability for verbal and visual behaviour, such as head movement, body movement, and eye contact, as these have a direct effect of increasing in-person experiences. A large body of research shows that humans will treat computers as equal social partners if they behave in a human-like manner, thus facilitating natural interaction.

There is still a long way to go before we produce a robot that can interact in a fully human-like manner, but advances in control and sensor technologies now make it possible for researchers to carry out experiments with different configurations of hardware and software. Based on the above observation, this study aimed to answer the following research question:

Do users perceive an enhancement to the telepresence robot interface, following the addition of nonverbal behaviour by the robot?

In order to answer this question, the following two specific sub-questions are to be answered:

1st Sub RQ: How does the enhanced appearance, following the addition of nonverbal behaviour by the telepresence robot interface, influence the users' sense of presence?

2nd Sub RQ: How do the perceived embodiment, perceived sociability, social presence, enjoyment, Temporal Dissociation and Focus Immersion influence attitudes towards the enhanced appearance, following the addition of nonverbal behaviour by the telepresence robot interface?

The MobiBot telepresence robot for socially interactive video calls was developed and progressively evaluated from different perspectives in an attempt to answer the aforementioned questions.

9.3 Research development

The exploratory study addressed the first sub research question. As it used a small monitor, as the most common forms of mediated human-human communication are based on standard computer monitors which can relay visual as well as audio information to users. For this reason, we used computer hardware to connect and synchronise two distant users, and at the same time, to offer an interactive platform by replicating the head movements for the users, in order to enhance their current communication. Head movement was chosen in order to evaluate the participants' perceptions during a task, as well as their general perceptions of the proposed platform.

The aim of this study was to explore the potential for replicating the head movements that support engagement and feedback functions between people, which could be useful in the development of a model of human gestures to implement in a telepresence platform. The system improvements that were required were defined by incorporating the qualitative and quantitative results, based on the exploratory study outcomes. Iterative system design and evaluation were utilised in order to integrate users' feedback into the development procedure.

We revisited earlier negative experimental results, and we presented some design heuristics for video displays in TP systems. We concluded that for systems that preserve gaze, facial expression, gesture and upper-body cues, there is no evidence of a deficit in communication effectiveness compared to face-to-face meetings. In addition, we found that the communication between the platforms and the users should be simple, and only the really necessary information should be transmitted. In the light of these new findings, taking into consideration the actual technology capabilities we have in the real world, we changed the concept of the current research, by developing a platform which is able to deliver nonverbal signals for video-mediated conversations as representations of real-world interaction. We proposed a socially expressive system that allows users to relate to the nonverbal behaviours that spontaneously arise during speech, for a reliable synchronous interaction.

Using the proposed socially-supportive movement method for MobiBot Telepresence system for video, the MobiBot system was evaluated with representative users in order to address second sub research question. A comparison was conducted between mimicking movement triggered by user input, and automated movement triggered by the user vocal stream, in order to understand the potential advancement in using this method in controlling such systems. Further, the attitudes of the gesturally-supported video communication used for the MobiBot system was examined, based on extended UTAUT. Proving the affect of social abilities on the he perceived sociability of the robot which in turns affects the Social Presence as part of the research hypotheses.

In another words, technology that is interacting in a social manner using nonverbal behaviours can resulted in a high feeling of social presence which can result in users perceiving the robot as a social entity. In addition, a meaningful relationship between perceived sociability and enjoyment was found which also promotes a positive attitudes towards the technology. By perceiving the system as social entity, a fun and pleasure will be associcated with, which in turns will tends to have a positive impact on attitude towards the system. Furthermore a positve impact from temporal dissociation and Focus Immersion was found on Social Presence. This means the natural users perceive an interaction, the more engage and absorb the users.

In general, the feedback of user trials of the system obtained by questionnaire concluded that the socially-supportive movement method for MobiBot system in the context of video calls is a useful way to improve how a user perceives it as a natural social partner, during a conversation that is highly interactive, and provides more real-world and more natural real-time interaction. In addition, movement has a direct influence on creating an enjoyable and pleasing experience, and as a consequence promotes positive attitudes towards the technology.

9.4 Research limitations

This study as stated above has several limitations, which in turn could offer potential avenues for future study

- The recruitment of larger numbers of users was hindered by the time limitations of a Ph.D. study, which resulted in very little real user experience being gathered, processed and fed back into the design concepts and the nature of the physical platform.
- The limited budget shaped the electronic hardware used at the beginning of the project, as we could easily have adopted a more accurate solution for voice analysis.

- Limited resources could be seen as an advantage in our present study, where the immediacy dimension is not applicable because of the nature of the game technology used in our with video analysis, however, this type of method needs at least two people to measure the reliability of the findings, which we could not afford. As a result, we only covered the quantifying behaviours in measuring user's engagement.
- Although our measurement and analysis methods used in this study appear to be a practical way to acquire data on the magnitude and length of a variety of gestures, we also believe that available technology can be used to further measure other specifics, including synchronizing the head movements with eye gaze, using an eye gaze tracking system, and a webcam as a video recorder.

9.5 Future work

Stemming from the current research, potential avenues for future study include:

- The **system functionality** could be upgraded by use of the latest available technologies, and this could have a major influence on the usability of the system, and therefore the results obtained. For example, the development/learning time and cost to use such mobile android/IOS/windows developments in voice recognition applications were discounted due to cost and the time to customise such apps, coupled with the fact that the mobile OS apparent lacks the ability to run multiple apps at the same time, to save on development time. However, future work might include considering mobile voice recognition applications, as they are available and are becoming more reliable.
- **A larger sample** could be considered to get more accurate data from user trials. Factor analysis could be run to further assess the model and the relationship between the factors of interest, in addition to examining in-depth the influence of the dimensions of each construct upon the attitudes of individual.

- **Improving and refining** the output for the Vocal-triggered automatic movement. This might include users' trial suggestions with regards to the movement types used in the system, as to a mix of both Vocal-triggered automatic and mimicking movements. Fewer large movements, and more frequent small and steady movements, which are more sensitive and accurate, might enhance the usability of the system.
- The challenges for Intelligent Technology and Human-Robot Interaction researchers are numerous, and in our case, are mainly related to long-term acceptance evaluation. **A long-term** evaluation in a non-case in laboratory settings which might influence user's attitude towards the system by eliciting novelty effect arises with such technology.
- Different **type of task and different experiment set up** where we could explore more questions whether the movement will have the same effect when these changes made.

9.6 Summary of contributions

This research has focused on studying the human-to-human interaction, and replicating this in the context of human-robot interaction. It has furthermore developed a platform which is able to deliver nonverbal signals for video-mediated conversations as representations of real world gestures, aiming to simultaneously exhibit competent behavior, convey attention and intentionality, and handle social interaction with a particular interest in triggering awareness of the presence of a person. The major contributions made in this research work are summarised below:

- In this thesis, a **HCI interface design philosophy** was presented which mainly involves the analysis of HHI. **Understanding human physiology and human psychology** as the only solution to help in understanding the dynamics of it in order to increase the overall functionality of telepresence robotic system and mediated system in general. Also, being able to understand human interaction will help in improving the direct human contact which can be seen in our daily activities which will provide more effective mediated video display communication.

- Our research **investigated the perspectives of adding nonverbal behaviour** to the systems, such as the capability for head nodding and adjustable body orientation. As they convey an independent message and have been shown to clarify the meaning of speech behaviours, to convey a certain state, and in the inference of another's attitudes from such behaviours.
- This thesis can help **guide the development of social Telepresence robots** in embodiment design of telepresence which based on the finding from the literature review. The best way to categories telepresence robotic system from engineers and media developer's perspective and practical viewpoint is to organize them based on the human social norms. The **overview of methods to communicate non-verbal cues** aims to provide a framework for increasing social presence, recognizing that good social relationships breed good working relationships.
- Implementing head movements in TP system is valuable for HCI field, as it provides users with feedback about their communication. The trope of a "face-to-face interaction" may be to blame for the emphasis on head-only video in the design of video-mediated systems. The trope is quite misleading: humans rarely, if ever, have head-only encounters. As we found from the exploratory study that a real-time dynamic communication is needed in order to get a significant result. Thus, body language plays a major role with in-person encounters. This suggests that body **language cues, beyond facial cues, to be extremely important** if we are to approximate in-person experiences. Many of these nonverbal behaviours are directional and, just as with gaze, it is important that they not be omitted.
- We found that it is impossible for video-mediated conversations to look just like face-to-face conversations, with high-quality image resolution, no transmission delays and simulated eye-to-eye contact. Most of the research conducted on improving interactions within telepresence robot platforms used an overall ideal condition in order to test their hypotheses and come up with suggestions, and although their findings added to the research and helped in

building knowledge, in the real world, users do not get clear benefits from them. As **the actual capabilities we have** in the real world do not match the capabilities that are used in the research.

- **A new protocol for the translation of the vocal stream** into gesture to generate human-like behaviour and support more natural interaction within the embodied technology system was presented and investigated in this thesis. Such a technique could incorporate with telepresence in order to compliment or exaggerate the affective expressions.
- **Custom UTAUT constructs were also developed** in order to cover the particular area of user acceptability. Their categories have been designed to reflect particular issues in user experience encountered in telepresence video-mediated technology, as described in the literature. Highlighting the effects of interacting in social manner on perceiving the technology as a social partner and therefore as a social entity. In addition, perceiving the technology as a social partner will features the system as an engage and absorb by the users. This experience will be highlighted as an enjoyable throughout. This in turns tends to have a positive impact on attitude towards the system.
- An **immediate consequence of the philosophy of design** used in the proposed HCI interface was that user satisfaction was significantly improved in respect of the social presence and enjoyment. This was largely due to the simplicity of the interface and its capacity for gesture recognition, which was able to clearly and responsively convey non-verbal behavioural information.
- A set of **guidelines for 'personifying' telepresence** conversations and development of such systems.

9.7 Conclusions

Although TP is generally believed that it is not as effective as face to face, we argue that careful design based on an understanding of video display aspects in TP, and how these can be used to improve non-verbal communication, can mitigate these differences. We hope that this work encourages a fresh look at video-mediated as an alternative to in-person meetings, especially given the economic and environmental cost of travel; all reasonable alternatives deserve consideration.

In summary, mobile telepresence is already becoming an important part of our lives. To drive forth research it is not simply a matter of developing higher resolution screens or better bandwidth connections, we must appreciate the nature of human communications. This work has identified, and utilised, key areas of human communications to inform telepresence system design. In personifying the technology through the sensitive use of human-centred form and movement, it is hoped that these initial explorations might form the foundations for further development within the area and underpin the creation of more realistic and engaging systems.

References

Adalgeirsson, S.O. and Breazeal, C. (2010) Published. 'MeBot: a robotic platform for socially embodied presence'. *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction,* 2010. IEEE Press, pp.15-22.

Agarwal, R. and Karahanna, E. (2000) 'Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage'. *MIS quarterly,* 665-694.

Agarwal, R., Sambamurthy, V. and Stair, R.M. (1997) Published. 'COGNITIVE ABSORPTION AND THE ADOPTION OF NEW INFORMATION TECHNOLOGIES'. *Academy of Management Proceedings,* 1997. Academy of Management, pp.293-297.

Ajzen, I. and Fishbein, M. (1988) 'Theory of reasoned action-Theory of planned behavior'. *University of South Florida*.

Akin, D.L. *et al.* (1983) 'Space Applications of Automation, Robotics and Machine Intelligence Systems (ARAMIS), phase 2. Volume 1: Telepresence technology base development'.

Anderson, A.H. *et al.* (1991) 'The HCRC map task corpus'. *Language and speech,* 34 (4), pp. 351-366.

Ankrum, D.R. (1996) 'Viewing distance at computer workstations'. *Workplace Ergonomics,* 2 (5), pp. 10-13.

Argyle, M. (1973) 'The syntaxes of bodily communication'. *Linguistics,* 11 (112), pp. 71-92.

Arras, K.O. and Cerqui, D. (2005) 'Do we want to share our lives and bodies with robots? A 2000 people survey'.

Ates, A. (2008) 'Fundamental concepts in management research and ensuring research quality: focusing on case study method'. *European Academy of Management,* 8 35.

Baldwin, D.A. and Baird, J.A. (2001) 'Discerning intentions in dynamic human action'. *Trends in cognitive sciences,* 5 (4), pp. 171-178.

Banse, R. and Scherer, K.R. (1996) 'Acoustic profiles in vocal emotion expression'. *Journal of personality and social psychology,* 70 (3), pp. 614.

Bar-Cohen, Y. (2003) 'Haptic devices for virtual reality, telepresence, and human-assistive robotics'. *Biol Inspired Intell Robots,* 122 73.

Baron-Cohen, S. (1991) 'Precursors to a theory of mind: Understanding attention in others'.

Bartneck, C. (2000) 'Affective expressions of machines'. *emotion,* 8 489-502.

Bartneck, C. (2003) Published. 'Interacting with an embodied emotional character'. *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces,* 2003. ACM, pp.55-60.

Bartneck, C. and Forlizzi, J. (2004) Published. 'Shaping human-robot interaction: understanding the social aspects of intelligent robotic products'. *CHI'04 Extended Abstracts on Human Factors in Computing Systems,* 2004. ACM, pp.1731-1732.

Bates, J. (1994) 'The role of emotion in believable agents'. *Communications of the ACM,* 37 (7), pp. 122-125.

Beattie, G.W. and Barnard, P. (1979) 'The temporal structure of natural telephone conversations (directory enquiry calls)'. *Linguistics,* 17 (3-4), pp. 213-230.

Beech, N. (2005) 'Research Methodology Course Notes: Strathclyde Business School'.

Beerends, J.G. and De Caluwe, F.E. (1999) 'The influence of video quality on perceived audio quality and vice versa'. *Journal of the Audio Engineering Society,* 47 (5), pp. 355-362.

Bickmore, T. and Cassell, J. (2001) Published. 'Relational agents: a model and implementation of building user trust'. *Proceedings of the SIGCHI conference on Human factors in computing systems,* 2001. ACM, pp.396-403.

Bickmore, T. and Schulman, D. (2006) Published. 'The comforting presence of relational agents'. *CHI'06 Extended Abstracts on Human Factors in Computing Systems,* 2006. ACM, pp.550-555.

Bickmore, T.W., Caruso, L. and Clough-Gorr, K. (2005) Published. 'Acceptance and usability of a relational agent interface by urban older adults'. *CHI'05 extended abstracts on Human factors in computing systems,* 2005. ACM, pp.1212-1215.

Biocca, F. and Harms, C. (2002) 'Defining and measuring social presence: Contribution to the networked minds theory and measure'. *Proceedings of PRESENCE,* 2002 1-36.

Birdwhistell, R.L. (1966) 'Some relations between American kinesics and spoken American English'. *Communication and Culture. New York: Holt, Rinehart and Winston,* 182-189.

Birdwhistell, R.L. (1983) 'Background to kinesics'. *Etc.,* 40 (3), pp. 352-361.

Birdwhistell, R.L. (2011) *Kinesics and context: Essays on body motion communication.* University of Pennsylvania press.

Boissy, P. *et al.* (2007) 'A qualitative study of in-home robotic telepresence for home care of community-living elderly subjects'. *Journal of Telemedicine and Telecare,* 13 (2), pp. 79-84.

Boomer, D.S. (1965) 'Hesitation and grammatical encoding'. *Language and speech,* 8 (3), pp. 148-158.

Bowling, A. (1991) *Measuring health: a review of quality of life measurement scales.* Open University Press Milton Keynes.

Boynton, P.M. and Greenhalgh, T. (2004) 'Selecting, designing, and developing your questionnaire'. *Bmj,* 328 (7451), pp. 1312-1315.

Breazeal, C. (2003) 'Toward sociable robots'. *Robotics and autonomous systems,* 42 (3), pp. 167-175.

Breazeal, C. (2004) 'Social interactions in HRI: the robot view'. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* 34 (2), pp. 181-186.

Breazeal, C.L. (2000) *Sociable machines: Expressive social exchange between humans and robots.* Massachusetts Institute of Technology.

Brink, H., Van der Walt, C. and Van Rensburg, G. (2006) *Fundamentals of research methodology for health care professionals.* Juta and Company Ltd.

Brown, E. and Cairns, P. (2004) Published. 'A grounded investigation of game immersion'. *CHI'04 extended abstracts on Human factors in computing systems,* 2004. ACM, pp.1297-1300.

Bruce, A., Nourbakhsh, I. and Simmons, R. (2002) Published. 'The role of expressiveness and attention in human-robot interaction'. *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on,* 2002. IEEE, pp.4138-4142.

Bryman, A. (2015) *Social research methods.* Oxford university press.

Bryman, A. and Cramer, D. (2005) *Quantitative data analysis with SPSS 12 and 13: a guide for social scientists.* Psychology Press.

Busso, C. *et al.* (2004) Published. 'Analysis of emotion recognition using facial expressions, speech and multimodal information'. *Proceedings of the 6th international conference on Multimodal interfaces,* 2004. ACM, pp.205-211.

Cassell, J. (2000) 'Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents'. *Embodied conversational agents,* 1-27.

Cegala, D.J. (1984) 'Affective and cognitive manifestations of interaction involvement during unstructured and competitive interactions'. *Communications Monographs,* 51 (4), pp. 320-338.

Cegala, D.J. *et al.* (1982) 'An elaboration of the meaning of interaction involvement: Toward the development of a theoretical concept'. *Communications Monographs,* 49 (4), pp. 229-248.

Chandra, S., Srivastava, S. and Theng, Y.-L. (2009) 'Role of cognitive absorption and trust for collaboration in virtual world'. *PACIS 2009 Proceedings,* 82.

Chapanis, A. *et al.* (1972) 'Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving'. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 14 (6), pp. 487-509.

Chesney, T. (2006) 'An acceptance model for useful and fun information systems'. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments,* 2 (2), pp. 225-235.

Codourey, A., Rodriguez, M. and Pappas, I. (1997) Published. 'A task-oriented teleoperation system for assembly in the microworld'. *Advanced Robotics, 1997. ICAR'97. Proceedings., 8th International Conference on,* 1997. IEEE, pp.235-240.

Coker, D.A. and BURGOON, J. (1987) 'The nature of conversational involvement and nonverbal encoding patterns'. *Human Communication Research,* 13 (4), pp. 463-494.

Condon, W.S. and Ogston, W.D. (1966) 'Sound film analysis of normal and pathological behavior patterns'. *The Journal of Nervous and Mental Disease,* 143 (4), pp. 338-347.

Condon, W.S. and Ogston, W.D. (1967) 'A segmentation of behavior'. *Journal of psychiatric research,* 5 (3), pp. 221-235.

Cook, D.A. and Beckman, T.J. (2006) 'Current concepts in validity and reliability for psychometric instruments: theory and application'. *The American journal of medicine,* 119 (2), pp. 166. e167-166. e116.

Cook, M. and Lalljee, M.G. (1972) 'Verbal substitutes for visual signals in interaction'. *Semiotica,* 6 (3), pp. 212-221.

Coradeschi, S. *et al.* (2011) Published. 'Towards a methodology for longitudinal evaluation of social robotic telepresence for elderly'. *Human Robot Interaction,* 2011.

Cramer, D. and Howitt, D.L. (2004) *The Sage dictionary of statistics: a practical resource for students in the social sciences.* Sage.

Csikszentmihalyi and Mihaly (1990) *Flow: The Psychology of Optimal Experience.* New York: Harper and Row.

Cui, J. *et al.* (2003) Published. 'A review of teleoperation system control'. *Proceedings of the Florida Conference on Recent Advances in Robotics,* 2003. Florida Atlantic University Boca Raton, FL.

Dario, P., Guglielmelli, E. and Laschi, C. (2001) 'Humanoids and personal robots: Design and experiments'. *Journal of robotic systems,* 18 (12), pp. 673-690.

Davis, F.D. (1989) 'Perceived usefulness, perceived ease of use, and user acceptance of information technology'. *MIS quarterly,* 319-340.

Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. (1992) 'Extrinsic and intrinsic motivation to use computers in the workplace1'. *Journal of applied social psychology,* 22 (14), pp. 1111-1132.

de Oliveira, H.F.P. *et al.* (2008) Published. 'Dynamical models for omni-directional robots with 3 and 4 wheels'. *Proceedings of the 5th International Conference on Informatics in Control, Automation and Robotics,* 2008.

De Ruyter, B. and Aarts, E. (2004) Published. 'Ambient intelligence: visualizing the future'. *Proceedings of the working conference on Advanced visual interfaces,* 2004. ACM, pp.203-208.

De Ruyter, B. *et al.* (2005) 'Assessing the effects of building social intelligence in a robotic interface for the home'. *Interacting with computers,* 17 (5), pp. 522-541.

DeCarlo, D. *et al.* (2004) 'Specifying and animating facial signals for discourse in embodied conversational agents'. *Computer Animation and Virtual Worlds,* 15 (1), pp. 27-38.

Dennett, D.C. (1978) 'Three kinds of intentional psychology'. *Perspectives in the Philosophy of Language: A Concise Anthology,* 163-186.

Diegel, O. *et al.* (2002) Published. 'Improved mecanum wheel design for omni-directional robots'. *Proc. 2002 Australasian Conference on Robotics and Automation, Auckland,* 2002. pp.117-121.

Doherty-Sneddon, G. *et al.* (1997) 'Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance'. *Journal of Experimental Psychology: Applied,* 3 (2), pp. 105.

Dunbar, R. (1998) *Grooming, gossip, and the evolution of language.* Harvard University Press.

Duran, R.L. and Kelly, L. (1988) 'An investigation into the cognitive domain of competence II: The relationship between communicative competence and interaction involvement'. *Communication Research Reports,* 5 (1), pp. 91-96.

Easterby-Smith, M., Thorpe, R. and Jackson, P.R. (2012) *Management research.* Sage.

Ekman, P. and Friesen, W.V. (1967) 'Head and body cues in the judgment of emotion: A reformulation'. *Perceptual and motor skills,* 24 (3), pp. 711-724.

Ekman, P. and Friesen, W.V. (1982) 'Felt, false, and miserable smiles'. *Journal of nonverbal behavior,* 6 (4), pp. 238-252.

Ellis, A.W. and Beattie, G. (1986) *The psychology of language and communication.* Weidenfeld and Nicolson.

Ellison, L.M. *et al.* (2004) 'Telerounding and patient satisfaction after surgery'. *Journal of the American College of Surgeons,* 199 (4), pp. 523-530.

Evaristo, J.R. and Karahanna, E. (1998) 'The impact of mental workload in the evaluation of innovations'. *Information systems innovation and diffusion: Issues and directions,* 48-70.

Ferrell, W.R. and Sheridan, T.B. (1967) 'Supervisory control of remote manipulation'. *IEEE spectrum,* 4 (10), pp. 81-88.

Field, A. (2013) *Discovering statistics using IBM SPSS statistics.* Sage.

Findlay, R.A. (2003) 'Interventions to reduce social isolation amongst older people: where is the evidence?'. *Ageing and Society,* 23 (05), pp. 647-658.

Fink, A. (2013) *Conducting research literature reviews: from the Internet to paper.* Sage Publications.

Fish, R.S. *et al.* (1992) 'Evaluating video as a technology for informal communication'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Monterey, California, USA, 142755: ACM.

Fong, T., Nourbakhsh, I. and Dautenhahn, K. (2003) 'A survey of socially interactive robots'. *Robotics and autonomous systems,* 42 (3), pp. 143-166.

Fowler, C. and Mayes, T. (1997) 'Applying telepresence to education'. *BT technology Journal,* 15 (4), pp. 188-195.

Fraenkel, J.R., Wallen, N.E. and Hyun, H.H. (1993) *How to design and evaluate research in education.* McGraw-Hill New York.

Freeman, J. and Avons, S.E. (2000) Published. 'Focus group exploration of presence through advanced broadcast services'. *Electronic Imaging,* 2000. International Society for Optics and Photonics, pp.530-539.

George, D. and Mallery, P. (2016) *IBM SPSS Statistics 23 Step by Step: A Simple Guide and Reference.* Routledge.

Ghani, J.A. and Deshpande, S.P. (1994) 'Task characteristics and the experience of optimal flow in human—computer interaction'. *The Journal of psychology,* 128 (4), pp. 381-391.

Given, L.M. (2008) *The Sage encyclopedia of qualitative research methods.* Sage Publications.

Goetz, J. and Kiesler, S. (2002) Published. 'Cooperation with a robotic assistant'. *CHI'02 Extended Abstracts on Human Factors in Computing Systems,* 2002. ACM, pp.578-579.

Goetz, J., Kiesler, S. and Powers, A. (2003) Published. 'Matching robot appearance and behavior to tasks to improve human-robot cooperation'. *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on,* 2003. IEEE, pp.55-60.

Goffman, E. (1999) *The Presentation of Self in Everyday Life.* Peter Smith Publisher, Incorporated.

Goodwin, C. (1981) *Conversational Organization: Interaction Between Speakers and Hearers.* Academic Press.

Gould, J.D. and Lewis, C. (1985) 'Designing for usability: key principles and what designers think'. *Communications of the ACM,* 28 (3), pp. 300-311.

Graf, H.P. *et al.* (2002) Published. 'Visual prosody: Facial movements accompanying speech'. *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on,* 2002. IEEE, pp.396-401.

Green, P.S. *et al.* (1995) 'Telepresence surgery'. *Engineering in Medicine and Biology Magazine, IEEE,* 14 (3), pp. 324-329.

Grosz, B.J. and Kraus, S. (1996) 'Collaborative plans for complex group action'. *Artificial Intelligence,* 86 (2), pp. 269-357.

Guerrero, L.K. (1994) 'An application of attachment theory to relational messages and nonverbal involvement behaviors in romantic relationships'.

Guerrero, L.K. (2005) 'Observer ratings of nonverbal involvement and immediacy'. *The sourcebook of nonverbal measures: Going beyond words,* 221-235.

Guizzo, E. (2010) 'When my avatar went to work'. *IEEE Spectr.,* 47 (9), pp. 26-31.

Gunn, J. (1996) *Isaac Asimov: the foundations of science fiction.* Scarecrow Press.

Hadar, U. *et al.* (1983) 'Kinematics of head movements accompanying speech during conversation'. *Human Movement Science,* 2 (1), pp. 35-46.

Hadar, U., Steiner, T. and Rose, F.C. (1985a) 'Head movement during listening turns in conversation'. *Journal of Nonverbal Behavior,* 9 (4), pp. 214-228.

Hadar, U., Steiner, T.J. and Rose, F.C. (1985b) 'Head movement during listening turns in conversation'. *Journal of Nonverbal Behavior,* 9 (4), pp. 214-228.

Heerink, M. (2011) Published. 'Exploring the influence of age, gender, education and computer experience on robot acceptance by older adults'. *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on,* 2011. IEEE, pp.147-148.

Heerink, M. *et al.* (2008a) Published. 'The influence of social presence on enjoyment and intention to use of a robot and screen agent by elderly users'. *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on,* 2008a. IEEE, pp.695-700.

Heerink, M. *et al.* (2009) Published. 'Measuring acceptance of an assistive social robot: a suggested toolkit'. *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on,* 2009. IEEE, pp.528-533.

Heerink, M. *et al.* (2008b) Published. 'Responses to a social robot by elderly users'. *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on,* 2008b. IEEE, pp.2724-2724.

Heerink, M. *et al.* (2006) Published. 'The influence of a robot's social abilities on acceptance by elderly users'. *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on,* 2006. IEEE, pp.521-526.

Heerink, M. *et al.* (2010) 'Assessing acceptance of assistive social agent technology by older adults: the almere model'. *International journal of social robotics,* 2 (4), pp. 361-375.

Heinlein, R.A. (1969) *Waldo, and Magic, Inc.* Alan G. Gates.

Heylen, D. (2005) 'Challenges ahead: head movements and other social acts during conversations'.

Hinds, P.J., Roberts, T.L. and Jones, H. (2004) 'Whose job is it anyway? A study of human-robot interaction in a collaborative task'. *Human-Computer Interaction,* 19 (1), pp. 151-181.

Hsu, H. and Lachenbruch, P.A. (2008) 'Paired t test'. *Wiley Encyclopedia of Clinical Trials*.

IJsselsteijn, W.A. *et al.* (2000) Published. 'Presence: concept, determinants, and measurement'. *Electronic Imaging,* 2000. International Society for Optics and Photonics, pp.520-529.

Imai, M., Ono, T. and Ishiguro, H. (2003) 'Physical relation and expression: Joint attention for human-robot interaction'. *Industrial Electronics, IEEE Transactions on,* 50 (4), pp. 636-643.

Ishiguro, H. and Trivedi, M. (1999) Published. 'Integrating a perceptual information infrastructure with robotic avatars: a framework for tele-existence'. *Intelligent Robots and Systems, 1999. IROS'99. Proceedings. 1999 IEEE/RSJ International Conference on,* 1999. IEEE, pp.1032-1038.

Iwano, Y. *et al.* (1996) Published. 'Analysis of head movements and its role in spoken dialogue'. *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on,* 1996. IEEE, pp.2167-2170.

Jack, B. and Clarke, A.M. (1998) 'The purpose and use of questionnaires in research'. *Professional nurse (London, England),* 14 (3), pp. 176.

Kantonen, T., Woodward, C. and Katz, N. (2010) Published. 'Mixed reality in virtual world teleconferencing'. *2010 IEEE Virtual Reality Conference (VR),* 2010. IEEE, pp.179-182.

Karvonen, K. (2000) Published. 'The beauty of simplicity'. *Proceedings on the 2000 conference on Universal Usability,* 2000. ACM, pp.85-90.

Kelley, J.F. (1983) Published. 'An empirical methodology for writing user-friendly natural language computer applications'. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems,* 1983. ACM, pp.193-196.

Kendon, A. (1970) 'Movement coordination in social interaction: Some examples described'. *Acta psychologica,* 32 101-125.

Kendon, A. (1972) 'Some relationships between body motion and speech'. *Studies in dyadic communication,* 7 177.

Kidd, C.D. and Breazeal, C. (2004) Published. 'Effect of a robot on user perceptions'. *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on,* 2004. IEEE, pp.3559-3564.

Kies, J.K., Williges, R.C. and Rosson, M.B. (1997) 'Evaluating desktop video conferencing for distance learning'. *Computers & Education,* 28 (2), pp. 79-91.

Knapp, M., Hall, J. and Horgan, T. (2013) *Nonverbal communication in human interaction.* Cengage Learning.

Knapp, M.L. and Daly, J.A. (2011) *The SAGE handbook of interpersonal communication.* Sage Publications.

Knoche, H., McCarthy, J.D. and Sasse, M.A. (2005) Published. 'Can small be beautiful?: assessing image resolution requirements for mobile TV'. *Proceedings of the 13th annual ACM international conference on Multimedia,* 2005. ACM, pp.829-838.

Kose-Bagci, H. *et al.* (2009) 'Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot'. *Advanced Robotics,* 23 (14), pp. 1951-1996.

Kristoffersson, A., Coradeschi, S. and Loutfi, A. (2013) 'A review of mobile robotic telepresence'. *Advances in Human-Computer Interaction,* 2013 3.

Laver, J. (1975) 'Communicative functions of phatic communion'. *Organization of behavior in face-to-face interaction,* 215-238.

Lee, K.M. *et al.* (2006) 'Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction'. *International Journal of Human-Computer Studies,* 64 (10), pp. 962-973.

Lee, K.M. and Nass, C. (2003) Published. 'Designing social presence of social actors in human computer interaction'. *Proceedings of the SIGCHI conference on Human factors in computing systems,* 2003. ACM, pp.289-296.

Leite, I. *et al.* (2008) Published. 'Are emotional robots more fun to play with?'. *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication,* 2008. IEEE, pp.77-82.

Li, L. *et al.* (1996) Published. 'Development of a telepresence controlled ambidextrous robot for space applications'. *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on,* 1996. IEEE, pp.58-63.

Libin, A.V. and Libin, E.V. (2004) 'Person-robot interactions from the robopsychologists' point of view: the robotic psychology and robotherapy approach'. *Proceedings of the IEEE,* 92 (11), pp. 1789-1803.

Lichtman, H.S. (2006) 'Telepresence, effective visual collaboration and the future of global business at the speed of light'. *HPL, Human Productivity Lab Magazine*.

Lidwell, W., Holden, K. and Butler, J. (2010) *Universal principles of design, revised and updated: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design.* Rockport Pub.

Lombard, M. and Jones, M.T. (2007) 'Identifying the (Tele) Presence Literature'. *PsychNology Journal,* 5 (2), pp. 197-206.

Looije, R., Cnossen, F. and Neerincx, M.A. (2006) Published. 'Incorporating guidelines for health assistance into a socially intelligent robot'. *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on,* 2006. IEEE, pp.515-520.

Looije, R., Neerincx, M.A. and Cnossen, F. (2010) 'Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors'. *International Journal of Human-Computer Studies,* 68 (6), pp. 386-397.

Mair, G.M. (1997) Published. 'Telepresence-the technology and its economic and social implications'. *Technology and Society, 1997.'Technology and Society at a Time of Sweeping Change'. Proceedings., 1997 International Symposium on,* 1997. IEEE, pp.118-124.

Mair, G.M. (2007) 'Towards transparent telepresence'. *Virtual Reality.* Springer, pp. 300-309.

Maletic, V. (1987) *Body-space-expression: The development of Rudolf Laban's movement and dance concepts.* Walter de Gruyter.

Malle, B.F., Moses, L.J. and Baldwin, D.A. (2001) *Intentions and intentionality: Foundations of social cognition.* MIT press.

Markar, S.R. *et al.* (2012) 'International survey study of attitudes towards robotic surgery'. *Journal of Robotic Surgery,* 6 (3), pp. 231-235.

Maulsby, D., Greenberg, S. and Mander, R. (1993) Published. 'Prototyping an intelligent agent through Wizard of Oz'. *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems,* 1993. ACM, pp.277-284.

Maxwell, J.A. (2008) 'Designing a qualitative study'. *The Sage handbook of applied social research methods,* 214-253.

McColl, D. and Nejat, G. (2014) 'Recognizing emotional body language displayed by a human-like social robot'. *International Journal of Social Robotics,* 6 (2), pp. 261-280.

McLellan, H. (1996) 'Virtual realities'. *Handbook of research for educational communications and technology,* 457-487.

McNeill, D. (1992) *Hand and mind: What gestures reveal about thought.* University of Chicago press.

Mehrabian, A. (1968) 'Some referents and measures of nonverbal behavior'. *Behavior Research Methods & Instrumentation,* 1 (6), pp. 203-207.

Mehrabian, A. (1972) 'Nonverbal communication'.

Mehrabian, A. (1977) *Nonverbal communication.* Transaction Publishers.

Michaud, F. *et al.* (2007) Published. 'Telepresence Robot for Home Care Assistance'. *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics,* 2007. California, USA, pp.50-55.

Minato, T. *et al.* (2006) 'Evaluating the human likeness of an android by comparing gaze behaviors elicited by the android and a person'. *Advanced Robotics,* 20 (10), pp. 1147-1163.

Minsky, M. (1980) 'Telepresence'.
Mitsunaga, N. *et al.* (2008) 'Adapting Robot Behavior for Human--Robot Interaction'. *IEEE Transactions on Robotics,* 24 (4), pp. 911-916.
Moore, M.M. (1985) 'Nonverbal courtship patterns in women: Context and consequences'. *Ethology and sociobiology,* 6 (4), pp. 237-247.
Mori and M (1970) 'The Uncanny Valley'. *Energy,* 7 (4), pp. 33-35.
Mutlu, B. *et al.* (2009) Published. 'Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior'. *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction,* 2009. ACM, pp.69-76.
Nah, F.F.-H., Eschenbrenner, B. and DeWester, D. (2011) 'Enhancing brand equity through flow and telepresence: A comparison of 2D and 3D virtual worlds'. *MIs Quarterly,* 35 (3), pp. 731-747.
Nakanishi, H., Murakami, Y. and Kato, K. (2009) Published. 'Movable cameras enhance social telepresence in media spaces'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2009. ACM, pp.433-442.
Nguyen, D.T. and Canny, J. (2009) Published. 'More than face-to-face: empathy effects of video framing'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2009. ACM, pp.423-432.
Nielsen, J. (1993) *Usability engineering.* San Francisco, Calif: Morgan Kaufmann Publishers.
Norris, A.E. *et al.* (2014) 'Quantifying engagement: Measuring player involvement in human–avatar interactions'. *Computers in human behavior,* 34 1-11.
Novak, T.P., Hoffman, D.L. and Yung, Y.-F. (2000) 'Measuring the customer experience in online environments: A structural modeling approach'. *Marketing science,* 19 (1), pp. 22-42.
O'Conaill, B., Whittaker, S. and Wilbur, S. (1993) 'Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication'. *Human-computer interaction,* 8 (4), pp. 389-428.
O'Malley, C. *et al.* (1996) 'Comparison of face-to-face and video-mediated interaction'. *Interacting with computers,* 8 (2), pp. 177-192.
Oestreicher, L. (2007) Published. 'Cognitive, social, sociable or just socially acceptable robots?'. *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication,* 2007. IEEE, pp.558-563.
Oliveira, H.P. *et al.* (2009) 'Modeling and assessing of omni-directional robots with three and four wheels'. *Contemporary Robotics: Challenges and Solutions*.
Oliver, R.L. (1997) *Satisfaction: a behavioral perspective on the consumer.* New York: McGraw Hill.
Olson, J.S., Olson, G.M. and Meader, D.K. (1995) Published. 'What mix of video and audio is useful for small groups doing remote real-time design work?'. *Proceedings of the SIGCHI conference on Human factors in computing systems,* 1995. ACM Press/Addison-Wesley Publishing Co., pp.362-368.
Oppenheim, A.N. (2000) *Questionnaire design, interviewing and attitude measurement.* Bloomsbury Publishing.
Orha, I. and Oniga, S. (2012) 'Assistance and telepresence robots: a solution for elderly people'. *Carpathian Journal of Electronic and Computer Engineering,* 5 (87), pp. 2012.

Palmer, M.T. (1995) 'Interpersonal communication and virtual reality: Mediating interpersonal relationships'. *Communication in the age of virtual reality,* 277-299.

Patrick, A. 'The use and relative benefits of Telepresence'.

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

Paulos, E. and Canny, J. (1998) Published. 'PRoP: personal roving presence'. *Proceedings of the SIGCHI conference on Human factors in computing systems,* 1998. ACM Press/Addison-Wesley Publishing Co., pp.296-303.

Payne Jr John, H. (1949) '*Remote-control manipulator'*. Google Patents.

Pittenger, R.E., Hockett, C.F. and Danehy, J.J. (1960) 'The first five minutes: A sample of microscopic interview analysis'.

Powers, A. and Kiesler, S. (2006) Published. 'The advisor robot: tracing people's mental model from a robot's physical attributes'. *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction,* 2006. ACM, pp.218-225.

Powers, A. *et al.* (2005) Published. 'Eliciting information from people with a gendered humanoid robot'. *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.,* 2005. IEEE, pp.158-163.

Reeves, B. (1991) 'Being there: Television as symbolic versus natural experience'. *Unpublished manuscript, Institute for Communication Research, Stanford University, Stanford, CA*.

Rheingold, H. (1991) *Virtual Reality: Exploring the Brave New Technologies.* Simon & Schuster Adult Publishing Group.

Richards and Lara (2011) *I, student . Homebound Teenis attending Classes Via Wheel-bound Robot.* Available at: http://www.timesrecordnews.com/news/columnists/i-student (Accessed: 11 September).

Rico, J. and Brewster, S. (2010) Published. 'Usable gestures for mobile interfaces: evaluating social acceptability'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 2010. ACM, pp.887-896.

Riek, L.D. (2012) 'Wizard of oz studies in hri: a systematic review and new reporting guidelines'. *Journal of Human-Robot Interaction,* 1 (1), pp.

Robins, B. *et al.* (2007) Published. 'Eliciting requirements for a robotic toy for children with autism-results from user panels'. *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication,* 2007. IEEE, pp.101-106.

Rogoff, B. (1991) 'Perspectives on socially shared cognition'.

Rosenfeld, H.M. (1978) 'Conversational control functions of nonverbal behavior'. *Nonverbal behavior and communication,* 291-328.

Rousseau *et al.* (1997) 'Interacting with Personality-Rich Characters'. *Technical Report No. KSL 97-06,* pp. doi.

Rubin, R.B., Perse, E.M. and Barbato, C.A. (1988) 'Conceptualization and measurement of interpersonal communication motives'. *Human Communication Research,* 14 (4), pp. 602-628.

Rutter, D.R. and Stephenson, G.M. (1977) 'The role of visual communication in synchronising conversation'. *European Journal of Social Psychology,* 7 (1), pp. 29-37.

Saadé, R. and Bahli, B. (2005) 'The impact of cognitive absorption on perceived usefulness and perceived ease of use in on-line learning: an extension of the technology acceptance model'. *Information & management,* 42 (2), pp. 317-327.

Salih, J.E.M. *et al.* (2006) 'Designing omni-directional mobile robot with mecanum wheel'. *American Journal of Applied Sciences,* 3 (5), pp. 1831-1835.

Sarantakos, S. (2012) *Social research.* Palgrave Macmillan.

Schloerb, D.W. (1995) 'A quantitative measure of telepresence'. *Presence: Teleoperators and Virtual Environments,* 4 (1), pp. 64-80.

schneider, K.P. (1988) *Small talk : analysing phatic discourse.* Marburg: Hitzeroth.

Sellen, A.J. (1995) 'Remote conversations: The effects of mediating talk with technology'. *Human-computer interaction,* 10 (4), pp. 401-444.

Shadish, W.R., Cook, T.D. and Campbell, D.T. (2002) *Experimental and quasi-experimental designs for generalized causal inference.* Houghton, Mifflin and Company.

Sharbafi, M.A., Lucas, C. and Daneshvar, R. (2010) 'Motion control of omni-directional three-wheel robots by brain-emotional-learning-based intelligent controller'. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* 40 (6), pp. 630-638.

Sheridan, T.B. (1992a) 'Musings on telepresence and virtual presence'. *Presence: Teleoper. Virtual Environ.,* 1 (1), pp. 120-126.

Sheridan, T.B. (1992b) *Telerobotics, automation and human supervisory control.* The MIT press.

Shieh, K.-K. (2000) 'Effects of reflection and polarity on LCD viewing distance'. *International Journal of Industrial Ergonomics,* 25 (3), pp. 275-282.

Shieh, K.-K. and Lee, D.-S. (2007) 'Preferred viewing distance and screen angle of electronic paper displays'. *Applied Ergonomics,* 38 (5), pp. 601-608.

Short, J., Williams, E. and Christie, B. (1976) 'The social psychology of telecommunications'.

Sidelinger, R.J. *et al.* (2008) 'Couples go online: Relational maintenance behaviors and relational characteristics use in dating relationships'. *Human communication,* 11 333-348.

Sidner, C.L. (2004) 'Building spoken-language collaborative interface agents'. *Practical Spoken Dialog Systems.* Springer, pp. 197-226.

Sidner, C.L., Lee, C. and Kidd, C. (2005) Published. 'Engagement during dialogues with robots'. *AAAI spring symposia,* 2005.

Sidner, C.L. *et al.* (2005) 'Explorations in engagement for humans and robots'. *Artificial Intelligence,* 166 (1), pp. 140-164.

Simmel, G. (1910) 'How is Society Possible?'. *American Journal of Sociology,* 16 (3), pp. 372-391.

Sirkin, D., Ju, W. and Cutkosky, M. (2012) 'Communicating Meaning and Role in Distributed Design Collaboration: How Crowdsourced Users Help Inform the Design of Telepresence Robotics'. *Design Thinking Research.* Springer, pp. 173-187.

Slater, M. and Usoh, M. (1993) 'Representations systems, perceptual position, and presence in immersive virtual environments'. *Presence: Teleoperators & Virtual Environments,* 2 (3), pp. 221-233.

Sun, H. and Zhang, P. (2006) 'Causal relationships between perceived enjoyment and perceived ease of use: An alternative approach'. *Journal of the Association for Information Systems,* 7 (9), pp. 24.

Tabachnick, B.G., Fidell, L.S. and Osterlind, S.J. (2001) 'Using multivariate statistics'.

Tapus, A. and Mataric, M.J. (2008) Published. 'Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance'. *AAAI Spring Symposium: Emotion, Personality, and Social Behavior,* 2008. pp.133-140.

Tellegen, A. and Atkinson, G. (1974) 'Openness to absorbing and self-altering experiences (" absorption"), a trait related to hypnotic susceptibility'. *Journal of abnormal psychology,* 83 (3), pp. 268.

Thabane, L. *et al.* (2010) 'A tutorial on pilot studies: the what, why and how'. *BMC medical research methodology,* 10 (1), pp. 1.

Thórisson, K.R. (1997) Published. 'Gandalf: An embodied humanoid capable of real-time multimodal dialogue with people'. *Agents,* 1997. pp.536-537.

Tojo, T. *et al.* (2000) Published. 'A conversational robot utilizing facial and body expressions'. *Systems, Man, and Cybernetics, 2000 IEEE International Conference on,* 2000. IEEE, pp.858-863.

Trager, G.L. and Smith, H.L. (2009) *An outline of English structure.* Рипол Классик.

Trevino, L.K. and Webster, J. (1992) 'Flow in computer-mediated communication electronic mail and voice mail evaluation and impacts'. *Communication research,* 19 (5), pp. 539-573.

Trout, D.L. and Rosenfeld, H.M. (1980) 'The effect of postural lean and body congruence on the judgment of psychotherapeutic rapport'. *Journal of Nonverbal Behavior,* 4 (3), pp. 176-190.

Tsui, K.M. *et al.* (2011) Published. 'Exploring use cases for telepresence robots'. *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on,* 2011. IEEE, pp.11-18.

Tsui, K.M. *et al.* (2012) Published. 'Robots in the loop: telepresence robots in everyday life'. *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction,* 2012. ACM, pp.317-318.

Van der Heijden, H. (2004) 'User acceptance of hedonic information systems'. *MIS quarterly,* 695-704.

Venkatesh, V. *et al.* (2003) 'User acceptance of information technology: Toward a unified view'. *MIS quarterly,* 425-478.

Villaume, W.A. and Cegala, D.J. (1988) 'Interaction involvement and discourse strategies: The patterned use of cohesive devices in conversation'. *Communications Monographs,* 55 (1), pp. 22-40.

Weiss, A. *et al.* (2009) Published. 'The usus evaluation framework for human-robot interaction'. *AISB2009: proceedings of the symposium on new frontiers in human-robot interaction,* 2009. pp.11-26.

Weiss, P.L.T. *et al.* (2001) 'PEBBLES: a personal technology for meeting educational, social and emotional needs of hospitalised children'. *Personal and Ubiquitous Computing,* 5 (3), pp. 157-168.

Weniger, S. and Loebbecke, C. (2010) Published. 'Cognitive absorption: Literature review and suitability in the context of hedonic IS usage'. 2010. Annual SIG-IS Cognitive Research Exchange Workshop, St. Louis, MO.

Whittaker, S. and O'Conaill, B. (1997) 'The role of vision in face-to-face and mediated communication'.

Williams, E. (1977) 'Experimental comparisons of face-to-face and mediated communication: A review'. *Psychological Bulletin,* 84 (5), pp. 963.

Yoerger, D.R. and Slotine, J.-J.E. (1987) Published. 'Supervisory control architecture for underwater teleoperation'. *Robotics and Automation. Proceedings. 1987 IEEE International Conference on,* 1987. IEEE, pp.2068-2073.

Yung, S.T. and Piebalga, A. (2013) *Design Directions: The Relationship Between Humans and Technology.* Cambridge Scholars Publishing.

Yung, S.T., Tang, M.-X. and Justice, L. (2009) *Modelling and Evaluation of Emotional Interfaces.*

Zaad, L. and Allouch, S.B. (2008) 'The influence of control on the acceptance of ambient intelligence by elderly people: An explorative study'. *Ambient Intelligence.* Springer, pp. 58-74.

Zhang, P., Li, N. and Sun, H. (2006) Published. 'Affective quality and cognitive absorption: Extending technology acceptance research'. *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on,* 2006. IEEE, pp.207a-207a.

Appendices

Appendix A

Appendix A presents the material used in the exploratory experiment. This covers the pre-test questionnaire, post-test questionnaire and the observational rating scheme.

Appendix A 1

PURPOSE OF THE EXPERIMENT

The aim of this study is to compare performance and effects on communication under two different conditions; video-mediated interaction with in-space and on-screen head movements, and video-mediated interaction with on-screen only head movements.

A short task will be completed under each set of conditions, and you will take part in the experiments alongside another participant. Each task will take no longer than 15 to 20 minutes (depending on how you manage to finish the given task) and in each pair of participants, there will be a Picker and a Guesser.

Remember, this is all totally voluntary. If you should become uncomfortable or find any session objectionable in any way, feel free to stop at any time.

If you agree to take part in the two tasks, please sign here: ……………………………………..…….

The Task

Thank you very much for agreeing to participate in this experiment, and for agreeing to work under the two conditions, with another participant.

In the task, one player (the Picker) will choose an item from among twelve items printed in A5 size, without identifying it to the other player (the Guesser). The Guesser will try to guess the item chosen by the Picker by asking a set of questions that can be answered with "Yes" and "No". However, the Picker can give the Guesser some hints if they want to do so. Please be aware that we will be recording everything that you do under both conditions, and the recordings will be kept strictly confidential.

First Condition

For the mediated conditions, you will not be able to see each other physically as a black curtain will be erected between you, adjusted to block any direct view of one another's faces. You will work collaboratively together and will be able to hear and talk to each other without wearing headsets. You will see one another's faces through the webcams (placed on top of each screen), and both sides will have a camera set up to record their interactions.

Second Condition

Similar to the first condition; however, the experimenter will take a seat beside the screen displaying the head movements of one person, and reproduce those movements on the other person's screen. The researcher will try to reproduce the exact head movements.

Further instructions

Remember, you are permitted to speak to each other, and the Guesser may ask as many questions as they wish during the exercise to determine the right answer and complete the task. The Picker will also have an equal chance to explain his/her choice, but we encourage them to make it difficult for the Guesser by not giving them a direct explanation.

As you work through the tasks, please note that we will not provide help or to answer any questions, as the exercise is about participants communicating with each other in order to complete the task.

After each task, we will ask you to complete a questionnaire to provide us with feedback on your experiences. The first questionnaire will pertain to both tasks; the second questionnaire will pertain to the second task only.

Could you please fill in your answers to these questions; we can assure you that all of your responses will be kept strictly confidential, so please answer as accurately and honestly as possible.

- How old are you (in years)? ____ ____
- What is your gender? ____ Male ____ Female
- Do you have any sight problems? Yes No
- Do you have any hearing problems? Yes No

Do you have any questions before we start?

Appendix A 2

INSTRUCTIONS

This questionnaire is designed to provide information about how subjects communicate during a given task. There are no right or wrong answers to any of the questions so please just respond to each item in a way that best describes your typical manner of communication.

How you would usually behave. If you cannot decide how a particular item applies to you, circle the "not sure" alternative, but please ensure you respond to all of the items.

Involvement

Please circle the responses that best represent your answers. All of your responses will be kept strictly confidential.

	Not at all like me	Not like me	Not sure	Like me	Very much like me
1. I am keenly aware of how my partner perceives me during my conversation.					
2. My mind wanders during the conversation and I often miss parts of what is going on.					
3. Often in conversation, I'm not sure what to say, I can't seem to find the appropriate words.					
4. I am very observant of my partner's reaction while I'm speaking.					
5. During the conversation I listen carefully to my partner and obtain as much information as I can.					
6. Often in conversation, I'm not sure what my role is, I'm not sure how I'm expected to relate to my partner.					
7. Often in conversation, I pretend to be listening when in fact, I am thinking of something else.					
8. Often during a conversation I feel like I know what should be said (like accepting, a compliment or					

asking a question) but I hesitate to do so.					
	Not at all like me	Not like me	Not sure	Like me	Very much like me
9. Sometimes during a conversation I'm not sure what my partner really mean or intend by certain comments.					
10. I carefully observe how partners is responding to me during the conversation.					
11. Often I feel withdrawn or distant during the conversation.					
12. Often in conversation, I'm not sure what my partner's needs are (e.g., a compliment, reassurance, etc.) until it is too late to respond appropriately.					
13. I feel confident during the conversation, I am sure of what to say and do.					
14. Often I'm preoccupied in my conversation and do not pay complete attention, my partner.					
15. Often I feel sort of "unplugged" during the conversation, I am uncertain of my role, partner's motive, and what is happening.					
16. In my conversation, I often do not accurately perceive partner's intention or motivations.					
17. In a conversation I am very perceptive of the meaning of my partner's behaviour in relation to myself and the situation.					
18. Often during a conversation I can't think of what to say, I just don't react quickly enough.					

Appendix A 3

Please indicate how much you disagree or agree with each statement below.

	Not at all like me	Not like me	Not sure	Like me	Very much like me
1. I was confused by where and when the screen moved.					
2. The screen moved at the appropriate times.					
3. The screen movement helped in improving the clarity of the conversation.					
4. The screen movement did not help in enhancing the communication.					

In your opinion, which experience did you think was the more successful (1 or 2), and why?

Please use the space below to provide your comments about the media experience in the second condition.

Appendix A 4

Behavioural observation

Parameters	Normal	Stimulated
Length		
Number of Questions		
Number of Answers		
Nods		
Smiling		
Adaptors		
Rocking or twisting		

Appendix B

Appendix B presents the material used in the MobiBot system evaluation

Appendix B 1

- ***The design of the System***

    - Very appealing
    - Somewhat appealing
    - Neither appealing nor unappealing
    - Somewhat unappealing
    - Very unappealing

- ***The size of the System***

    - Strongly like
    - Somewhat like
    - Neither like nor dislike
    - Somewhat dislike
    - Strongly dislike

- ***The height of the System***

    - Strongly dislike
    - Somewhat dislike
    - Neither dislike nor dislike
    - Somewhat like
    - Strongly like

- ***The colour of the System***

    - Strongly dislike
    - Somewhat dislike
    - Neither oppose nor favour
    - Somewhat like
    - Strongly like

- ***The System looked similar to a human***

    - Strongly agree
    - Somewhat agree
    - Neither agree nor disagree
    - Somewhat disagree
    - Strongly disagree

- ***The System has human-like behaviour features***

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

- ***The System's forward movement features***

  - Very appealing
  - Somewhat appealing
  - Neither appealing nor unappealing
  - Somewhat unappealing
  - Extremely unappealing

- ***The System's forward tilt movement features***

  - Very appealing
  - Somewhat appealing
  - Neither appealing nor unappealing
  - Somewhat unappealing
  - Very unappealing

- ***The System's backward tilt movement features***

  - Very unappealing
  - Somewhat unappealing
  - Neither appealing nor unappealing
  - Somewhat appealing
  - Very appealing

- ***The System's ability to nod***

  - Very unappealing
  - Somewhat unappealing
  - Neither appealing nor unappealing
  - Somewhat appealing
  - Very appealing

- ***The System's ability to shrug***

- Very appealing
- Somewhat appealing
- Neither appealing nor unappealing
- Somewhat unappealing
- Very unappealing

- ***I consider the System a pleasant conversational partner***

  - Definitely false
  - Probably false
  - Neither false nor true
  - Probably true
  - Definitely true

- ***I find the System pleasant to interact with***

  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree

- ***I feel the System understands me***

  - Definitely true
  - Probably true
  - Neither true nor false
  - Probably false
  - Definitely false

- ***I think the System is***

  - Very unappealing
  - Somewhat unappealing
  - Neither unappealing nor appealing
  - Somewhat appealing
  - Very appealing

- ***While using the System, the interaction with my partner feels***

- Very personal
- Somewhat personal
- Neither personal nor impersonal
- Somewhat impersonal
- Very impersonal

- ***While using the System, the interaction with my partner feels***

  - Very unnatural
  - Somewhat unnatural
  - Neither unnatural nor natural
  - Somewhat natural
  - Very natural

- ***While using the System, the interaction with my partner feels close***

  - Definitely false
  - Somewhat false
  - Neither false nor true
  - Somewhat true
  - Definitely true

- ***While using the System, the interaction with my partner feels***

  - Very human
  - Somewhat human
  - Neither human nor inhuman
  - Somewhat inhuman
  - Very inhuman

- ***My experience of interacting with my partner using the System is consistent with a real world experience***

  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree

- ***The forward movement of the System makes me feel as if I am having a face to face interaction with my partner***

- Definitely true
- Probably true
- Neither true nor false
- Probably false
- Definitely false

- ***The nod movement of the System makes me feel as if I am having a face to face interaction with my partner***

  - Strongly disagree
  - Somewhat disagree
  - Neither disagree nor agree
  - Somewhat agree
  - Strongly agree

- ***The forward tilt of the System makes me feel as if I am having a face to face interaction with my partner***

  - Definitely false
  - Probably false
  - Neither false nor true
  - Probably true
  - Definitely true

- ***The backward tilt of the System makes me feel as if I am having a face to face interaction with my partner***

  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree

- ***Interacting with my partner using the System is enjoyable***

  - Strongly disagree
  - Somewhat disagree
  - Neither disagree nor agree
  - Somewhat agree
  - Strongly agree

- ***Interacting with my partner using the System is not boring***

- Definitely false
- Somewhat false
- Neither false nor true
- Somewhat true
- Definitely true

- ***Interacting with my partner using the System is interesting***

  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree

- ***Interacting with my partner using the System is entertaining***

  - Definitely not
  - Probably not
  - Might not or might
  - Probably yes
  - Definitely yes

- ***Sometimes I lose track of time while I am interacting with my partner using the System***

  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree

- ***Time flies while I am interacting with my partner using System***

  - Definitely false
  - Somewhat false
  - Neither false nor true
  - Somewhat true
  - Definitely true

- ***While interacting with my partner using the System I am able to block out most other distractions***

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

- ***While interacting with my partner using the System, I am absorbed in what I am doing***

  - Definitely false
  - Somewhat false
  - Neither false nor true
  - Somewhat true
  - Definitely true

- ***While interacting with my partner using the System, I am immersed in the task I am performing***

  - Strongly disagree
  - Somewhat disagree
  - Neither disagree nor agree
  - Somewhat agree
  - Strongly agree

- ***Using the System is a***

  - Very good idea
  - Somewhat good idea
  - Neither good idea nor poor
  - Somewhat poor
  - Very poor

- ***The System will make conversation more interesting***

  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree

- ***Q40 The System is***

- Very inspiring
- Somewhat inspiring
- Neither inspiring nor uninspiring
- Somewhat uninspiring
- Very uninspiring

- ***Do not prefer to do video calling using this System***

  - Strongly agree
  - Somewhat agree
  - Neither agree nor disagree
  - Somewhat disagree
  - Strongly disagree

- ***I would like to experience this System again***

  - Definitely true
  - Probably true
  - Neither true nor false
  - Probably false
  - Definitely false

- ***This experience is not consistent with my expectations from the System***

  - Strongly disagree
  - Somewhat disagree
  - Neither disagree nor agree
  - Somewhat agree
  - Strongly agree

- ***I am satisfied with my experience of the System***

  - Definitely yes
  - Probably yes
  - Not sure
  - Probably not
  - Definitely not

- ***In your opinion, which experience did you think was the more successful, and why?***

- ***With which gender do you identify?***

- Male
- Female

- ***How old are you?***

  - 18-20 years old
  - 21-25 years old
  - 26 to 30 years
  - 31 to 35 years
  - 36 to 40 years
  - 41 to 45 years